



HAL
open science

Dynamiques des associations entre les fourmis amazoniennes et leurs communautés microbiennes

Anaïs Chanson

► To cite this version:

Anaïs Chanson. Dynamiques des associations entre les fourmis amazoniennes et leurs communautés microbiennes. Zoologie des invertébrés. Université de Guyane, 2021. Français. ⟨NNT : ⟩. ⟨tel-04820551⟩

HAL Id: tel-04820551

<https://theses.hal.science/tel-04820551v1>

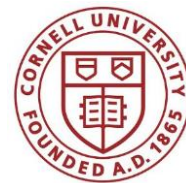
Submitted on 5 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



Thèse

En vue de l'obtention du grade de
Docteur de l'université de Guyane

Domaine : Biologie, médecine et santé

Discipline : Physiologie et biologie des organismes-populations-interactions

Présentée et soutenue publiquement le 28 juin 2021 par :

ANAÏS CHANSON

Dynamiques des associations entre les fourmis amazoniennes et leurs communautés microbiennes

JURY

Didier DEBROAS	PU	Rapporteur
Ludwig JARDILLIER	PU	Rapporteur
Joël MEUNIER	HDR, CR	Examineur
Heidy SCHIMANN	HDR, CR	Examinatrice
Christophe DUPLAIS	HDR, CR	Directeur de Thèse
Corrie MOREAU	PU	Co-Directrice de Thèse

Sommaire

Remerciements.....	6
Table des figures.....	8
Introduction générale	10
Etat de l'art général des connaissances	14
I. Structure des communautés.....	14
A. Généralités.....	14
B. Mesure de la structure d'une communauté	15
II. Facteurs façonnant la structures des communautés	17
A. Généralités.....	17
B. Facteurs abiotiques	18
C. Facteurs biotiques	19
D. Processus écologiques et évolutifs.....	20
E. Perturbations de la structure des communautés.....	22
F. Cas particulier des communautés microbiennes	22
III. Les différents types de symbioses avec micro-organismes	23
A. Généralités.....	23
B. Les types d'associations symbiotiques	24
C. Bénéfices de l'association symbiotique.....	26
IV. Bactéries productrices de métabolites secondaires.....	29
A. Généralités.....	29
B. Fonction des métabolites secondaires au sein des communautés	29
C. Intérêts pharmaceutiques.....	31
Etat de l'art général des méthodes.....	33
I. Séquençage de l'ADN	33
A. Préparation des échantillons	33
B. Techniques de séquençage	34
II. Techniques d'identification moléculaires post-séquençage	40
A. Le métabarcoding.....	40
B. Approche du multi-marqueur métabarcoding.....	41
C. La génomique structurale	43
D. La métagénomique.....	43
III. Techniques bio-informatiques d'analyse de données de séquençage.....	44
A. Analyse de données de métabarcoding.....	44
B. Analyse de données génomiques et métagénomiques.....	49

IV.	Techniques de mesure et de comparaison de la structure des communautés	55
A.	Diversité alpha.....	55
B.	Diversité bêta	57
C.	Analyse de pourcentage de similarité (SIMPER)	58
Chapitre I : Etude des facteurs structurant les communautés bactériennes associées aux fourmis amazoniennes.....		60
I.	Introduction	60
II.	Impact of Amazonian ants nesting mode, diet and taxonomy in structuring their associated bacterial communities.....	62
A.	Abstract.....	63
B.	Introduction	63
C.	Material and Methods	66
D.	Results	70
E.	Discussion	80
III.	Conclusion.....	91
Chapitre II : Etude des communautés eucaryotes associées aux fourmis amazoniennes.....		94
I.	Introduction	94
II.	Potential impact of host nesting mode in structuring their associated eukaryote communities 96	
A.	Abstract.....	97
B.	Introduction	97
C.	Material and Methods	99
D.	Results	102
E.	Discussion	109
III.	Conclusion.....	116
Chapitre III : Etude des clusters de gènes de biosynthèse des bactéries associées aux fourmis du genre <i>Cephalotes</i>		118
I.	Introduction	118
II.	Assessing biosynthetic gene cluster diversity of specialized metabolites in the conserved gut symbionts of herbivorous turtle ants.....	120
A.	Abstract.....	121
B.	Introduction	121
C.	Material and Methods	124
D.	Results	128
E.	Discussion	135
III.	Conclusion.....	142
Conclusion générale et perspectives		144

Annexes.....	148
Annexe 1. Supporting Information de l'article Impact of Amazonian ants nesting mode, diet and taxonomy in structuring their bacterial communities	148
Annexe 2. Supporting Information de l'article Potential impact of host nesting mode in structuring their associated eukaryote communities.....	209
Annexe 3. Supporting Information de l'article Assessing biosynthetic gene cluster diversity of specialized metabolites in the conserved gut symbionts of herbivorous turtle ants	236
Bibliographie	278

Remerciements

Tout d'abord, j'aimerais remercier les membres du jury de thèse qui ont accepté de prendre du temps pour juger mes travaux : les Dr. Didier Debroas, Ludwig Jardilier, Joël Meunier et Heidi Schimann. Je remercie les membres de mon comité de thèse, les Dr. Christophe Lucas, Jérôme Orivel et Kira Weissman pour avoir accepté de juger les avancements de mes recherches chaque année, ainsi que pour leurs nombreux conseils et remarques constructives. Je remercie également infiniment mes directeurs de thèse, les Dr. Christophe Duplais et Corrie Moreau. Je leur suis reconnaissante pour leur soutien et leur aide tout au long de ma thèse.

Au cours de cette thèse, j'ai eu la chance de profiter d'une grande mobilité et de travailler dans plusieurs laboratoires. En Guyane, mon bureau principal se trouvait dans l'unité de chimie d'EcoFoG à Cayenne. Je remercie les Dr. Yannick Estevez, Emeline Houël et Nadine Amusant ainsi que Cyrielle Sophie et Arnaud Jahn pour leur soutien pendant ces 4 années de thèse. En particulier, je remercie Yannick pour toute son aide dans le domaine informatique. Je remercie également Emeline et Yannick qui m'ont souvent déposé/ramené de l'aéroport quand je rentrais en métropole ou que je partais en visite dans un laboratoire aux USA. J'ai eu la chance passer une semaine dans la réserve naturelle des Nouragues pour collecter mes fourmis. Je remercie mes directeurs de thèse, ainsi que le Dr. Axel Touchard, Peter Flynn et Philippe Gaucher pour leur aide pendant cette semaine de collection.

J'ai également pu passer quelques semaines au laboratoire de génétique d'EcoFoG à Kourou. Je remercie le Dr. Niklas Tysklind de m'avoir permis de passer quelques semaines dans l'équipe de génétique d'EcoFoG, ainsi que le Dr. Jérôme Orivel pour m'avoir laissé utiliser un bureau auprès de ses autres étudiants pendant ces quelques semaines. Je remercie beaucoup Elianne Louisanna et Valérie Troispoux qui m'ont accueillies dans le laboratoire et m'ont prodiguées de précieux conseils pour la réalisation de mes PCR et les conseils en biologie moléculaire. Je remercie également Anthony Chassaing de m'avoir permis de loger sur le campus de Kourou pendant les quelques semaines où j'y étais présente. Je remercie également Josie Santini pour avoir pris en charge tous les bons de

commandes pour mes réactifs, ainsi que la réservation de mes différents billets d'avion. En plus de ma visite au laboratoire de génétique, j'ai eu la possibilité de faire plusieurs visites dans le laboratoire d'entomologie d'EcoFoG. Je remercie le Dr. Mélanie Fichaux ainsi que Darouny Outsama pour m'avoir appris à reconnaître les différents genres de fourmis, et m'avoir aidé à identifier les fourmis que j'avais collecté pour mes recherches.

Grâce à ma co-directrice Corrie Moreau, j'ai également eu la chance de faire plusieurs séjours dans ses laboratoires aux USA. Je remercie tous les membres du Moreau lab de Chicago et de Cornell pour leur accueil, leur bonne humeur et leur soutien pendant ces nombreuses visites. En particulier, je remercie le Dr Manuela Ramalho pour m'avoir appris à réaliser de nombreuses expériences en laboratoire, ainsi qu'à utiliser la plateforme Qiime2 pour analyser mes résultats de métabarcoding. Je remercie également le Moreau lab pour m'avoir permis de faire un entraînement pour ma défense de thèse, ainsi que pour tous leurs conseils pour les différentes présentations que j'ai fait au cours de cette thèse. Je remercie le Lucky Lab dans lequel j'ai eu la chance de passer une semaine pour apprendre l'écologie et la collection de fourmis invasives, et en particulier je remercie le Dr. Andrea Lucly pour m'avoir accueillie dans son laboratoire pendant cette semaine. Je remercie infiniment l'université de Guyane pour avoir financé mes différentes visites aux USA. Je remercie Cheryl Gombas de l'université de Cornell pour m'avoir aidé dans ma demande de visa.

Je remercie également mes directeurs de thèse pour m'avoir permis de participer à différentes conférences aux USA pendant ma thèse. J'ai pu ainsi présenter mes recherches ainsi que développer mon réseau scientifique. Je les remercie également pour leur aide pendant les analyses de résultats, leurs nombreuses idées, ainsi que toutes les relectures de mon manuscrit de thèse et de mes différents articles. Enfin, je remercie mes colocataires, collègues et ami-e-s en Guyane et aux USA, ainsi que mes ami-e-s en métropole et ma famille pour leur soutien tout au long de ma thèse

Table des figures

Figure 1. Exemple d'analyse de la structure d'une communauté.

Figure 2. Exemples de facteurs abiotiques et biotiques structurant une communauté.

Figure 3. Différentes voies de transmission pour les symbiontes.

Figure 4. Exemples de molécules antimicrobiennes produites par des bactéries.

Figure 5. Principe des différents types de séquençage.

Figure 6. Principes du métabarcoding et de la métagénomique

Figure 7. Résumé du processus de travail et exemples de commandes utilisées pour l'analyse de séquences de métabarcoding sur la plateforme Qiime2.

Figure 8. Résumé du processus de travail et exemples de logiciels utilisés pour l'analyse de séquences génomiques et métagénomiques.

Figure 9. Résumé du processus de travail et des logiciels utilisés dans le chapitre 3 pour l'analyse des clusters de gènes de biosynthèse à partir de génomes et métagénomes bactériens associés à l'intestin des fourmis du genre *Cephalotes*.

Figure 10. Résumé du processus de travail et exemples de commandes utilisées pour analyser la structure des communautés dans différents échantillons sur la plateforme Qiime2.

Figure 11. Differences in bacterial alpha diversity by ant nesting mode.

Figure 12 Differences in bacterial alpha diversity by ant genus.

Figure 13. Taxa barplot based on relative abundance ordered by ant subfamily, genus, habitat, nesting mode and diet.

Figure 14. Differences in bacterial communities across ant habitat, diet and nesting mode.

Figure 15. Differences in bacterial communities across ant subfamily, genus and phylogeny.

Figure 16. Differences in eukaryote alpha diversity by ant nesting mode.

Figure 17. Differences in eukaryote communities across ant habitat and nesting mode.

Figure 18. Differences in eukaryote communities across ant subfamily, genus and phylogeny.

Figure 19. Taxa barplot based on relative abundance ordered by ant subfamily, genus, habitat and nesting mode.

Figure 20. Taxa barplot based on relative abundance ordered by ant subfamily, genus, habitat nesting mode after removing all the undetermined ASVs.

Figure 21. Diversity of biosynthetic gene clusters of bacteria associated with *Cephalotes* turtle ants.

Figure 22. Genetic networks of the BGCs identified in bacterial genomes and metagenomes from *Cephalotes* gut.

Figure 23. CORASON phylogenetic reconstruction of arylpolyene, NRP, T1PK and siderophore biosynthetic gene clusters.

Figure 24. Phylogeny of retrieved C and KS elongation domain sequences in the NRP and PK assemblies respectively.

Introduction générale

L'écologie des communautés est l'étude des interactions entre différentes communautés et leur écosystème. Ces interactions sont soumises à différents facteurs, comme les conditions environnementales ou la disponibilité des ressources présentes dans un habitat. En effet, dans un contexte où les ressources sont limitées, les organismes utilisant les mêmes ressources vont entrer en compétition pour ces ressources. De plus, la composition des communautés peut varier grandement en termes d'abondance et de diversité. Ainsi, pour comprendre comment les interactions entre communautés sont structurées et maintenues, il est nécessaire de prendre en compte ces paramètres.

Parmi ces interactions, les interactions entre certains organismes et leurs communautés microbiennes sont particulièrement étudiées, en partie pour les avantages conférés par de nombreux microbes à leur hôte dans ces associations. Un de ces avantages est la supplémentation nutritionnelle. Chez certaines insectes par exemple, les communautés bactériennes mutualistes peuvent compléter le régime alimentaire de leur hôte lorsque celui-ci est nutritionnellement insuffisant, comme c'est le cas pour les insectes herbivores¹ ou suceurs de sang². Chez d'autres insectes, une partie de leur régime alimentaire est basée sur un champignon mutualiste cultivé par les individus directement dans le nid³⁻⁵. Un autre avantage de ces associations est la protection de l'hôte. Par exemple, certaines bactéries peuvent renforcer le système immunitaire de leur hôte⁶. De plus, de nombreux microbes sont capables de produire des molécules à activités antimicrobiennes, dont certaines peuvent aider à défendre l'hôte contre des pathogènes⁷. A l'inverse, les communautés microbiennes peuvent aussi être nocives aux hôtes auxquels elles sont associées. En effet, certains microbes sont capables de modifier le comportement de leur hôte^{8,9}. De plus, de nombreuses classes de microbes contiennent des organismes pathogènes des insectes, comme chez les champignons¹⁰, les nématodes¹¹, ou encore les straménophiles¹².

Dans le cadre de cette thèse, nous avons choisi d'étudier les différents facteurs structurant les communautés microbiennes en se focalisant sur les communautés microbiennes associées aux fourmis

amazoniennes. Nous avons ciblé ces systèmes, pour différentes raisons. Tout d'abord, l'Amazonie abrite une grande diversité d'organismes. En effet, la forêt Amazonienne est la forêt contenant le plus de diversité au monde, et sa biodiversité contribue majoritairement à la biodiversité totale présente dans les néo-tropiques¹³. Ensuite, les fourmis sont les principaux responsables de plusieurs processus écologiques majeurs des écosystèmes, comme la redistribution des nutriments sur la litière du sol¹⁴. Enfin, les fourmis sont extrêmement diverses en Amazonie, en termes de phylogénies et d'écologies. Au niveau phylogénique, les néo-tropiques abritent 30% des espèces connues sur les 13,000 espèces décrites, et les fourmis et termites représentent 30% de la biomasse animale totale présente dans les néo-tropiques¹⁵. Au niveau écologique, les fourmis amazoniennes ont des régimes alimentaires et des modes de nidification très variés. Par exemple, les fourmis sont les principaux prédateurs d'arthropodes et les principaux herbivores présents dans les néo-tropiques. De plus, les communautés microbiennes, et en particulier communautés bactériennes, associées à ces fourmis pourraient être en partie responsables de ces diverses écologies des fourmis. En effet, les communautés bactériennes pourraient avoir permis aux fourmis de dominer la canopée des forêts tropicales¹⁶, et les communautés bactériennes associées aux fourmis sont différentes entre fourmis herbivores et fourmis carnivores¹⁷. Pour ces différentes raisons, étudier les associations entre communautés microbiennes et fourmis amazoniennes nous permet d'étudier ces interactions le long d'une large phylogénie d'hôte et d'y associer ces facteurs écologiques diversifiés pour ainsi déterminer quels facteurs interviennent dans la structure et la maintenance de ces associations.

L'objectif général de cette thèse est d'étudier les différents facteurs intervenant dans la structure et la maintenance des communautés microbiennes associées aux fourmis amazoniennes. Ce travail de thèse est organisé sur trois chapitres, chacun répondant à des sous-objectifs particuliers mais s'intégrant dans l'objectif général de la thèse. Chaque chapitre correspond à un article soumis ou en préparation de soumission, et par conséquent les trois chapitres sont rédigés en anglais dans ce manuscrit. Le chapitre 1 est une étude des communautés bactériennes associées à différentes espèces de fourmis amazoniennes. Les objectifs de ce chapitre sont (1) d'étudier les différentes communautés

bactériennes associées à ces fourmis, et (2) de déterminer quels facteurs environnementaux, écologiques et évolutifs participent à structurer les communautés bactériennes au sein de ces associations. Pour cela, nous avons collecté 38 espèces de fourmis différentes dans la forêt Amazonienne et dans la ville de Cayenne. Nous avons réparti ces fourmis en plusieurs catégories selon leur régime alimentaire et type de nid. Nous avons ensuite utilisé la technique du métabarcoding en ciblant le gène 16S rRNA, qui est un gène marqueur des procaryotes, pour déterminer les communautés bactériennes associées à chaque échantillon de fourmi.

Le chapitre 2 est une étude des communautés eucaryotes associées à différentes espèces de fourmis amazoniennes. La méthodologie est similaire à celle utilisée dans le chapitre 1, mais cette fois le gène cible du métabarcoding est le 18S rRNA, qui est un gène marqueur des eucaryotes. Les objectifs de ce chapitre sont (1) d'étudier les différentes communautés eucaryotes associées à ces fourmis, (2) de déterminer quels facteurs environnementaux, écologiques et évolutifs participent à structurer les communautés bactériennes au sein de ces associations, et (3) de mettre en évidence de potentiels patrons de co-occurrence entre communautés bactériennes et communautés eucaryotes associées aux mêmes échantillons de fourmis.

Enfin, le chapitre 3 est l'étude des clusters de gènes de biosynthèse présents dans les génomes et métagénomes de communautés bactériennes associées aux fourmis du genre *Cephalotes*. Les clusters de gènes de biosynthèse sont responsables de la production de métabolites secondaires, dont certains jouent un rôle dans les interactions entre communautés bactériennes, ainsi que dans les interactions entre ces communautés bactériennes et leur hôte. Les objectifs de ce chapitre sont d'étudier (1) la présence et (2) la diversité des clusters de gènes de biosynthèse présents dans les communautés bactériennes associées à 17 espèces de fourmis *Cephalotes*. Pour cela, nous avons utilisé les génomes et métagénomes de bactéries intestinales des fourmis *Cephalotes* qui ont été publiés récemment¹⁸. L'originalité de cette étude est l'analyse de clusters de gènes de biosynthèse au sein de métagénomes. En effet, bien que l'analyse de ces clusters soit fréquente dans les études génomiques, l'utilisation de métagénomes pour rechercher ces mêmes clusters est encore très peu

répandue. Ce chapitre continue l'exploration fascinante des symbiontes du tube digestif avec les fourmis *Cephalotes*.

Etat de l'art général des connaissances

I. Structure des communautés

A. Généralités

En écologie, une communauté est définie par l'ensemble des populations (animales, végétales et microorganismes) qui vivent et interagissent ensemble dans une zone géographique spécifique¹⁹. La composition d'une communauté se traduit par le nombre et l'abondance des différentes espèces et populations qu'elle contient, et également par les différentes interactions entre populations : c'est la structure de la communauté.

Une communauté peut être composée de plusieurs types de populations : les populations fondatrices, les populations clés de voûte et les populations invasives. Les populations fondatrices sont les populations qui ont fondé une communauté spécifique²⁰. Elles jouent un rôle primordial dans la structure de la communauté et ce sont des producteurs primaires. C'est le cas des coraux dans les récifs par exemple. Les populations clés de voûte jouent également un rôle très important dans le maintien des communautés car elles sont souvent en interaction avec la plupart des autres populations composant la communauté²¹. Elles permettent à la communauté de conserver son équilibre écologique et sans ces populations la communauté serait grandement menacée. C'est par exemple le rôle que jouent les prédateurs dans la régulation de la population de leurs proies, mais également dans la population des plantes dont se nourrissent les proies. Les populations invasives sont des populations qui ne sont pas natives de la communauté mais y ont été introduites, souvent par des moyens humains (bateaux, avions)²². Les populations invasives vont souvent causer un déséquilibre dans la communauté en entrant en compétition avec les espèces natives, allant à risquer l'extinction des espèces natives.

B. Mesure de la structure d'une communauté

1. Définitions

Pour décrire la structure d'une communauté et les possibles changements que cette communauté subit, il faut pouvoir analyser sa composition. La composition d'une communauté comprend deux mesures importantes : la richesse spécifique d'espèce et l'uniformité des espèces (Figure 1). La richesse spécifique d'espèces est le nombre d'espèces différentes que l'on peut trouver dans une communauté. L'uniformité des espèces est la mesure de l'abondance relative de chaque espèce. Ensemble ces deux mesures permettent d'analyser la diversité des espèces dans une communauté.

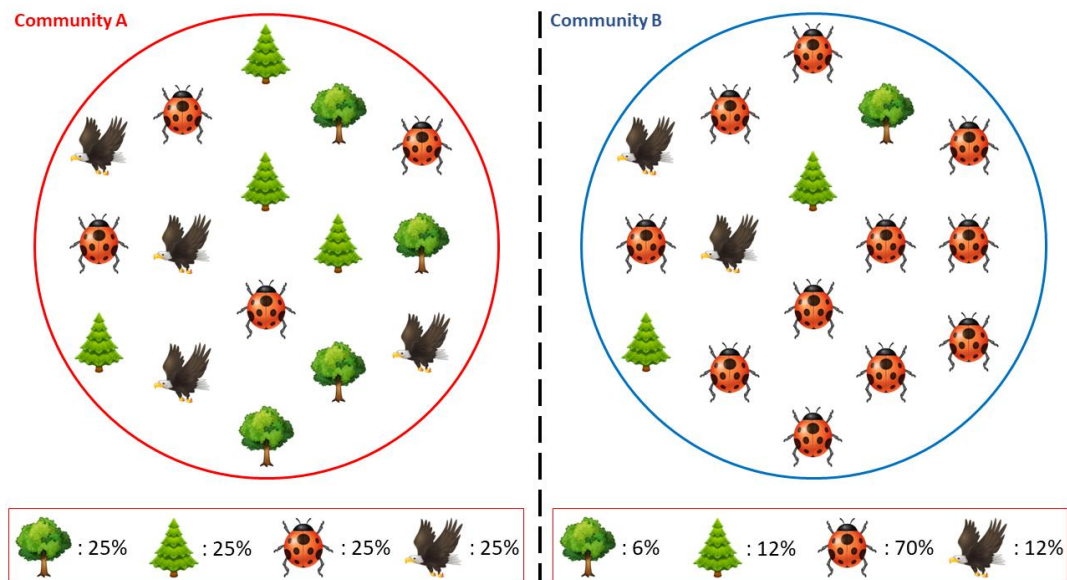


Figure 1. Exemple d'analyse de la structure d'une communauté. Les communautés A et B ont la même richesse spécifique d'espèce car elles contiennent quatre espèces différentes. En revanche, l'uniformité des espèces dans les communautés A et B est différente. La communauté A est uniforme car elle contient le même nombre d'individus de chaque espèce, alors la communauté B n'est pas uniforme car les espèces qu'elle contient ne possèdent pas toutes le même nombre d'individus.

2. Richesse spécifique d'espèces

La mesure de la richesse spécifique d'espèces permet de calculer des indices de richesse spécifique d'espèces, qui sont ensuite utilisés pour étudier la composition de la communauté (Figure 1). Il existe plusieurs indices de richesse spécifique d'espèces. Les plus simples sont calculés uniquement en utilisant le nombre d'espèces différentes (S) et le nombre total d'individus (N). On peut citer 2 exemples de ces indices de richesse spécifique d'espèces :

- l'indice de Margalef²³ : $D = S - 1 \ln N$
- l'indice de Menhinick²⁴ : $D = SN$

Ces différents indices utilisent différentes fonctions pour évaluer la richesse spécifique d'espèces d'une communauté, et généralement il convient de comparer et d'intégrer plusieurs de ces indices lorsque l'on étudie une communauté.

Les indices de richesse spécifique d'espèces ont cependant deux limites majeures. La première limite est que le nombre d'espèce différentes que l'on peut trouver dans une communauté dépend fortement de l'échantillonnage²⁵. En effet, plus la taille de l'échantillonnage est grande, plus les chances d'obtenir toutes les espèces présentes dans une communauté seront grandes. Cela est d'autant plus vrai si certaines de ces espèces sont rares dans la communauté. Il est donc nécessaire d'utiliser des techniques d'échantillonnage standardisées, surtout si le but est de comparer la richesse spécifique d'espèces entre plusieurs communautés. La seconde limite est le besoin de pouvoir identifier à quelle espèce appartient chaque organisme de la communauté. Cela peut être particulièrement problématique dans certaines communautés végétales et microbiennes où la taxonomie des différents organismes peut être difficile à déterminer. Dans ce cas, il convient d'utiliser un niveau taxonomique plus large (comme la famille ou l'ordre) afin de pouvoir assigner le niveau taxonomique choisit au plus grand nombre d'organismes de la communauté avant de calculer la richesse spécifique.

3. L'uniformité des espèces

Une communauté est uniforme si toutes les espèces qu'elle contient possèdent le même nombre d'organismes (Figure 1). Tout comme pour la richesse spécifique d'espèces vue précédemment, il existe plusieurs indices de calcul d'uniformité des espèces. On peut citer deux exemples de ces indices d'uniformité des espèces, utilisant également le nombre d'espèces différentes (S) et le nombre total d'individus (N) :

- l'indice d'uniformité de Pielou²⁶ : $E = \frac{-\sum_{i=1}^R \frac{n_i}{N} \ln \frac{n_i}{N}}{\ln S}$
- l'indice d'uniformité de McIntosh²⁷ : $E = \frac{\sqrt{\sum_{i=1}^R n_i^2}}{\sqrt{((N-S+1)^2 + S-1)}}$

Tout comme les indices de richesse spécifique, les indices d'uniformité des espèces doivent être comparés et intégrés lorsque l'on étudie une communauté. De plus, les indices d'uniformité des espèces sont soumis aux mêmes limites que les indices de richesse spécifique : la taille de l'échantillonnage et l'assignation taxonomique des organismes.

Pour étudier une communauté, on utilise généralement des indices qui combinent la richesse spécifique et l'uniformité des espèces pour calculer la diversité des espèces d'une communauté. C'est ce qu'on appelle la diversité alpha. Des indices ont également été développés pour comparer la diversité des espèces entre plusieurs communautés, appelée diversité bêta. Certains de ces indices seront présentés dans la partie État de l'art général des méthodes.

II. Facteurs façonnant la structures des communautés

A. Généralités

La science qui étudie les interactions entre populations et les facteurs façonnant la structure des communautés s'appelle l'écologie des communautés. Les communautés sont en constante évolution et ces évolutions sont liées à plusieurs facteurs : les facteurs abiotiques, les facteurs biotiques, les

processus écologiques et les grandes perturbations. Les facteurs abiotiques sont les interactions de l'environnement sur les communautés (Figure 2). Les facteurs biotiques sont les interactions des différentes populations entre elles dans les communautés (Figure 2). Les processus écologiques ont une grande influence sur la structure des communautés mais ils ne font partie ni des facteurs abiotiques, ni des facteurs biotiques. Ils sont parfois appelés facteurs aléatoires. Enfin, les grandes perturbations sont des processus qui peuvent grandement affecter la structure d'une communauté. Ces perturbations peuvent être influencées par les facteurs abiotiques et biotiques.

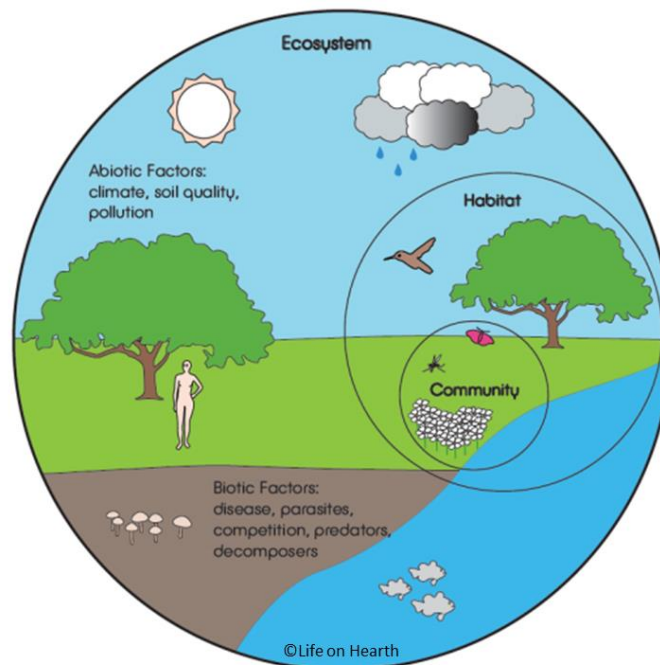


Figure 2. Exemples de facteurs abiotiques et biotiques structurant une communauté.

B. Facteurs abiotiques

Les facteurs abiotiques sont une importante composante de la structure des communautés, puisqu'ils représentent tous les facteurs de l'environnement qu'une population doit pouvoir supporter pour survivre dans cette communauté (Figure 2). Les facteurs abiotiques sont nombreux et de différents types.

Parmi les plus étudiés, on retrouve les facteurs climatiques, comme la température et la lumière. Les facteurs chimiques sont également importants certaines communautés, comme les plantes : le pH, le dioxygène et le dioxyde de carbone, et l'apport en certains éléments minéraux du sol essentiels. Enfin, certains facteurs abiotiques sont particulièrement essentiels dans certains écosystèmes. C'est par exemple le cas de la salinité et des courants dans les communautés marines. La pression peut également jouer un rôle important dans les communautés marines ou en altitude.

L'ensemble de ces facteurs abiotiques représentent ce qu'on appelle en écologie le filtre environnemental²⁸. Cela signifie que pour qu'une population puisse s'établir ou persister dans une communauté, elle doit pouvoir s'adapter aux conditions environnementales locales. Ainsi, les populations qui ne sont pas capables de s'adapter à cet environnement seront « filtrées » de l'environnement et des communautés locales.

C. Facteurs biotiques

Une fois le filtre environnemental passé, les différentes populations de la communauté vont entrer en interaction les unes avec les autres (Figure 2). Les populations dans une communauté peuvent être classés selon leur niveau trophique.

Les niveaux trophiques sont classés en pyramide du premier, comportant les producteurs primaires comme les plantes, au quatrième ou cinquième contenant les prédateurs supérieurs²⁹. Les populations de différents niveaux trophiques ont une relation de proies-prédateurs, où les populations des niveaux trophiques supérieures se nourrissent des populations des niveaux trophiques inférieures. Les populations des niveaux trophiques inférieures sont souvent plus nombreuses que les populations des niveaux trophiques supérieures. En revanche, les populations de niveaux trophiques similaires se nourrissent du même type de ressources et peuvent donc entrer en compétition pour ces ressources. Si deux populations sont en compétition pour la même ressource, alimentaire ou territoriale, une des deux populations finira par exclure l'autre, diminuant son abondance et pouvant aller jusqu'à l'exclure de la communauté.

Afin de diminuer la compétition entre populations, les populations ayant des écologies similaires peuvent différencier leurs niches, que ce soit leur niche temporelle, territoriale ou alimentaire. Par exemple, une étude portant sur la coexistence locale entre différentes espèces de fourmis en Guyane a montré que certaines de ces espèces ont une préférence pour un type de ressources alimentaires spécifique, ou un moment de la journée différent pour collecter leurs ressources alimentaires³⁰. Cette différenciation de niche alimentaire et temporelle pourrait participer à expliquer la coexistence de ces différentes espèces de fourmis. Un autre système de coexistence entre populations compétitrices est la théorie du « cher voisin ». Cette théorie stipule que dans une communauté où les territoires sont bien établis, les populations occupant des territoires voisins seront de moins en moins agressives l'une envers l'autre au fur et à mesure du temps³¹. Ces populations finiront par totalement tolérer leur voisin habituel, mais deviendront extrêmement agressives envers un nouveau voisin inconnu. Enfin, certaines populations peuvent aussi développer des relations de symbiose entre elles. Ce concept sera expliqué plus en détail dans la troisième partie de cet état de l'art.

D. Processus écologiques et évolutifs

Il existe quatre processus écologiques et évolutifs principaux qui peuvent influencer sur la structure des communautés : la spéciation, la sélection naturelle, la dispersion et la dérive écologique.

La spéciation est la différenciation d'une population en plusieurs, créant ainsi de nouvelles espèces³². La spéciation est allopatrique si les nouvelles populations s'éloignent physiquement de la population initiale. La spéciation est péripatrique si une des nouvelles populations est beaucoup plus petite que la seconde. La spéciation est parapatricque si les deux nouvelles populations s'éloignent physiquement, mais continuent de partager certaines niches. Enfin, la spéciation est sympatrique si elle n'est pas issue d'une séparation géographique mais d'une différenciation des traits par la sélection naturelle.

La sélection représente les différences de succès reproducteurs entre plusieurs populations. Il existe plusieurs formes de sélection selon si cette sélection est dépendant d'autres facteurs³³. La

première forme de sélection est la sélection constante, où le succès reproducteur des différentes populations dans une communauté est indépendant de conditions spatiales, temporelles ou de la densité des populations. La deuxième forme de sélection est la sélection densité-dépendante, si le succès reproducteur d'une population dépend de la densité de cette population et des autres populations de la communauté. La dernière forme de sélection est la sélection spatiotemporelle-dépendante, si le succès reproducteur d'une population dépend de conditions spatiales ou temporelles.

La dispersion représente un déplacement des populations dans l'espace et potentiellement entre plusieurs communautés³⁴. Cela entraîne des différences dans l'abondance des populations. Les causes de dispersion peuvent être diverses : reproduction, compétition, migration. La dispersion peut également être active pour les organismes motiles, comme les animaux, ou passive pour les organismes statiques, comme les plantes ou les champignons. Dans le cas de la dispersion passive, la dispersion est effectuée par les conditions environnementales (courants et vents) ou par le déplacement de communautés animales vecteurs.

La dérive écologique représente les changements de taille de populations dans une communauté, dus au nombre arbitraire de naissance et de décès dans les différentes populations³⁵. La dérive écologique existe dans toutes les populations, mais son importance est plus grande dans les populations de plus petite taille. Quel que soit la taille des populations, à terme seulement une population prévaudrait et les autres risqueraient de s'éteindre.

En outre, ces quatre processus ne sont pas indépendants les uns des autres : chaque processus peut avoir une influence sur les autres. Par exemple, la spéciation dans une communauté peut créer des populations qui n'ont pas le même succès reproductif, et donc la sélection aura des effets différents sur ces nouvelles populations. Une population avec un plus faible succès reproductif pourrait avoir une population diminuante, et cette diminution serait accélérée par les processus de dispersion et de dérive écologique.

E. Perturbations de la structure des communautés

Certaines perturbations particulières peuvent aussi drastiquement modifier la structure d'une communauté. Par exemple, les perturbations naturelles comme les feux de forêt ou les inondations peuvent totalement détruire certaines populations et complètement déstructurer leur communauté. Ces perturbations sont suivies d'un processus appelé successions écologiques secondaires. Ce processus est une série de nombreux changements des structures communautaires suivant une grande perturbation³⁶. Par exemple, après un feu de forêt, les communautés bactériennes vont modifier le sol, puis une nouvelle végétation va pouvoir s'établir. Ensuite, cette végétation va attirer de nouvelles communautés animales, et la communauté va se restructurer avec toutes ces nouvelles populations.

Une seconde perturbation majeure qui peut drastiquement modifier les structures communautaires est l'urbanisation. L'urbanisation est le fait de modifier un terrain naturel en terrain développé pour être utilisé par l'homme, comme une ville ou une parcelle agricole. L'urbanisation peut modifier de nombreux facteurs abiotiques d'un écosystème, comme le drainage, la température, le cycle des nutriments du sol, le pH ou la lumière³⁷. Comme vu précédemment, la modification de ces facteurs abiotiques va modifier les différentes populations qui peuvent passer le filtre environnemental. L'urbanisation va aussi modifier les facteurs biotiques en altérant ou même supprimant la végétation. Tous ces changements peuvent donc modifier les populations vivantes dans une communauté urbanisée.

F. Cas particulier des communautés microbiennes

Les communautés microbiennes sont également soumises aux facteurs vus précédemment. Cependant, les communautés microbiennes auraient une capacité de dispersion illimitée, ce qui les différencierait des autres communautés animales et végétales³⁸. Les communautés microbiennes ont une dispersion passive et un taux de reproduction élevé, augmentant leurs facultés de dispersion.

Grâce à ces facultés de dispersion, les communautés microbiennes pourraient donc théoriquement s'établir n'importe où. Cependant, l'établissement des communautés microbiennes est tout de même soumis à deux facteurs principaux : le filtre environnemental et les interactions avec d'autres communautés microbiennes. En effet, même si une communauté bactérienne peut se disperser n'importe où, elle n'est pas forcément adaptée à la survie de n'importe quel environnement, et les conditions météorologiques (température, humidité, pH, pression) ou la disponibilité des nutriments nécessaires peuvent empêcher une communauté microbienne de s'établir à un endroit spécifique. De plus, les interactions avec d'autres communautés microbiennes, comme la compétition ou la prédation, vont également limiter les possibilités des communautés microbiennes à s'établir³⁹.

III. Les différents types de symbioses avec micro-organismes

A. Généralités

Comme expliqué précédemment, une communauté est définie par l'ensemble des populations qui vivent et interagissent ensemble. Cependant, au sein d'une communauté, certaines populations sont étroitement associées par des interactions spécifiques : les interactions symbiotiques.

La définition de symbiose en biologie a longtemps été controversée, entre la définition d'Anton de Bary en 1879 qui considère une symbiose comme la vie en association entre deux espèces différentes⁴⁰, et la définition restrictive qui considère qu'une symbiose est une association forcément bénéfique aux deux espèces associées. Cependant, il semblerait que la définition restrictive soit désormais moins utilisée par la communauté scientifique au profit de la définition de de Bary⁴¹. C'est donc la définition de de Bary que nous utiliserons dans cette thèse.

L'ensemble des micro-organismes vivant en symbiose avec un même hôte est appelée le microbiote, tandis que la niche écologique regroupant ces micro-organismes est appelée le microbiome. Le terme microbiome désigne également l'ensemble du microbiote. Durant cette thèse,

nous nous sommes concentrés sur les associations symbiotiques entre 3 populations spécifiques : les hôtes et les différents organismes (bactéries et eucaryotes) qu'ils peuvent héberger sur eux et en eux.

Les associations symbiotiques peuvent être transitoire et de courte durée ou maintenue évolutivement depuis des centaines, voire des milliers d'années. Ces associations peuvent également être facultatives si elles ne sont pas nécessaires à la survie des micro-organismes, ou obligatoires si les micro-organismes ne sont pas capables de survivre en dehors de leur hôte.

B. Les types d'associations symbiotiques

Les associations symbiotiques hôtes-symbionte sont très généralisées et extrêmement étudiées. Il existe trois grands types d'associations symbiotiques, classées selon le bénéfice ou le désavantage que chaque partenaire retire de l'association : le parasitisme, le mutualisme et le commensalisme⁴².

Le parasitisme est une association où le symbionte tire profit de son hôte pour ses propres besoins. L'hôte ne retire aucun bénéfice de cette association, et souvent en subit des effets négatifs pouvant le rendre malade et aboutir à sa mort. Le mutualisme est une association où le symbionte et l'hôte retirent tous deux des effets bénéfiques de l'association. Enfin, le commensalisme est une association où le symbionte utilise son hôte sans lui induire d'effet ni positif ni négatif. L'absence d'effet étant impossible à démontrer, les associations symbiotiques causant un faible effet positif ou négatif à l'hôte sont également souvent qualifiées de commensalisme⁴³. De plus, une apparente absence d'effet représente souvent un manque de connaissance actuel sur les effets positifs ou négatifs d'une associations symbiotiques sur l'hôte, et ces effets peuvent évoluer pendant le cycle de vie du symbionte ou de l'hôte. En outre, les symbioses sont souvent multipartite, ce qui signifie qu'elles impliquent de nombreux micro-organismes qui peuvent avoir des effets coordonnés ou contraires sur leur hôte⁴⁴. Ces limites rendent la transition entre mutualisme, commensalisme et parasitisme d'autant plus fréquente.

Les micro-organismes symbiotiques sont acquis et transmis à leur hôte de différentes façons. Certains micro-organismes sont transmis par voie horizontale de la mère à sa progéniture (Figure 3A).

C'est le cas par exemple des endosymbiontes, qui sont des micro-organismes intracellulaires⁴⁵. D'autres micro-organismes peuvent être transmis socialement entre individus d'une même population. Chez les insectes sociaux, les micro-organismes sont transmis socialement dans la colonie de quatre façons : contact social, inoculation sur les œufs, coprophagie et trophallaxie⁴⁶ (Figure 3B). La dernière voie de transmission est la transmission horizontale par l'environnement (Figure 3C).

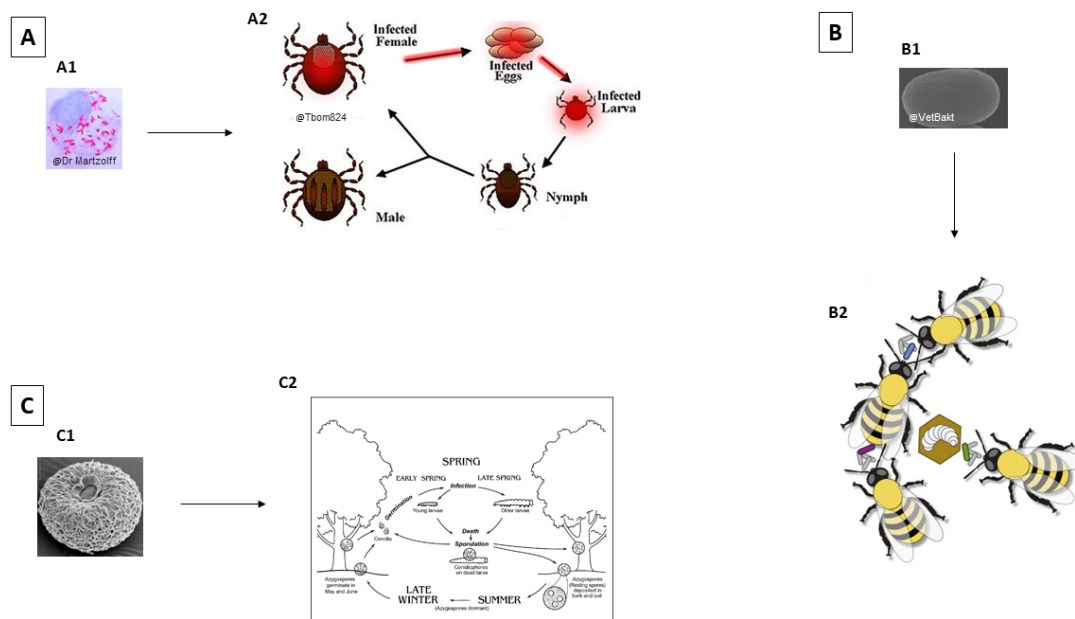


Figure 3. Différentes voies de transmission pour les symbiontes. (A) Exemple de vertical transmission : la bactérie *Rickettsia rickettsii* (A1) est transmise par transmission transovarienne chez certaines tiques⁴⁷ (A2). (B) Exemple de transmission sociale : la bactérie *Gilliamella apicola* (B1) est transmise par coprophagie chez certaines abeilles⁴⁸ (B2). (C) Exemple de transmission horizontale : les spores de champignons pathogènes de l'embranchement Entomophthoromycotina (C1) sont transmis par l'environnement chez de nombreux insectes⁴⁹ (C2).

C. Bénéfices de l'association symbiotique

1. Bénéfices pour les symbiotes

Les bénéfices des associations symbiotiques pour les symbiotes ont été moins étudiés que les bénéfices pour les hôtes et ceci pour deux raisons principales. La première raison est que l'hôte est souvent plus grand que ses symbiotes et plus facile à manipuler. La seconde raison est qu'il est souvent sous-entendu qu'un micro-organisme tirera forcément des bénéfices d'une symbiose, que ce soit du parasitisme, du commensalisme ou du mutualisme.

Un des bénéfices de vivre en symbiose pour les micro-organismes serait une plus faible compétition entre micro-organismes chez un hôte que dans la nature. Ceci est particulièrement vrai dans les systèmes contenant des populations de symbiotes monoclonaux⁵⁰, mais la plupart des symbioses contiennent des populations de symbiotes polyclonaux. Dans les systèmes polyclonaux, la compétition peut entraîner la diminution d'abondance de toutes les espèces présentes⁵¹. Cependant cette compétition peut aussi être bénéfique pour certains symbiotes dont l'abondance augmente si des symbiotes de génotypes différents sont présents⁵². En outre, la vie en symbiose peut permettre aux symbiotes d'être moins la cible de prédateurs et pathogènes. Ceci est surtout vrai dans les systèmes mutualistes, où l'hôte bénéficie aussi de la symbiose et donc a besoin de garder son symbiote en bonne santé. Par exemple, certains insectes cultivateurs de champignons utilisent des bactéries Actinomycètes pour protéger leur champignon cultivé de champignons pathogènes^{53,54}, ainsi que l'élimination manuelle de champignons pathogènes⁵⁵.

Un second bénéfice souvent rapporté pour les symbiotes, et particulièrement pour les bactéries intestinales, est l'assimilation de nutriments dérivés de la nourriture de l'hôte. Ce bénéfice est illustré par la théorie de la niche nutritionnelle proposée par Rol Freter en 1983^{56,57}. Selon cette théorie, les bactéries ne peuvent coloniser l'intestin d'un hôte que si elles sont capables de capter des nutriments spécifiques et limités. Cette théorie a été prouvée par l'utilisation de probiotiques pour modifier expérimentalement les nutriments disponibles dans l'intestin d'un hôte, résultant dans la modification de ses bactéries intestinales⁵⁸. Dans les systèmes parasites, les symbiotes peuvent aussi profiter de

leur hôte pour se nourrir. Par exemple, les champignons ascomycètes pathogènes des insectes se nourrissent du cadavre de leur hôte après sa mort⁵⁹.

Un dernier bénéfice très important pour les symbiotes est la multiplication et la dispersion. Chez les insectes, de nombreuses espèces mutualistes de champignons ont évolués des organes spécifiques pour transporter les spores de leur symbiotes⁶⁰. Dans les systèmes parasites, le champignon peut également utiliser son hôte pour se disperser. Par exemple, un champignon de l'ordre des Sordariomycètes peut infecter certaines fourmis, se développant dans leur cerveau puis modifiant le comportement de l'hôte pour le forcer à grimper sur une feuille près de sa colonie⁹. Cela permet au champignon de disperser ses spores sur toute la colonie de son hôte. Dans un autre exemple, de nombreuses bactéries parasites d'arthropodes sont connus pour modifier la reproduction de leur hôte à leur avantage de différentes manières⁸ : asexualité de l'hôte, féminisation de l'hôte, incompatibilité entre mâles infectés et femelles non infectées, mort des mâles, ou être nécessaire pour l'ovogenèse.

2. Bénéfices pour les hôtes

Comme expliqué précédemment, lors d'une association symbiotique l'hôte ne reçoit un bénéfice que si l'association est mutualiste. Un des bénéfices les plus étudiés des associations mutualistes est la supplémentation nutritionnelle de l'hôte par son symbiote. En effet, de nombreux symbiotes participe à la décomposition des nutriments et fournissent de nombreux substrats énergétiques à leur hôte, comme des acides aminés ou des vitamines. Chez de nombreux herbivores vertébrés et invertébrés par exemple, il a été démontré que leur communauté bactérienne leur permet de compléter leur régime alimentaire en aidant à digérer la cellulose⁶¹, en diminuant les défenses des plantes cibles⁶², en fournissant de la vitamine B², ou en recyclant l'azote pour le synthétiser en acide aminés⁶³. Chez les carnivores, leur communauté bactérienne leur permet de digérer leur haut niveau de protéines et de matières grasses⁶⁴, et de décomposer la chitine de leurs proies⁶⁵. Enfin, certains hôtes se nourrissent directement de leur symbiote. C'est le cas des insectes cultivateurs de

champignons, qui cultivent un champignon directement dans leur nid et s'en nourrissent exclusivement³⁻⁵.

En plus de la supplémentation nutritionnelle, les symbiotes permettent également à leur hôte une adaptation environnementale, leur permettant d'habiter des niches territoriales très diverses en fonction du symbiote. Par exemple, de nombreuses bactéries et champignons ont un effet sur la tolérance de températures extrêmes, qu'elles soient très froides ou très chaudes, chez les plantes⁶⁶ ou les insectes⁶⁷. Les bactéries et champignons mutualistes procurent également à leur hôte plante ou insecte une meilleure résistance aux fluctuations d'humidité afin de survivre à des épisodes de sécheresse⁶⁸ ou d'inondations⁶⁹. Ces résistances environnementales conférées par les symbiotes pourraient s'avérer encore plus importante du fait des phénomènes de changements climatiques toujours plus nombreux désormais.

Enfin, un troisième bénéfice important des symbioses mutualistes est le renforcement des défenses immunitaires et de la santé de l'hôte. L'importance des bactéries, et des bactéries intestinales en particulier, dans le maintien du système immunitaire et du développement des organismes ont été démontrés à de nombreuses reprises chez les vertébrés⁷⁰ comme chez les insectes⁷¹. De plus, les hôtes peuvent profiter de la production de molécules secondaires à activité antimicrobiennes produites par de nombreuses bactéries et champignons⁷. Cette défense chimique permet aux hôtes de réduire leurs infections par de nombreux pathogènes, comme les bactéries ou les champignons, mais également de protéger leurs ressources alimentaires⁷², leur nid⁷³ ainsi que leurs œufs^{74,75}.

IV. Bactéries productrices de métabolites secondaires

A. Généralités

La plupart des organismes vivants sont capables de produire des métabolites secondaires. Ces métabolites secondaires, aussi appelées métabolites spécialisés, sont des molécules organiques qui ne sont pas directement nécessaires à la survie des organismes, mais qui offrent aux organismes les produisant des avantages leur permettant de maximiser leur croissance, leur survie, et leur succès reproducteur.

Chez les bactéries les métabolites secondaires sont produits en réponse à un stress environnemental, comme un manque de nutriments ou en réponse à des interactions biologiques avec d'autres organismes. Les métabolites secondaires sont produits par des groupes de gènes spécifiques appelés clusters de gènes de biosynthèse⁷⁶. Comme leur nom l'indique, ces clusters de gènes de biosynthèse sont constitués de plusieurs gènes localisés les uns à côté des autres dans le génome bactérien, et ensemble ils codent pour des enzymes spécifiques utilisés dans les différentes voies de biosynthèse pour produire des métabolites secondaires.

Il existe différentes classes de métabolites secondaires d'origine bactérienne, produites par différentes voies de biosynthèse. Parmi les métabolites secondaires les plus étudiées notamment pour leurs activités biologiques et leurs applications en médecine, on compte les peptides non ribosomiques et les polycétides. Les métabolites secondaires ont un grand nombre de fonctions biologiques connues mais de nombreuses restent à découvrir.

B. Fonction des métabolites secondaires au sein des communautés

Les métabolites secondaires ont d'importantes fonctions pour le maintien des populations bactériennes au sein des communautés. En effet, les populations bactériennes vivent dans des environnements compétitifs, et la multiplication de défenses chimiques et biochimiques est une stratégie robuste pour s'établir durablement dans ces environnements.

Vivant en communautés, une des fonctions importantes pour les bactéries est la communication. Les bactéries peuvent communiquer en utilisant des métabolites secondaires appelées auto-inducteurs⁷⁷. Lorsque la densité de la population bactérienne est faible, peu d'auto-inducteurs sont produits. En revanche, plus la population bactérienne grandit, plus la concentration en auto-inducteurs augmente dans l'environnement. Quand cette concentration d'auto-inducteurs atteint un certain seuil, les bactéries alentours deviennent capables de la détecter : ce mécanisme est appelé la détection du quorum. La détection du quorum ainsi que la production d'auto-inducteurs ont un effet sur la régulation des gènes bactériens, et ils permettent aux bactéries de se synchroniser pour effectuer ensemble plusieurs activités métaboliques, comme la création de biofilms, l'augmentation de la virulence, ou encore un échange de gène via conjugaison.

De plus, pour s'établir et se maintenir dans une communauté, les bactéries doivent faire preuve de compétitivité les unes envers les autres. Cette compétitivité implique plusieurs facettes. Tout d'abord, dans les systèmes symbiotiques, les bactéries doivent être capable de coloniser leur hôte. Les mécanismes de colonisation d'un hôte par les bactéries ne sont pas encore bien compris, mais la potentielle capacité des arylpolyènes à protéger les bactéries de stress oxydant exogène pourrait faciliter la colonisation de l'hôte⁷⁸. Une seconde facette importante de la compétitivité bactérienne est la capacité à capter les nutriments de l'environnement. Parmi ces nutriments, le fer est essentiel pour la croissance des bactéries. Les bactéries sont capables de capter le fer dans leur environnement en utilisant des sidérophores⁷⁹ qui sont des molécules chélatrices du fer. Ainsi, une bactérie capable de produire un grand nombre de sidérophores pourrait capter plus de fer de son environnement au détriment d'autres espèces bactériennes et être ainsi plus compétitive. Enfin, comme expliqué dans la partie précédente, certaines métabolites secondaires produites par les bactéries ont des activités antimicrobiennes. C'est par exemple le cas de beaucoup de peptides non ribosomiques et de polycétides. Ces molécules antimicrobiennes permettent aux bactéries de lutter contre divers microbes, et notamment d'autres bactéries. Dans un contexte de forte compétitivité bactérienne dans

une même communauté, les bactéries produisant ces molécules antimicrobiennes pourraient se montrer plus aptes à se maintenir dans la communauté.

C. Intérêts pharmaceutiques

Grâce à leurs nombreuses activités antimicrobiennes, les métabolites secondaires produites par les bactéries ont également été très utiles dans le domaine pharmaceutique. En effet, la majorité des antibiotiques utilisés de nos jours sont basés sur une molécule antimicrobienne. Parmi les bactéries produisant des molécules antimicrobiennes, les plus étudiées sont les actinobactéries, notamment car elles peuvent facilement isolées et cultivées en laboratoire⁸⁰. Par exemple, pendant l'âge d'or de la découverte des antibiotiques dans les années 1970, 70 à 80% de tous les antibiotiques découverts provenaient de la bactérie *Streptomyces*, qui appartient aux actinobactéries. Cependant, les actinobactéries ne sont pas les seuls à produire des molécules antimicrobiennes. Par exemple, certaines bactéries de l'ordre Bacillales⁸¹ et certaines *Pseudomonas*⁸² sont étudiées pour leur capacité à produire des molécules antimicrobiennes. La Figure 4 montre quelques exemples de molécules antimicrobiennes produites par des bactéries.

Cependant, la surutilisation des antibiotiques en médecine et en agriculture a conduit à l'émergence de résistance bactérienne⁸³. Certaines bactéries sont maintenant résistantes aux antibiotiques utilisés couramment, et rendent leur élimination très difficile. Par exemple, certaines souches de *Staphylococcus aureus* et *Enterococcus faecium* sont devenues multirésistantes et elles sont couramment trouvées en milieu hospitalier, et certaines personnes hospitalisées pour d'autres raisons se retrouvent infectées par ces bactéries⁸⁴.

Pour lutter contre les bactéries multirésistantes, de nouvelles méthodes pour identifier de production d'antibiotiques doivent être développées. Le séquençage de gènes ciblés et l'étude des clusters de gènes de biosynthèse semblent être des méthodes prometteuses. En effet, les techniques de bio-informatiques actuelles nous permettent d'analyser rapidement un grand nombre de génomes

bactériens et de détecter spécifiquement gènes de biosynthèse ciblés qu'ils contiennent, sans avoir besoin de cultiver et d'isoler les bactéries. Ces approches permettent de détecter de nouvelles séquences de gène de biosynthèse qui sont ensuite introduit dans le génome de bactéries par des approches de biotechnologie pour produire les métabolites d'intérêt. Par exemple, en ciblant le domaine d'adénylation des peptides non ribosomiques, Hover *et al.*⁸⁵ ont découvert une nouvelle classe d'antibiotiques.

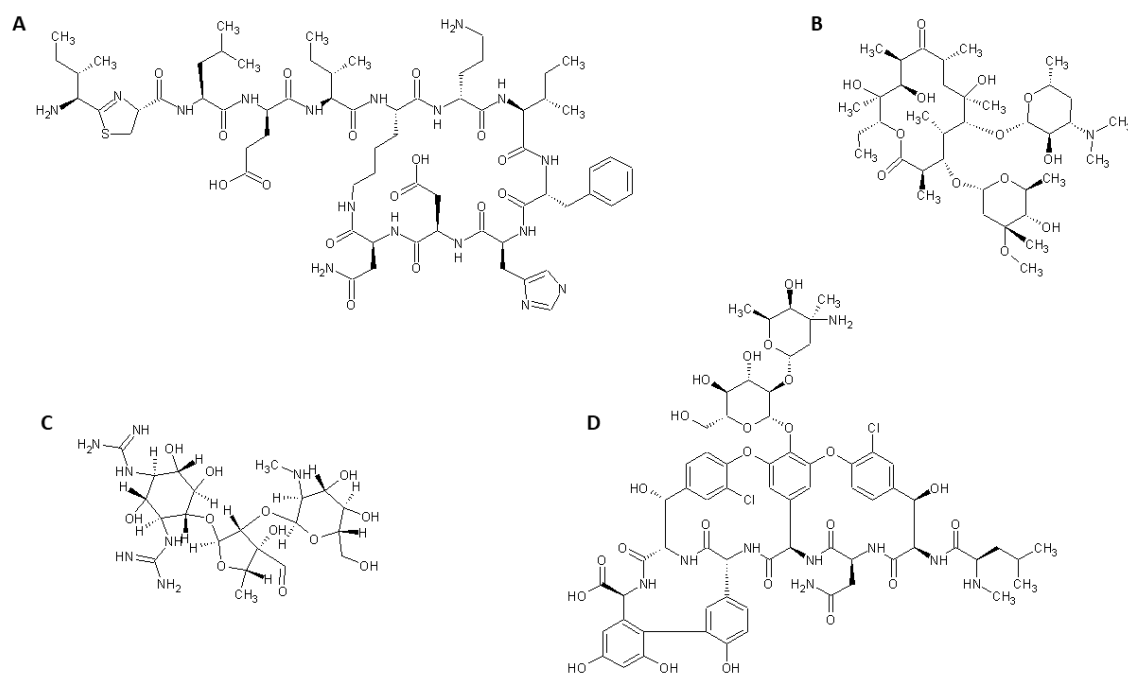


Figure 4. Exemples de molécules antimicrobiennes produites par des bactéries. (A) Bacitracine A produite par *Bacillus subtilis*. (B) Erythromycine produite par *Saccharopolyspora erythraea*. (C) Streptomycine produite par *Streptomyces griseus*. (D) Vancomycine produite par *Amycolatopsis orientalis*. Les structures de ces molécules ont été obtenues sur la base de données ChemSpider (www.chemspider.com) et redessinées à partir du logiciel ChemSketch.

Etat de l'art général des méthodes

I. Séquençage de l'ADN

A. Préparation des échantillons

1. Extraction de l'ADN

Il existe une multitude de méthodes et de kit d'extraction d'ADN. Le choix de la méthode doit se faire selon plusieurs facteurs, tels que le type d'échantillon, le matériel disponible, le coût, le temps et le risque de contamination.

Dans le cadre des chapitres 1 et 2 de cette thèse, nous désirions extraire l'ADN bactérien et eucaryote associé à plusieurs échantillons de fourmis ou de leur nid. Il était donc nécessaire d'utiliser une méthode qui permettent de bien lyser toutes cellules de bactéries d'eucaryotes et de fourmis contenues dans les échantillons pour en extraire l'ADN.

Plusieurs comparaisons de méthodologies ont été publiées pour extraire l'ADN associé aux insectes. En particulier, nous nous sommes intéressés à un article qui présente les différentes méthodes d'extraction d'ADN chez les insectes⁸⁶, ainsi qu'à deux articles qui comparent la diversité bactérienne associée aux fourmis obtenue selon la méthode d'extraction d'ADN utilisée^{87,88}. Suivant les recommandations de ces articles, nous avons décidé d'utiliser le kit DNeasy Powersoil (Quiagen, USA) pour extraire l'ADN associé aux échantillons de fourmis et environnementaux collectés.

2. Réaction de polymérase en chaîne

La réaction de polymérase en chaîne (PCR) est à la base des méthodes de séquençage actuelles. Le but est d'amplifier les échantillons d'ADN extraits. Les réactifs nécessaires à la réaction sont l'ADN extraits, les amorces, l'ADN polymérase, le buffer et les déoxynucléotides triphosphates (dNTPs). Le principe de base est composé de trois étapes. La première étape est la dénaturation de l'ADN, dans laquelle l'ADN double brin est chauffé à forte température pour séparer les deux brins. La deuxième

étape est la ligation, dans laquelle l'amorce choisie se combine à l'extrémité de l'ADN simple brin par complémentarité. Cette amorce est propre à l'ADN de l'organisme cible. Pour les bactéries il s'agit d'une amorce de XXX a YYY paires de bases (selon le type de séquençage) correspondant à la sous unité 16S de l'ARN ribosomique. La dernière étape est l'extension, dans laquelle l'ADN polymérase étend l'extrémité 3' de l'amorce le long de l'ADN simple brin par complémentarité pour reformer un ADN double brin. Ces trois étapes sont ensuite répétées sur un certain nombre de cycles pour amplifier l'ADN extrait. Après amplification, il est possible de purifier l'ADN amplifié par différentes méthodes, les hybrides, les amorces et nucléotides en surplus, et de ne conserver que les séquences ADN amplifiées.

Le principe de base est le même pour toutes les méthodes de séquençage, cependant la composition de la réaction peut varier en fonction de la méthode de séquençage désirée. De ce fait, beaucoup de plateformes de séquençage effectuent eux-mêmes la PCR ainsi que le séquençage. Il existe de nombreuses méthodes de séquençage différentes. Dans la partie suivante, nous présenterons quelques-unes de ces méthodes parmi les plus utilisées.

B. Techniques de séquençage

1. Le séquençage Sanger

La méthode de Sanger développée dans les années 1970, aussi appelée méthode de terminaison de chaîne ou méthode des didéoxynucléotides triphosphates (ddNTPs), est la méthode de base de séquençage d'ADN utilisée dans les techniques de séquençage actuelles⁸⁹. Dans cette méthode, la PCR réalisée en amont est appelée PCR de terminaison de chaîne. La différence majeure avec la PCR classique est l'ajout de ddNTPs en quantité moindres que la quantité de dNTPs ajoutés. Les ddNTPs sont des dNTPs modifiés pour leur retirer le groupe -OH de l'extrémité 3'. Ce groupe -OH permet la liaison entre deux dNTPs pendant la phase d'extension de la PCR classique. Ainsi, quand l'ADN

polymérase ajoute un aléatoirement un ddNTP au lieu d'un dNTP, la réaction s'arrête, ce qui crée de multiples brins d'ADN de différentes tailles.

Dans le cas du séquençage Sanger manuel, quatre réactions de PCR sont préparées, chacune contenant un type de ddNTP (ddATP, ddCTP, ddGTP ou ddTTP) différent. Le produit de réaction de chaque PCR est ensuite placé dans un puits différent et séparé par gel d'électrophorèse. Dans le gel, les fragments d'ADN les plus petits migreront plus rapidement que les fragments plus grands, et se retrouveront en bas du gel alors que les plus grands se retrouveront en haut. Ainsi, pour déterminer la séquence d'ADN il suffit de lire les quatre puits du gel de bas en haut pour déterminer l'ordre des nucléotides dans la séquence ADN.

Dans le cas du séquençage Sanger automatisé, une seule réaction de PCR contenant les 4 types de ddNTPs est préparée, mais un fluorochrome de couleur différente est ajouté à chaque type de ddNTP (Figure 5A). Le produit de réaction PCR est ensuite placé dans un tube capillaire rempli de gel d'électrophorèse, éclairé par un laser et relié à un ordinateur. Quand le laser éclaire une bande, le fluorochrome excité formera sur l'ordinateur un pic d'une couleur spécifique en fonction du type de ddNTP. Ainsi, un chromatogramme sera généré montrant les pics colorés ainsi que la séquence ADN correspondante.

Le développement de méthodes de séquençage à haut débit qui seront présentées ensuite a considérablement accéléré la vitesse de séquençage et la réduction des coûts. Cependant, la méthode de Sanger est toujours considérée comme un standard pour le séquençage d'ADN, y compris pour valider des séquences d'ADN d'intérêt séquencées au préalable par les méthodes de séquençage à haut débit, car elle possède une précision de 99.99% et un coût très bas pour séquencer des gènes d'intérêt.

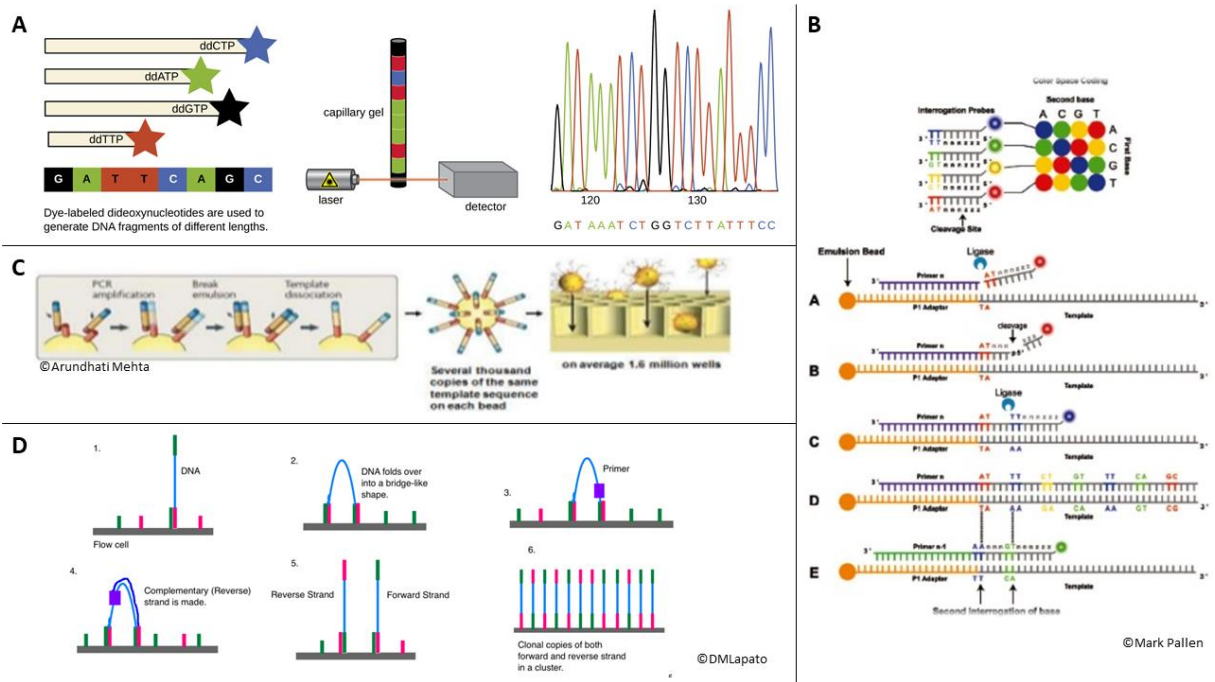


Figure 5. Principe des différents types de séquençage. (A) Dans le séquençage Sanger, des ddNTPS associés à des fluorochromes sont utilisés pour générer les fragments d'ADN, qui sont ensuite séparés par électrophorèse pour que la séquence d'ADN puisse être décodée par ordinateur⁹⁰. (B) Dans le séquençage SOLiD, les fragments d'ADN se fixe à des billes par complémentarité, qui sont mises en émulsion PCR. Les billes contenant les fragments d'ADN amplifiés sont déposées sur une plaque de verre contenant des amorces. Ces amorces sont liguées par complémentarité aux deux premières bases d'un bloc de huit bases associées à un fluorochrome. Ce cycle se reproduit 5 fois avec des amorces de plus en plus petites, puis la séquence d'ADN est identifiée par recoupement. (C) Dans le pyroséquençage, les fragments d'ADN se fixe à des billes par complémentarité, qui sont mises en émulsion PCR. Les billes contenant les fragments d'ADN amplifiés sont déposées dans des puits auxquels sont ajoutés séquentiellement des dNTPs, et la séquence d'ADN produite est décodée par pyrogramme. (D) Dans le séquençage Illumina, les brins d'AND s'attachent par complémentarité sur les amorces d'une microplaque, puis un brin se courbe et s'attache à un second brin pour former un pont. Ensuite, la polymérase synthétise le brin complémentaire et les deux brins se détachent, puis reforment des ponts avec de nouveaux brins et le principe recommence.

2. Le séquençage « shotgun »

La méthode de Sanger ne permet d'identifier que des morceaux d'ADN mesurant entre 100 et 1000 paires de bases. Ainsi, pour séquencer des morceaux d'ADN plus grands, il faut auparavant les couper en fragments plus petits et les séquencés séparément. La stratégie « shotgun » consiste à couper aléatoirement plusieurs copies de longs morceaux d'ADN en petits morceaux⁹¹. Ces morceaux d'ADN sont ensuite séquencés en utilisant la méthode de Sanger, et le résultat est appelé une lecture (read en anglais). On obtient ainsi de nombreuses lectures dont certaines contiennent des extrémités qui se chevauchent. En analysant et combinant par ordinateur les extrémités des lectures qui se chevauchent, on peut réassembler la longue séquence ADN initiale. Ces longues séquences d'ADN reformées par chevauchement de séquences sont appelées des « contigs ».

A l'origine, cette stratégie n'analysait qu'une extrémité de chaque fragment d'ADN et a été la première utilisée pour séquencer des génomes entiers. Cependant, le besoin de séquencer des génomes de plus en plus grands et complexes a entraîné la modification de cette stratégie pour pouvoir analyser les deux extrémités de chaque fragment d'ADN. Cette nouvelle stratégie s'appelle le séquençage des extrémités appariées. Cette stratégie est plus compliquée à analyser, mais fournit beaucoup plus d'informations que la stratégie shotgun originelle.

De plus le séquençage shotgun était auparavant basé sur la méthode de Sanger. Cependant, l'arrivée du séquençage à haut débit a permis de produire un grand nombre de lectures en un temps très court. Ainsi, en utilisant la stratégie shotgun sur des lectures provenant de séquençage à haut débit au lieu de lectures provenant de la méthode de Sanger, on obtient une couverture beaucoup plus grande. C'est cette stratégie qui est généralement utilisée pour séquencer des génome entiers et complexes, comme le génome humain⁹².

3. Le séquençage à haut-débit

Il existe deux types de séquençage à haut-débit : le séquençage de deuxième génération permettant uniquement le séquençage de fragments courts et datant du début des années 2000, et le séquençage de troisième génération datant du début des années 2010. Le séquençage de troisième génération a la particularité de pouvoir séquencer des fragments d'ADN beaucoup plus longs que le séquençage de deuxième génération : entre 6000 et 8000 paires de bases pour le séquençage de troisième génération, contre 200 à 300 paires de bases pour le séquençage de seconde génération. Cependant le séquençage de troisième génération est toujours une technologie en plein développement. Ici, nous ne présenterons que le séquençage de deuxième génération car c'est celui que nous avons utilisé dans cette thèse.

Le séquençage de deuxième génération regroupe plusieurs technologies. Ici nous présenterons trois des technologies les plus utilisées : le séquençage SOLiD⁹³, le pyroséquençage 454⁹⁴ et le séquençage Illumina⁹⁵. Ces trois techniques suivent le même principe général avec quelques différences pour l'amplification et le séquençage. La méthode se déroule en trois étapes.

La première étape est la préparation de la librairie, dans laquelle les fragments d'ADN sont ligués à chacune de leurs extrémités par des adaptateurs spécifiques qui contiennent les amorces et codes-barres nécessaires à l'amplification et au séquençage des fragments.

La deuxième étape est l'amplification, dans laquelle le produit de ligation est déposé sur des billes magnétiques dans le cas du séquençage SOLiD et du pyroséquençage 454, ou sur un support solide dans le cas du séquençage Illumina. Dans les trois cas, l'amplification se fait en parallèle. Dans le cas du séquençage SOLiD et du pyroséquençage 454, chaque fragment d'ADN va se fixer à une bille par complémentarité, puis les billes sont mises en émulsion avec les composés nécessaires à la PCR. Ainsi, après plusieurs cycles d'amplification, chaque bille contiendra des fragments d'ADN amplifiés (Figure 5B et C). Dans le cas du séquençage Illumina, chaque fragment d'ADN va se fixer par complémentarité sur une microplaque contenant les mêmes amorces. L'amplification et la génération de « clusters » se fait en pont. Les fragments d'ADN vont se courber et s'attacher au support solide grâce à leur adaptateur,

et l'ADN polymérase va créer le brin complémentaire. Les deux brins vont alors se dénaturer et le processus recommence sur chaque brin. Ce processus se fait sur des milliers de cycles (Figure 5D).

La dernière étape est le séquençage qui se fait également en parallèle sur tout le support. Dans le cas du séquençage SOLiD, les billes contenant les fragments d'ADN amplifiés sont déposées sur une plaque de verre et une amorce est fixée aux adaptateurs des fragments d'ADN (Figure 5B). Ensuite, des blocs de huit bases sont ligés aux amorces. Dans ces blocs de huit bases, les deux premières bases vont se fixer par complémentarité aux amorces correspondantes, et la base huit est fixée à un fluorochrome. A la fin du cycle, le produit est éliminé par une étape de clivage du fluorochrome, permettant de détecter le signal lumineux, et une étape de rinçage, puis un deuxième cycle commence en utilisant une amorce plus petite que dans le cycle précédent avec une base en moins. Ce système utilise donc une succession de cinq cycles avec des amorces de taille différentes (d'une à quatre bases de moins que la première). Seulement quatre types de fluorochromes différents peuvent être utilisés, donc la détection du premier signal lumineux ne donne qu'une information partielle sur la nature de la paire de base détectée. Il est donc nécessaire de faire des suites de cinq cycles avec des amorces de plus en plus petites. Cela permet de tester chaque base dans deux cycles différents en utilisant deux amorces différentes. A chaque fin de suite de cinq cycles, par recoupement il est possible d'identifier les bases de la séquence ADN originelle. Dans le cas du pyroséquençage 454, les billes contenant les fragments d'ADN amplifiés sont déposées dans des puits (Figure 5C). A chaque cycle, les dNTPs sont introduits séquentiellement dans un ordre précis. L'ajout d'un ou plusieurs nucléotides entraîne une émission de lumière dont la force est proportionnelle au nombre de nucléotides ajoutés, et les résultats sont présentés sous forme d'un pyrogramme. Dans le cas du séquençage Illumina, un fluorochrome de couleur différente est ajouté à chaque types de dNTP (Figure 5D). Ce système fonctionne comme l'utilisation de ddNTPs dans la méthode de Sanger, car le fluorochrome agit comme un blocage et empêche d'autres dNTPs de se fixer ensuite. Cependant, ce blocage est réversible par clivage. A chaque cycle, un dNTP spécifique est donc ajouté à chaque fragment d'ADN, puis les dNTPs non fixés sont éliminés par une étape de rinçage. Après chaque cycle, la fluorescence permet à

l'ordinateur de détecter quel dNTP a été ajouté à chaque fragment d'ADN sur la plaque. Ensuite, l'extrémité bloquante des dNTPs fixés est supprimée par clivage et un nouveau cycle de séquençage commence.

II. Techniques d'identification moléculaires post-séquençage

A. Le métabarcoding

Le métabarcoding est une méthode moléculaire pour identifier les microorganismes présents dans un échantillon en utilisant un gène contenant un marqueur taxonomique spécifique (Figure 6). Un marqueur taxonomique est un gène très conservé dans un groupe taxonomique défini, et permettant d'identifier précisément un organisme choisit dans ce groupe taxonomique. Il existe de nombreux marqueurs taxonomiques communément utilisés dans le métabarcoding, comme l'ARN ribosomique 18S⁹⁶ (18S rRNA, présent dans le génome de tous les organismes eucaryotes), l'ARN ribosomique 16S⁹⁷ (16S rRNA, présent dans le génome de tous les organismes procaryotes), le cytochrome oxydase I⁹⁸ (COI, présent dans le génome de tous les animaux), ou l'espaceur interne transcrit⁹⁹ (ITS, présent dans le génome des champignons et des cyanobactéries).

Les marqueurs utilisés pour le métabarcoding doivent avoir une forte résolution taxonomique, afin de détecter et distinguer la majorité des espèces provenant du groupe taxonomique étudié. Pour cela, les marqueurs doivent être petits (pour pouvoir les utiliser dans les techniques de séquençage haut-débit), ainsi qu'avoir une faible variabilité intraspécifique (afin que les organismes venant d'une même espèce soit classés ensemble), une forte variabilité interspécifique (pour pouvoir différencier les organismes de différentes espèces) et des extrémités conservées (afin de pouvoir développer des amorces universelles pour étudier l'ensemble du groupe taxonomique d'intérêt)¹⁰⁰.

La méthodologie se fait en trois étapes. La première étape est l'extraction d'ADN d'un échantillon. La deuxième étape est l'amplification par PCR de l'ADN extrait en utilisant une amorce faite à partir du

marqueur taxonomique choisi pour amplifier ce marqueur dans l'échantillon. Le résultat de la PCR contiendra toutes les séquences du gène marqueur des différents microorganismes présents dans l'échantillon initial. La dernière étape est le séquençage à haut-débit des séquences amplifiées. De plus, en ajoutant au marqueur taxonomique des combinaisons de code-barres spécifiques, il est possible de réaliser le métabarcoding sur plusieurs échantillons en parallèle et de réattribuer chaque séquence à l'échantillon d'origine après séquençage.

Dans le cadre de cette thèse, nous avons utilisé cette méthode en ciblant le gènes 16S rRNA, qui est un marqueur spécifique des bactéries¹⁰¹ (chapitre 1), ainsi que le gène 18S rRNA, qui est un marqueur spécifique des eucaryotes¹⁰² (chapitre 2). Les amplifications et séquençage des gènes cibles ont été effectuées par Argonne National Laboratory (Lemont, Illinois, USA), en utilisant la technique de séquençage haut-débit Illumina présentée dans la partie précédente.

B. Approche du multi-marqueur métabarcoding

Certains groupes taxonomiques peuvent être identifiés par plusieurs marqueurs taxonomiques, avec différents degrés de résolution en fonction des marqueurs taxonomiques utilisés. Par exemple, le gène 18S rRNA est un marqueur des eucaryotes. Les animaux peuvent donc être identifiés en utilisant ce gène. Mais les animaux peuvent également être identifiés en utilisant le gène mitochondrial COI, avec une meilleure résolution⁹⁸. De même, les champignons peuvent être identifiés en utilisant le gène 18S rRNA, mais également en utilisant le gène ITS avec une meilleure résolution⁹⁹.

L'approche du multi-marqueur métabarcoding est une variante de la méthode métabarcoding présentée auparavant. Cette approche consiste à utiliser différents marqueurs du même groupe taxonomique, afin d'augmenter la résolution et la possibilité d'obtenir une correspondance taxonomique précise. En effet, pour chaque marqueur taxonomique on utilise une base de données différente pour identifier la taxonomie des organismes. Ces bases de données se remplissent au fur et à mesure par les recherches et les différentes espèces identifiées. Ainsi, plus un marqueur

taxonomique sera utilisé dans les recherches, plus la base de données correspondante s'étoffera, et de plus en plus d'espèces pourront être identifiées par ce marqueur taxonomique dans les recherches suivantes.

Le succès de cette approche a déjà été démontré dans l'étude des eucaryotes associés à des échantillons de sol^{103,104}. Dans ces études, l'utilisation de marqueurs multiples plutôt que l'utilisation du gène eucaryote 18S rRNA a permis d'augmenter de moitié la diversité globale des eucaryotes identifiés.

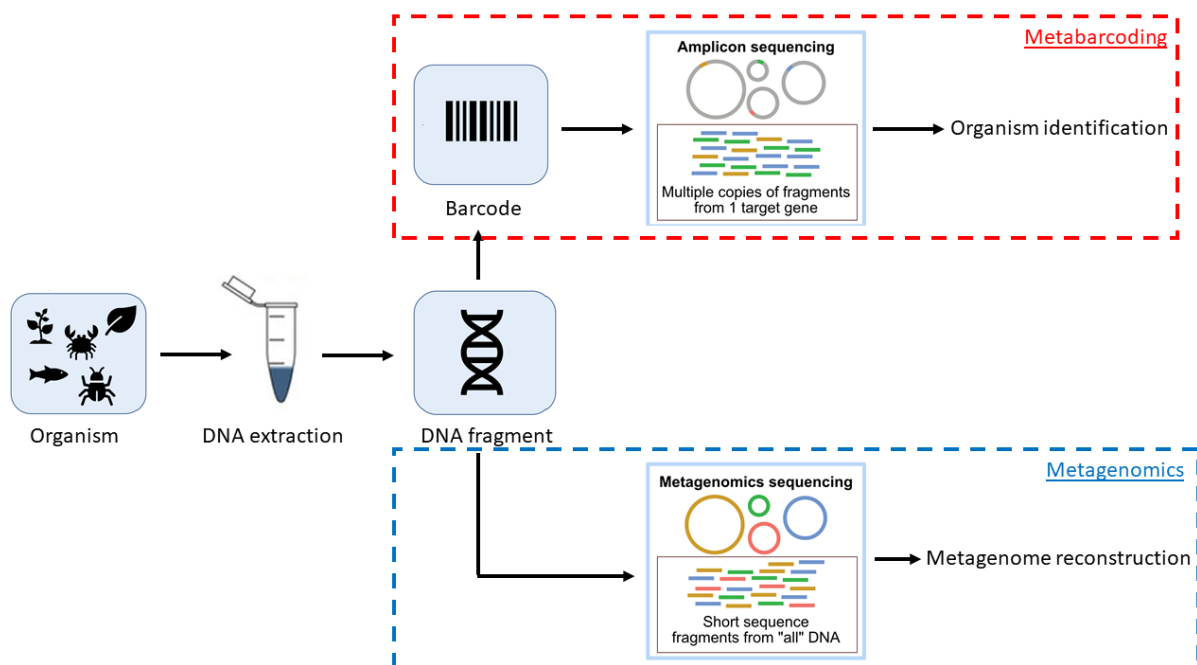


Figure 6. Principes du métabarcoding et de la métagénomique. Le rôle du métabarcoding est d'utiliser un gène marqueur spécifique pour identifier la taxonomie de différents organismes composant un échantillon grâce à leur ADN. Le rôle de la métagénomique est de séquencer le génome complet de tous les organismes composant un échantillon.

C. La génomique structurale

La génomique structurale est l'analyse de la structure d'un génome, c'est-à-dire l'identification des séquences géniques composant ce génome. Cette identification est possible en séquençant et réassemblant complètement le génome d'intérêt. Le développement des nombreuses techniques de séquençage shotgun et à haut-débit a grandement réduit les coûts et permet le séquençage de génomes de plus en plus longs et de plus en plus complexes. En effet, avant l'avènement des techniques de séquençage à haut-débit, le séquençage de génomes complexes était très coûteux en temps et en argent. Par exemple, en utilisant la technique de séquençage Sanger, le Human Genome Project (<https://www.genome.gov/human-genome-project>) a duré plus de 10 ans et coûté plus de 2 billions de dollars pour publier le séquençage complet du premier génome humain en 2003. Aujourd'hui, en utilisant les techniques de séquençage à haut-débit, il est possible de séquencer un génome humain en quelques jours pour moins de 1000 dollars.

La génomique structurale est très largement utilisée dans le domaine de la recherche fondamentale, mais a également beaucoup d'intérêt dans le domaine de la recherche clinique. Par exemple, cette méthode a été utilisée pour contrôler le développement d'épidémie provoquée par des pathogènes, comme le virus Zika¹⁰⁵. Cette méthode est également couramment utilisée dans le domaine du contrôle alimentaire, pour vérifier la possible présence de bactéries pathogènes dans la nourriture ou l'eau potable. Enfin, cette méthode permet de réaliser différents diagnostics médicaux pour la recherche de mutations ou de variants rares¹⁰⁶.

D. La métagénomique

Suivant l'avènement du séquençage de génomes complets, le séquençage de microbiome s'est largement développé. La métagénomique est l'étude de l'ensemble du matériel génétique d'un échantillon. Il est désormais possible de séquencer un échantillon, qu'il soit environnemental ou un organisme, et d'obtenir l'ensemble des génomes (prokaryotes et eucaryotes) présents dans cet

échantillon (Figure 6). En utilisant la métagénomique, plusieurs projets internationaux de grande envergure dont le but est d'identifier les différents microorganismes présents dans différents échantillons ont été lancés, comme le Global Ocean Sampling Expedition démarré en 2004 (<https://www.jcvi.org/research/gos>), le Human Microbiome Project démarré en 2008 (<https://hmpdacc.org>) et le Earth Microbiome Project démarré en 2010 (<http://earthmicrobiome.org>).

Un des nombreux avantages de cette méthode est qu'en plus de pouvoir identifier tous les microorganismes présents dans un échantillon, les études métagénomiques donnent accès à l'ensemble du matériel génomique présent dans cet échantillon, et il est donc possible d'étudier un gène ou une voie métabolique des différents microorganismes présents dans l'échantillon. Dans une récente étude, Hu *et al.* ont utilisé cette méthode pour étudier les microorganismes symbiotiques des intestins de différentes espèces de fourmis *Cephalotes*, et ont mis en évidence que les bactéries symbiotiques majeures de l'intestin chez les fourmis *Cephalotes* possèdent des gènes encodant les voies métaboliques responsables du recyclage de l'azote provenant de l'hôte en acides aminés essentiels¹⁰⁷. Dans le cadre du chapitre 3 de cette thèse, nous avons réutilisé les données métagénomiques produits par Hu *et al.* dans leur étude, afin d'identifier les différents clusters de gènes de biosynthèse présents chez les bactéries symbiotiques de l'intestin des fourmis *Cephalotes*.

III. Techniques bio-informatiques d'analyse de données de séquençage

A. Analyse de données de métabarcoding

1. Démultiplexage

Il existe de nombreux logiciels d'analyse de données de métabarcoding. Ainsi, pour faciliter les analyses de données de métabarcoding, la plateforme Qiime a été créée en 2010¹⁰⁸, suivie de Qiime2 en 2018¹⁰⁹. La plateforme Qiime2 permet d'analyser des communautés microbiennes qui ont été obtenues par l'utilisation de gènes marqueurs, comme les gènes 16S rRNA et 18S rRNA. Elle regroupe

donc les logiciels et outils les plus utilisés dans l'analyse des communautés microbiennes. Dans le cadre des chapitres 1 et 2 de cette thèse, nous avons utilisé la plateforme Qiime2 pour étudier les communautés microbiennes associées à une trentaine d'espèces différentes de fourmis néotropicales. Pour cette raison, seules les méthodes et outils communément utilisés sur cette plateforme seront présentés dans cette partie.

Dans le cas où plusieurs échantillons ont été séquencés en même temps, des code-barres ont été utilisés pendant le séquençage haut-débit pour identifier les différents échantillons après séquençage. Dans ce cas, la première étape nécessaire est le démultiplexage des séquences (Figure 7). Les fichiers de sortie typiques de séquençage à haut-débit de plusieurs échantillons sont généralement au nombre de trois : un fichier contenant toutes les lectures avant, un fichier contenant toutes les lectures arrière, et un fichier contenant la liste des code-barres, où chaque code-barre correspond à un échantillon différent. Le démultiplexage consiste à utiliser les code-barres présents dans les séquences pour créer des fichiers contenant uniquement les séquences avant et arrière liées à un code-barre spécifique. Ainsi, chaque fichier créé contiendra toutes les lectures avant et arrière concernant un échantillon différent. Ce démultiplexage peut parfois être effectué par la plateforme de séquençage elle-même.

Dans le cas où le démultiplexage n'est pas effectué directement par la plateforme de séquençage, sur Qiime2 le démultiplexage se fait après importation des fichiers par les commandes « demux emp-single »¹¹⁰ ou « demux emp-paired »¹¹¹, selon si les extrémités sont appariées ou non. Les commandes « cutadapt demux-single »¹¹² et « cutadapt demux-paired »¹¹² sont utilisées à la place, si les code-barres utilisés pour le démultiplexage sont présents dans les fichiers de lectures provenant du séquençage à haut-débit et non dans un fichier séparé.

2. Contrôle de qualité des séquences

Après démultiplexage, la seconde étape est le contrôle de la qualité des séquences. Sur Qiime2, il existe 2 méthodes pour cela : DADA2¹¹³ et Deblur¹¹⁴ (Figure 7). Les 2 méthodes ont leurs avantages et

leurs inconvénients, et l'utilisation de l'une ou de l'autre est dépendant des données à analyser et des objectifs de l'étude.

Les objectifs principaux des deux méthodes sont les mêmes : supprimer les séquences bruit, supprimer les séquences chimériques, supprimer les singletons, réduire les extrémités des séquences, supprimer les séquences trop courtes et de-répliquer les séquences. Cependant les deux méthodes ont aussi des différences importantes. Une de ces différences est que les séquences appariées sont jointes par DADA2, alors qu'elles ne le sont pas par Deblur. Dans ce cas, il est donc nécessaire de joindre les séquences appariées (en utilisant VSEARCH¹¹⁵) et de filtrer les séquences de faible qualité (en utilisant le q-score¹¹⁶) avant d'utiliser la méthode Deblur. Une autre différence importante entre ces deux méthodes est que les séquences marginales sont corrigées par DADA2 alors qu'elles sont supprimées par Deblur.

Quel que soit la méthode utilisée, les séquences obtenues après cette étape de filtration sont appelées des ASVs, signifiant Amplicon Sequence Variants. Auparavant, des techniques de clustérisations étaient utilisées en utilisant les séquences de gènes cibles afin de rassembler les organismes similaires. Les clusters ainsi formés sont appelés des OTUS, signifiant Operating Taxonomic Unit, et ils sont généralement inférés avec similarité d'identité de séquence de 97%. Cette approche entraîne le risque de rassembler dans le même OTU des espèces similaires mais différentes, pouvant causer la perte de l'identification de certaines espèces. Pour leur part, les ASVs sont très utiles dans les études de communautés microbiennes, parce qu'elles permettent d'inférer les séquences biologiques présentes dans chaque échantillon collecté. De plus, il est possible de distinguer des ASVs qui diffèrent par seulement un nucléotide. Cela permet de pouvoir comparer les ASVs entre différentes études, ce qui n'est pas possible avec les OTUs.

3. Assignement taxonomique des séquences

Pour obtenir l'assignement taxonomique des séquences, on utilise des bases de données de référence. Il existe de nombreuses bases de données différentes, et le choix d'une base de données

dépend des données que l'on a, de la question biologique que l'on se pose et des préférences de chacun. Quelques-unes des bases de données les plus utilisées dans l'étude de communautés microbiennes sont : SILVA¹¹⁷, Greengenes¹¹⁸ et NCBI-RefSeq¹¹⁹. En comparant les ASVs obtenues à une base de données de référence, il est possible d'identifier la taxonomie des organismes présents dans une communauté microbienne étudiée. Pour cela, on utilise un classificateur taxonomique.

QIIME2 propose trois méthodes de classification taxonomique : « classify-consensus-blast »¹²⁰, « classify-consensus-vsearch »¹¹⁵ and « classify-sklearn »¹²¹ (Figure 7). Les 2 premiers classificateurs utilisent des méthodes d'alignements entre les ASVs et la base de données de référence, tandis que le dernier utilise une méthode d'entraînement à partir des gènes marqueurs présents dans les ASVs pour les comparer à la base de données de référence. Cette dernière méthode est souvent plus longue et requiert des capacités et une puissance informatique plus grande que les méthodes d'alignement. Le choix de la méthode dépend donc du temps, du nombre d'ASVs à analyser, de la taille de la base de données de référence, de la puissance de l'ordinateur et des préférences de chacun. Les résultats obtenus diffèrent généralement peu en fonction de la méthode de classification utilisée.

Il est également possible d'évaluer les relations phylogénétiques entre les différents ASVs obtenus (Figure 7). Cela se fait par la commande « phylogeny »¹²² qui est basée sur des méthodes d'alignement de séquence en utilisant le gène marqueur des ASVs. La construction de cette phylogénie des ASVs est fondamentale pour certaines méthodes d'analyse de diversité alpha et beta qui seront présentées dans la partie suivante.

4. Filtration des ASVs

Avant de pouvoir utiliser les ASVs obtenus dans l'analyse des communautés microbiennes, il est nécessaire de les filtrer afin de retirer le plus possible de contaminants et de séquences non ciblées (Figure 7). L'étape de décontamination est la seule étape fondamentale qui n'est pas encore prise en charge par QIIME2, elle peut cependant être réalisée sur le logiciel R¹²³ à l'aide du package « decontam »¹²⁴. Ce package permet d'éliminer de nos ASVs toutes les séquences provenant de

sources contaminantes, comme les réactifs utilisés. Le package « decontam » propose deux méthodes de décontamination : la méthode de fréquence et la méthode de prévalence. La méthode de fréquence nécessite d'avoir auparavant quantifié l'ADN présent dans chaque échantillon. Elle utilise la quantité d'ADN présente dans chaque échantillon pour calculer la fréquence des différents ASVs dans chaque échantillon et ainsi identifier les contaminants. La méthode de prévalence nécessite d'avoir auparavant amplifié et séquencé des contrôles négatifs en même temps que les échantillons de communauté bactérienne testés. Elle utilise la présence/absence des ASVs dans les contrôles négatifs afin d'identifier les contaminants dans les échantillons. Il est possible d'utiliser l'une ou l'autre des méthodes en fonction des données à décontaminer, ou même de combiner les deux méthodes.

Après avoir retiré les contaminants, il est possible dans Qiime2 de filtrer les séquences non ciblées, comme les séquences de l'hôte, les mitochondries et les chloroplastes. Pour cela on utilise les commandes « taxa filter-table » et « taxa filter-seqs » pour supprimer les séquences non ciblées dans la table des ASVs et dans les séquences de référence respectivement. Si plusieurs échantillons sont analysés en même temps, il est possible que pour certains échantillons l'amplification ou le séquençage aient échoués, résultant dans un plus faible nombre d'ASVs que dans les autres échantillons. Afin que ce problème n'affecte pas les comparaisons de communautés bactériennes entre plusieurs échantillons, il est également possible de supprimer les échantillons dont le nombre d'ASVs présents est inférieur à un certain seuil. Pour cela, on utilise la commande « feature-table summarize » afin d'obtenir le nombre d'ASVs dans chaque échantillon, puis la commande « diversity alpha-rarefaction » afin de déterminer le seuil en dessous duquel le nombre d'ASVs pourrait être dû à un défaut d'amplification ou de séquençage. Enfin, la commande « feature-table filter-samples » est utilisée pour éliminer les échantillons concernés.

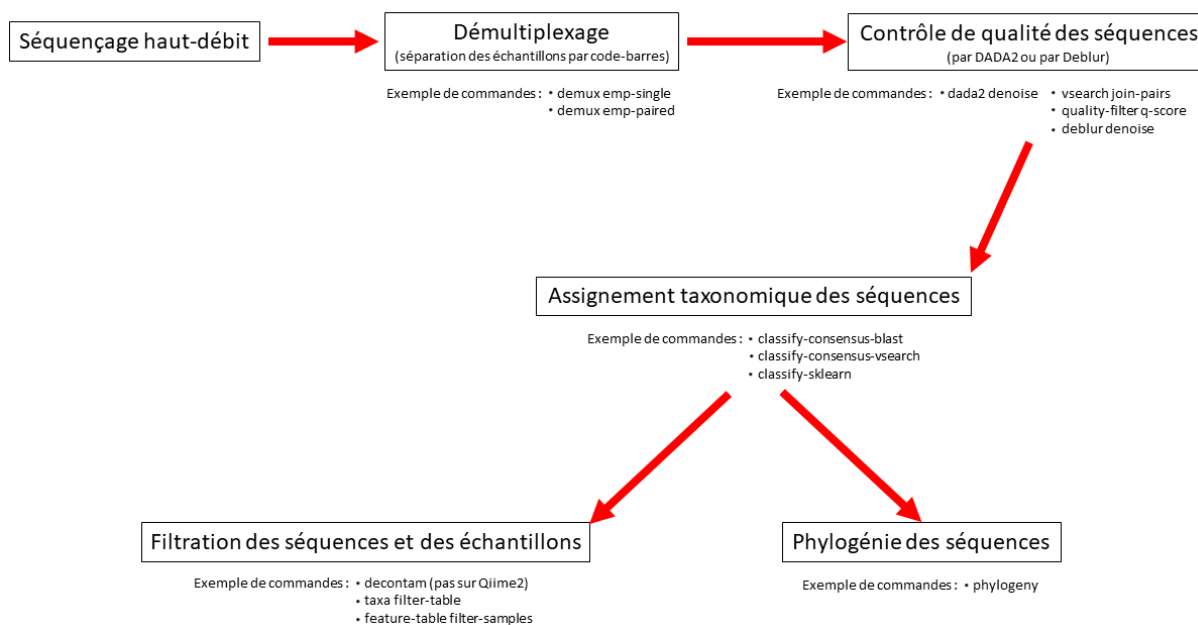


Figure 7. Résumé du processus de travail et exemples de commandes utilisées pour l’analyse de séquences de métabarcoding sur la plateforme Qiime2.

B. Analyse de données génomiques et métagénomiques

1. Contrôle de qualité des séquences

Du séquençage à l’annotation des séquences, l’analyse bio-informatique de données génomiques et métagénomiques suivent les mêmes étapes et peuvent utiliser les mêmes logiciels. La première étape d’analyse de données génomiques et métagénomiques est le démultiplexage, si plusieurs échantillons ont été séquencés en parallèle. Un exemple de logiciel pour effectuer le démultiplexage est la commande `fastx_demux` du logiciel `Usearch`¹²⁵.

Après démultiplexage, il est possible d’utiliser des logiciels comme `Trimmomatic`¹²⁶, pour réduire les extrémités des séquences, afin d’enlever les amorces, code-barres et adaptateurs utilisés (Figure 8). Ces logiciels peuvent également être utilisés pour réduire de quelques bases les extrémités des lectures afin d’améliorer leur qualité avant assemblage. Cependant, cette étape n’est pas forcément requise et doit être effectuée avec prudence, car réduire trop fortement les extrémités des lectures

pourrait diminuer les possibilités de chevauchements de séquences pour former les contigs, et cela pourrait conduire à une diminution de qualité de l'assemblage du génome/métagénome.

2. Assemblage du génome/métagénome

Les lectures obtenues par séquençage haut-débit sont généralement courtes de quelques centaines de paires de bases. Ainsi, pour pouvoir annoter le génome/métagénome il est nécessaire d'assembler les lectures en contigs par chevauchement des extrémités. Il existe deux types d'assemblage de génomes/métagénomes : le co-assemblage ou l'assemblage *de novo* (Figure 8). L'utilisation du type d'assemblage dépend du type de génome/métagénome à assembler. Typiquement, pour traiter des métagénomes provenant de séquençage de microbiome du même individu mais après un certain temps ou après différents traitements, il est préférable d'utiliser le co-assemblage. En revanche, pour traiter des métagénomes provenant de séquençage de microbiome de différents individus, il est préférable d'utiliser le séquençage *de novo*.

Le co-assemblage est le fait d'assembler plusieurs génomes/métagénomes à partir du même fichier, c'est-à-dire sans effectuer l'étape de démultiplexage des lectures. Ce type d'assemblage souvent permet d'obtenir une meilleure profondeur des lectures, et de faciliter la différenciation des génomes grâce à une meilleure couverture. Ce type d'assemblage peut aussi faciliter les comparaisons entre génomes/métagénomes. L'assemblage *de novo* consiste à effectuer l'assemblage des différents génomes/métagénomes séparément après démultiplexage des lectures. L'assemblage *de novo* est plus long et nécessite des performances informatiques beaucoup plus lourdes que le co-assemblage. Certains logiciels peuvent être utilisés pour effectuer les deux types d'assemblage, comme SPAdes¹²⁷ et Megahit¹²⁸. Le choix d'un logiciel d'assemblage peut influencer la qualité de l'assemblage. Il est généralement conseillé de tester plusieurs logiciels, et de comparer leurs performances avec MetaQUAST¹²⁹ pour déterminer quel assemblage donnent les meilleurs résultats pour un jeu de données spécifique.

3. Annotation du génome/métagénome

Si les génomes/métagénomes ont été obtenus par co-assemblage, il est nécessaire de cartographier les contigs assemblés en différents fichiers qui contiendront les séquences des différents génomes/métagénomes. Cela permet d'aligner les séquences et d'obtenir certaines informations sur la qualité d'assemblage des génomes/métagénomes, comme le pourcentage de recouvrement de chaque génome/métagénome. Un exemple de logiciel souvent utilisé pour cette étape est Bowtie2¹³⁰.

Après assemblage, le génome/métagénome est souvent annoté. L'annotation est dépendante d'un bon assemblage. En effet, plus la séquence d'un contig assemblé est longue, plus elle contiendra d'informations, ce qui facilitera les comparaisons avec des séquences connues dans les bases de données. Certains logiciels d'annotations, comme RAST¹³¹ ou IMG¹³², requièrent des contigs assemblés d'au moins 30 000 paires de bases.

L'annotation se fait en deux étapes : la prédiction des gènes et l'annotation fonctionnelle. La prédiction des gènes consiste à étiqueter la séquence des éléments génomiques (Figure 8). Différents logiciels ont été mis au point pour prédire les séquences codantes génomiques et métagénomiques, comme GeneMark¹³³ ou MetaGeneAnnotator¹³⁴ respectivement, ainsi que les séquences non codantes, comme tRNAscan-SE¹³⁵ pour l'annotation des tRNAs. L'annotation fonctionnelle est l'annotation des fonctions putatives des gènes, et elle peut être réalisée à partir de comparaison à différentes bases de données, comme KEGG¹³⁶ et PFAM¹³⁷. Aucune base de données ne peut couvrir l'ensemble des fonctions biologiques des gènes, cependant plusieurs plateformes ont été conçues pour faciliter les annotations des séquences géniques et les comparaisons à différentes bases de données en même temps, comme MG-RAST¹³⁸ et IMG/M¹³⁹. Ces plateformes permettent également le stockage des données génomiques et métagénomiques assemblées et annotées. Dans le cadre du chapitre 3 de cette thèse, nous avons utilisé des génomes et des métagénomes de bactéries symbiotiques de l'intestin de fourmis du genre *Cephalotes*, dont les séquences génomiques et métagénomique étaient stockés sur la plateforme IMG/M¹⁸.

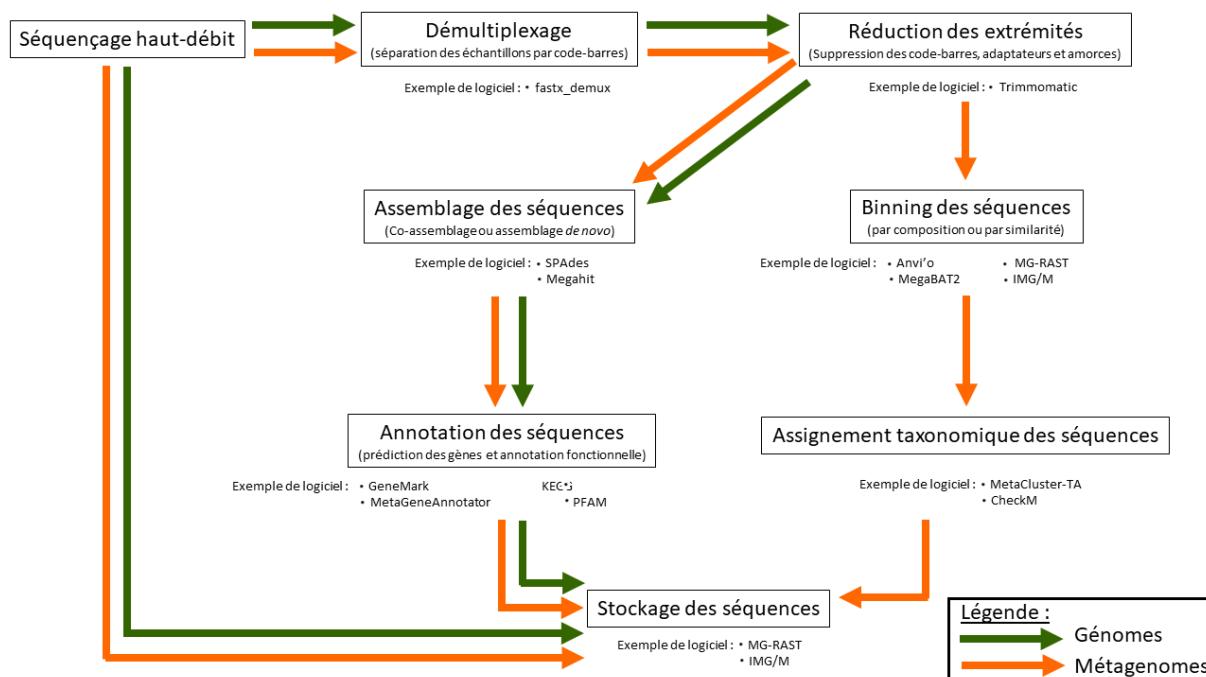


Figure 8. Résumé du processus de travail et exemples de logiciels utilisés pour l'analyse de séquences génomiques et métagénomiques.

4. Recouvrement des génomes à partir d'un métagénome

Le recouvrement des différents génomes composant les métagénomes est appelé « binning ». Ce processus consiste à analyser les séquences ADN contenues dans un métagénome et à les séparer dans différents groupes, appelés bins, qui représenteraient différents génomes ou qui représenteraient des génomes provenant d'organismes phylogénétiquement très proches.

Ce processus peut être fait de deux façons différentes : basé sur la composition ou basé sur la similarité (Figure 8). Le binning basé sur la composition utilise la conservation de la composition des nucléotides dans les génomes. Le binning basé sur la similarité compare les séquences du métagénome à des séquences géniques connues dans les bases de données. Des exemples de logiciels utilisés dans le binning par composition sont CONCOCT¹⁴⁰ et MetaBAT2¹⁴¹, tandis que des exemples de plateformes utilisées dans le binning par similarité sont MG-RAST¹³⁸ et IMG/M¹³⁹. Certaines plateformes permettent également un affinement manuel dans la création des bins. C'est le cas de la plateforme

Anvi'o¹⁴² qui a été utilisée pour réaliser le binning des métagénomés dans le cadre du chapitre 3 de cette thèse (Figure 9).

Le binning peut être fait sur un métagénome assemblé ou non, cependant logiciels nécessitent un métagénome assemblé car ils sont incapables de traiter des séquences courtes. C'est notamment le cas des logiciels utilisant le binning basé sur la composition. Le binning basé sur la similarité peut traiter des séquences plus courtes. Cependant, si les bases de données ne contiennent aucune séquence de gènes d'organismes phylogénétiquement proches d'une séquence contenue dans le métagénome, l'assignation dans une bin ne pourra être fait qu'à un niveau taxonomique élevé comme le phylum ou l'ordre. Si le métagénome contient plusieurs génomes qui ne peuvent être associés par similarité qu'à un niveau taxonomique élevé, ces génomes pourraient être placés dans la même bin. Cela pourrait entraîner la création de bins chimériques, qui sont constituées de génomes différents. Dans ce cas, utiliser le binning par composition sur une bin chimérique peut permettre de mieux séparer les différents génomes la composant.

Après binning, il est important d'effectuer l'assignement taxonomique des bins créées (Figure 8). Pour cela, de nombreux logiciels ont été développées, comme MetaCluster-TA¹⁴³ et CheckM¹⁴⁴. C'est ce dernier logiciel qui a été utilisé pour déterminer l'assignement taxonomique des bins dans le cadre du chapitre 3 de cette thèse (Figure 9).

5. Recherche de clusters de gènes de biosynthèse

L'objectif du chapitre 3 de cette thèse était d'identifier les clusters de gènes de biosynthèse présents dans les génomes et métagénomés de bactéries symbiotiques de l'intestin de fourmis du genre *Cephalotes* (Figure 9). Pour cela, après binning des métagénomés, les différents génomes et métagénomiques bins ont été analysés par la plateforme AntiSMASH¹⁴⁵. Cette plateforme fonctionne en plusieurs étapes et rassemble différents logiciels afin de repérer et identifier les différents clusters de gènes de biosynthèse. La première étape consiste à utiliser le logiciel HMMER¹⁴⁶ pour identifier les signatures des gènes principaux composant des clusters de gènes de biosynthèse. La deuxième étape

est d'analyser les séquences de chaque côté du gène signature identifié afin de détecter les gènes additionnels composant le cluster. La troisième étape consiste à utiliser le logiciel BLAST¹⁴⁷ pour chercher des composés connus possédant des gènes similaires aux clusters de gènes de biosynthèse identifiés.

Les clusters de gènes de biosynthèse peuvent être de tailles très variées. On estime que les clusters de gènes de biosynthèse les plus petits doivent mesurer au moins 5000 paires de bases, les clusters plus petits étant certainement incomplets, mais ils peuvent également mesurer plus de 60 000 paires de bases. Il peut également y avoir des disparités de tailles dans un même type de cluster.

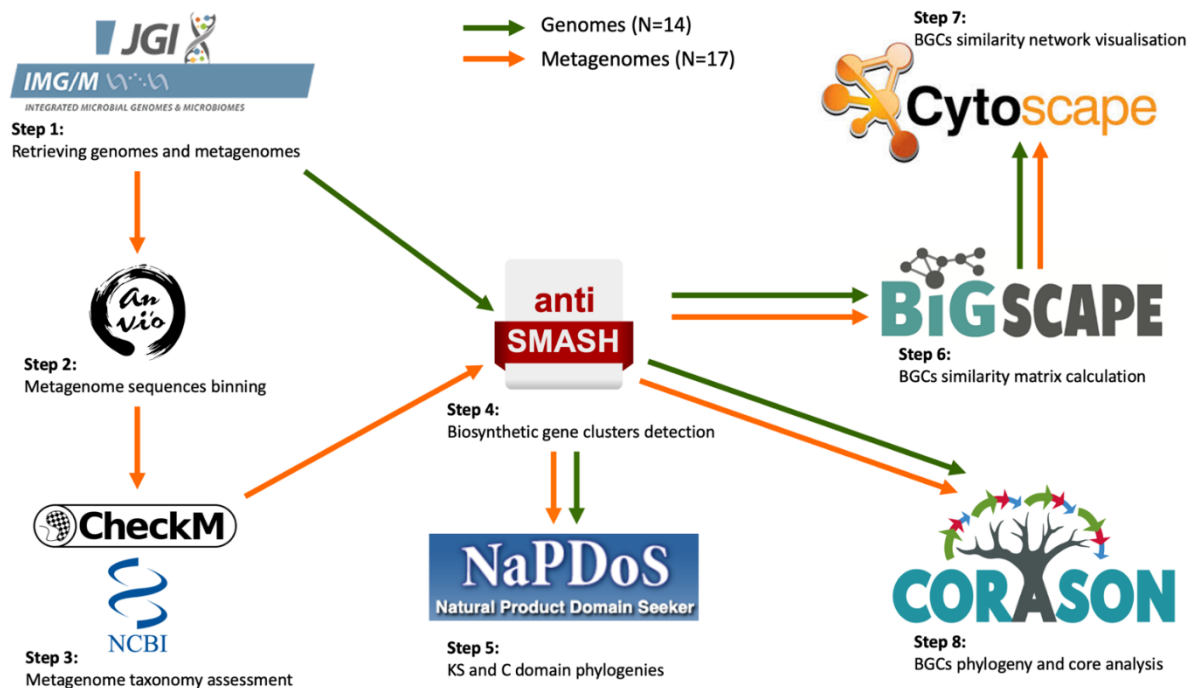


Figure 9. Résumé du processus de travail et des logiciels utilisés dans le chapitre 3 pour l'analyse des clusters de gènes de biosynthèse à partir de génomes et métagénomés bactériens associés à l'intestin des fourmis du genre *Cephalotes*.

IV. Techniques de mesure et de comparaison de la structure des communautés

A. Diversité alpha

Dans la première partie de cette thèse, nous avons montré que la structure d'une communauté se mesure par la richesse spécifique et l'uniformité des espèces, en utilisant le nombre d'espèces différentes (S) et le nombre total d'individus (N). Il est également possible de mesurer le niveau de diversité dans une communauté et entre plusieurs communautés. Dans les études d'analyse de microbiome, les communautés sont représentées par les différents échantillons analysés. Tout comme pour l'analyse des données bio-informatiques, la plateforme Qiime2 rassemble plusieurs méthodes de mesure de niveau de diversité dans les différents échantillons analysés. Ce sont certaines de ces méthodes de mesure et de comparaison de diversité microbienne que nous avons utilisé dans les chapitres 1 et 2 de cette thèse.

La diversité alpha est la mesure de la diversité des espèces dans un seul échantillon. La commande « diversity core-metrics-phylogenetic » permet d'obtenir les vecteurs de diversité alpha de quatre méthodes couramment utilisées dans l'analyse de microbiome (Figure 10) :

- l'indice de Shannon¹⁴⁸ : $H' = - \sum_{i=1}^R \frac{n_i}{N} \ln \frac{n_i}{N}$
- l'indice de diversité phylogénétique de Faith¹⁴⁹ : $PD = \sum_{branches\ b} \ell(b)$
- l'indice d'uniformité de Pielou²⁶ : $E = \frac{- \sum_{i=1}^R \frac{n_i}{N} \ln \frac{n_i}{N}}{\ln S}$
- l'indice de richesse en ASVs¹¹⁸ : $R = S$

L'indice de Shannon est une mesure de la diversité calculée en utilisant la richesse spécifique et la répartition des espèces présentes dans un échantillon. L'indice de diversité phylogénétique de Faith entre deux espèces est la somme des longueurs de branches entre ces espèces. C'est donc une mesure de la diversité qui prend en compte les différences phylogénétiques entre les espèces présentes dans un échantillon. L'indice d'uniformité de Pielou utilise l'indice de Shannon présenté précédemment pour calculer la mesure d'uniformité des espèces présentes dans un échantillon. Enfin, l'indice de richesse en ASVs calcule le nombre d'espèces différentes dans un échantillon. Comme ces différents

indices ne calculent pas la diversité alpha de la même façon, il est intéressant d'en utiliser plusieurs sur nos échantillons afin d'avoir une idée plus complète de la structure de la communauté présente dans chaque échantillon.

Après avoir obtenu ces vecteurs pour nos différents échantillons, il est possible de statistiquement calculer si la diversité alpha dans chaque échantillon est différente en fonction des échantillons, pour déterminer si certains échantillons ont une richesse d'espèces plus grandes que d'autres. Pour cela sur Qiime2 on utilise la commande « diversity alpha-group-significance » (Figure 10), qui est basée sur le test de rang de Kruskal-Wallis¹⁵⁰.

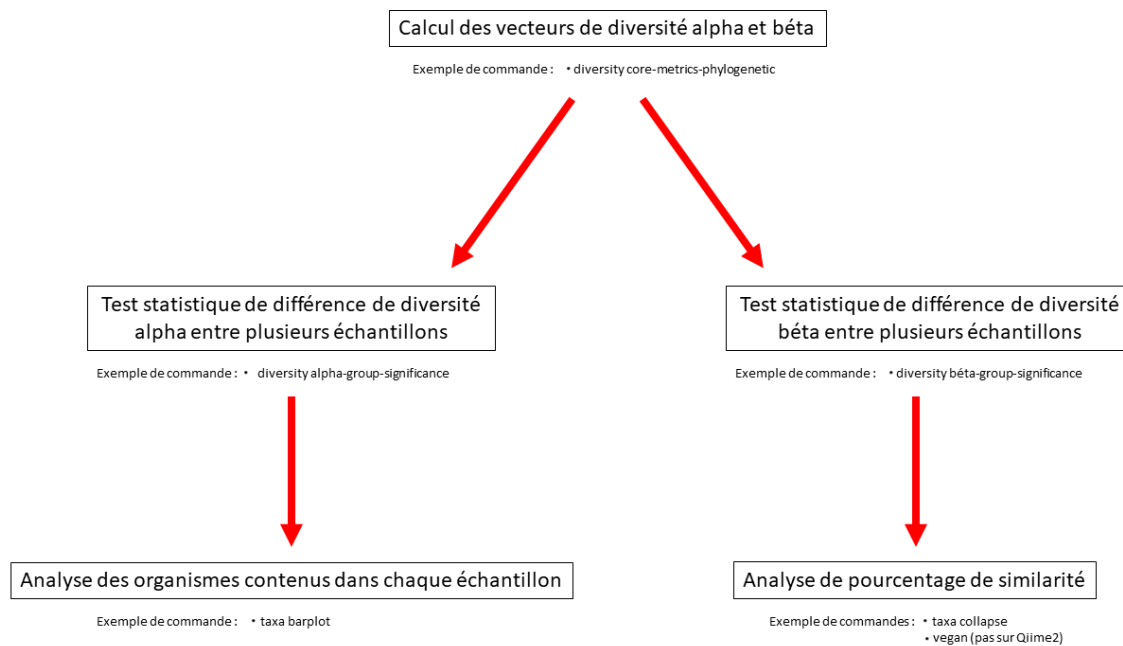


Figure 10. Résumé du processus de travail et exemples de commandes utilisées pour analyser la structure des communautés dans différents échantillons sur la plateforme Qiime2.

B. Diversité bêta

La diversité bêta est la mesure de la diversité des espèces entre plusieurs échantillons. La commande « diversity core-metrics-phylogenetic » permet également d'obtenir les vecteurs de diversité bêta de quatre méthodes couramment utilisées dans l'analyse de microbiome (Figure 10) :

- l'indice de similarité de Jaccard¹⁵¹ : $J(A,B) = |A \cap B| / |A \cup B|$
- l'indice de dissimilarité de Bray-Curtis¹⁵² : $PD = 1 - \frac{2 \sum_{i=1}^p \min(N_{ij}, N_{ik})}{\sum_{i=1}^p (N_{ij} + N_{ik})}$
- la distance unifrac qualitative¹⁵³ : $u = \frac{\sum_{i=1}^N l_i |A_i - B_i|}{\sum_{i=1}^N l_i \max(A_i, B_i)}$
- la distance unifrac quantitative¹⁵⁴ : $W = \frac{\sum_i^n l_i \left| \frac{A_i - B_i}{A_T - B_T} \right|}{\sum_{j=1}^S L_j}$

L'indice de similarité de Jaccard permet d'évaluer la similarité entre plusieurs échantillons en se basant uniquement sur la présence/absence des espèces présentes sans tenir compte de leur abondance respective. A l'inverse, l'indice de dissimilarité de Bray-Curtis permet d'évaluer la dissimilarité entre plusieurs échantillons en se basant sur l'abondance des espèces présentes. La distance unifrac qualitative est une mesure de la longueur des branches entre les différentes espèces présentes en se basant uniquement sur la présence/absence des espèces présentes. Elle tient donc compte des différences phylogénétiques entre les espèces présentes dans un échantillon. Enfin, la distance unifrac quantitative est aussi une mesure de la longueur des branches entre les différentes espèces présentes, mais prend également en considération l'abondance de chaque espèce présente. Tout comme pour la diversité alpha, il est intéressant d'utiliser plusieurs mesures de la diversité bêta entre différents groupes d'échantillons afin d'avoir une idée plus complète des différences de structure de communauté entre ces différents groupes d'échantillons.

Après avoir obtenu ces vecteurs pour nos différents échantillons, il est possible de statistiquement comparer la diversité bêta entre plusieurs échantillons en fonction d'un facteur spécifique, comme l'habitat où l'échantillon a été collecté par exemple, afin de former des groupes d'échantillons. Cela permet de déterminer si les échantillons présents dans un même groupe ont une structure de communauté plus similaire entre eux que par rapport à un autre groupe. Pour cela sur Qiime2 on utilise

la commande « diversity beta-group-significance » (Figure 10), qui est basée sur le d'analyse multivariée des variances par permutation (PERMANOVA)¹⁵⁵.

C. Analyse de pourcentage de similarité (SIMPER)

Après avoir étudié les diversités alpha et bêta de différents échantillons, il peut être intéressant de regarder quels organismes contiennent chaque échantillon, et quels organismes contribuent le plus à la diversité observée dans chaque échantillon.

Pour analyser quels organismes sont contenues dans chaque échantillon, sur Qiime2 on peut utiliser la commande « taxa barplot » (Figure 10). Un des avantages de cette commande est qu'elle permet d'obtenir un fichier dans lequel on peut spécifier le rang taxonomique d'intérêt. Cela permet de grouper et d'afficher les organismes contenus dans chaque échantillon en fonction de ce rang taxonomique défini, comme le phylum ou l'ordre par exemple.

Ensuite, pour tester quels organismes contribuent le plus à la diversité observée dans chaque échantillon, on utilise le pourcentage de similarité (SIMPER)¹⁵⁶. Le pourcentage de similarité est basé sur l'indice de dissimilarité de Bray-Curtis. Il calcule le pourcentage de contribution de chaque espèce à cet indice de dissimilarité de Bray-Curtis, selon la formule suivante : $d[jk] = \frac{abs(N[ij]-N[ik])}{sum(N[ij]+N[ik])}$.

Après avoir calculé le pourcentage de contribution de chaque espèce à l'indice de dissimilarité de Bray-Curtis, il est important de déterminer si l'abonde de ces espèces est différentes entre plusieurs groupes d'échantillons. Pour cela on utilise le test de rang de Kruskal-Wallis¹⁵⁰. Le pourcentage de dissimilarité présente généralement les espèces qui contribuent à au moins 70% à la différence de diversité entre plusieurs groupes d'échantillons. Cette méthode est donc très sensible à l'abondance des espèces. Le défaut majeur est que les espèces avec une très faible abondance ne seront pas présentées dans les résultats, même si leur présence et abondance et différente entre les différents groupes d'échantillons.

Le pourcentage de similarité se base donc sur la présence et l'abondance des différentes espèces. Cependant, dans les analyses de microbiome il est souvent difficile d'identifier toutes les organismes présentes au rang d'espèces. Ainsi, pour prendre en considération le plus d'organismes possibles, il peut être intéressant d'utiliser un rang taxonomique plus large mais identifié dans un plus grand nombre d'organismes. Sur Qiime2, il est possible de fusionner les différents ASVs déterminés dans chaque échantillon s'ils partagent un assignement taxonomique commun en utilisant la commande « taxa collapse » (Figure 10). Par exemple, en utilisant cette commande on peut fusionner les ASVs selon leur phylum, puis réaliser l'analyse de pourcentage de similarité. L'analyse de pourcentage de similarité en elle-même n'est pas disponible sur Qiime2, mais elle peut cependant être réalisée sur le logiciel R¹²³ à l'aide du package « vegan »¹⁵⁷.

Chapitre I: Etude des facteurs structurant les communautés bactériennes associées aux fourmis amazoniennes

I. Introduction

Les fourmis possèdent une grande diversité d'espèces, notamment en Amazonie. En effet, on estime qu'il y a plus de 1000 espèces de fourmis différentes en Amazonie. De plus, les fourmis Amazoniennes possèdent également différentes écologies, comme différents modes de nidification (terrestre ou arboricole), différents régimes alimentaires (carnivores ou herbivores), et différents habitats (forestiers ou urbains).

Cette variété d'espèces et d'écologies pourrait être renforcée par la diversité des communautés bactériennes associées à ces fourmis. En effet, certaines études ont montré que les communautés bactériennes pourraient jouer un rôle important dans la maintenance des fourmis vivant dans la canopée^{16,158}, et que les communautés bactériennes associées aux fourmis diffèrent en fonction du régime alimentaire de la fourmi¹⁷. De plus, les communautés bactériennes associées aux fourmis varient entre les genres de fourmis. En effet, de nombreuses études sur une variété de genres de fourmis ont mis en évidence différentes communautés bactériennes associées à ces fourmis^{18,159-165}

Ainsi, l'étude des communautés bactériennes associées à un genre de fourmis ou un facteur écologique spécifique des fourmis ne permet pas d'analyser la diversité des communautés bactériennes associées aux fourmis. De plus, si plusieurs facteurs jouent un rôle dans la structuration des communautés bactériennes associées aux fourmis, on peut se demander si certains facteurs ont un rôle plus important que d'autres. Par exemple, deux espèces de fourmis appartenant au même genre et ayant différents régimes alimentaires pourraient avoir des communautés bactériennes similaires entre elles. Cependant, ces fourmis pourraient également avoir des communautés bactériennes différentes entre elles mais similaires à des fourmis d'autres genres ayant le même régime alimentaire. Pour déterminer l'importance des facteurs taxonomiques et écologiques dans la

structuration des communautés bactériennes, il est donc nécessaire d'étudier les communautés bactériennes associées à des fourmis issues de différents taxons et possédant différentes écologies.

La technique du métabarcoding présentée précédemment dans l'état de l'art permet d'obtenir la diversité des bactéries associées à différents échantillons de fourmis en utilisant comme marqueur taxonomique l'ARN ribosomique 16S, qui est un gène spécifique des procaryotes. Cette technique permet de recenser les communautés bactériennes associées à chaque échantillon de fourmis, et d'étudier comment ces communautés bactériennes sont structurées.

Dans ce contexte, nous avons mené une étude d'analyse des communautés bactériennes associées à 38 espèces de fourmis différentes, récoltées dans la forêt de la réserve naturelle des Nouragues et dans plusieurs endroits de la ville de Cayenne. L'objectif de ce travail est de déterminer si les communautés bactériennes associées aux fourmis sont différentes en fonction de l'espèce de la fourmi, de son régime alimentaire, de son nid ou de l'habitat dans lequel elle a été collectée. Nous avons reçu une invitation à publier un article dans une édition spéciale « Microbiota in insects » dans le journal « Microorganisms » (ISSN 2076-2607). Les chapitres 1 et 2 de cette thèse seront assemblés dans un seul article pour être soumis dans cette issue spéciale fin juin 2021. C'est pourquoi ce chapitre 1 est rédigé en anglais dans ce manuscrit de thèse.

II. Impact of Amazonian ants nesting mode, diet and taxonomy in structuring their associated bacterial communities

Anaïs Chanson¹, Corrie S. Moreau², Christophe Duplais^{3*}

¹Université de Guyane, UMR8172 EcoFoG, AgroParisTech, CNRS, Cirad, INRAE, Université des Antilles, Kourou, France

²Departments of Entomology and Ecology & Evolutionary Biology, Cornell University, Ithaca, NY, USA

³CNRS, UMR8172 EcoFoG, AgroParisTech, Cirad, INRAE, Université des Antilles, Université de Guyane, Kourou, France

***Corresponding author**

Christophe Duplais

CNRS-UMR EcoFoG at the Pasteur Institute of French Guiana

23 avenue Pasteur, 97300 Cayenne, France

christophe.duplais@cnrs.fr

Phone: +594 594 293 134

Keywords: Formicinae, insect, microbes, 16S rRNA, amplicon sequencing

A. Abstract

Host-associated bacterial communities have been demonstrated to be impacted by numerous host factors, like habitat, diet and phylogeny. In insects some genera are known to rely on their bacterial communities to supplement their diet. Herein, we studied the structure of bacterial communities associated with Amazonian ants collected in the rainforest and in the city. We collected a wide taxonomic range of ants and we used 16S rRNA amplicon sequencing to examine the impact of ants ecological and phylogenetic factors on their bacterial communities. We investigated the roles of ant host habitat, nesting mode, diet and taxonomy on structuring the ant bacterial communities. Our results reveal that (1) habitat does not structure bacterial communities, (2) diet and nesting mode structure bacterial communities, (3) host phylogeny is mostly not correlated to bacterial diversity except for a few ant genera in which evolutionary history has an impact on the structure of their bacterial communities. This work contributes to our understanding how host ecology structures bacterial communities.

B. Introduction

One of the important goals of community ecology is to understand the maintenance and consequences of interactions between different communities, especially in ecosystems becoming more and more fragmented¹⁶⁶. Every organism has biological needs which can be fulfilled only under specific environmental conditions, such as the temperature, altitude, and humidity. Habitat filtering is one of the process invoked to explain why in a given habitat only individual species possessing suitable traits for this specific habitat can persist^{167,168}. Organisms rely on specific resources, like their diet, nesting options, dispersal ability, or even the size of their territory, to not only survive but also to have thriving populations. When organisms living in the same local environment rely on the same resources, they will engage in competition for these resources, often leading to competitive exclusion¹⁶⁹. In contrast, when some organisms living in the same local environment do not rely on the same resources, they can coexist without competition. This process is described as niche

differentiation^{168,170}. Furthermore, most communities are very complex, making it even more challenging to study them. For example, different communities often vary in the dominant species and the number of species they contain. As such, to understand the factors impacting the way these interacting communities are structured, it is necessary to take into account these different parameters.

Bacterial communities are ubiquitous and can be found in all types of environments, even the most extreme^{171,172}. Environmental factors have been shown to have strong impact on bacterial richness and diversity in species rich biomes. For example, temperature, light, and water acidity correlated to different bacterial distributions in marine environments^{173–176} and soil^{177–180}. Bacterial communities are also often found in associations with diverse host organisms. In insects, symbiotic bacteria have shaped the evolution of their host and are key components for providing basic functions to their host^{181,182}. Mutualistic bacteria are known to complement nutritionally insufficient diet for their host, like in wood-feeding insects¹⁸³, blood-sucking insects², sap-sucking insects¹⁸⁴, and herbivorous insects¹. They also help maintain and enhance their host health and immune system^{71,185}, and are involved in enlarging their host environmental tolerance, such as temperature range^{186,187} and humidity level^{188,189}. On the other hand, parasitic bacteria can manipulate host reproduction⁸ and dispersion^{190,191}.

Among insects, ants represent a species rich clade with different ecologies¹⁵ and provide many ecosystem services¹⁴. Symbiotic bacteria are thought to be in part responsible for their evolutionary success. For example, it has been hypothesized that symbiotic bacteria enable ants to dominate the rainforest canopies^{16,158}. Numerous studies have shown that symbiotic bacteria in ants differ with the ant diet, especially between herbivorous ants and carnivorous ants¹⁷. Predatory army ants possess specialized Firmicutes and Entomoplasmatales gut bacteria, which are common across army ant lineages with a symbiosis dating from the Cretaceous, and which are likely socially transmitted (by trophallaxis or coprophagy)^{192,193}. The functional role of these bacteria in army ants is not yet known due to the lack of available symbiont genomes from army ants, but a nutritional role for these symbionts have been suggested¹⁹³. Herbivorous turtle ants rely on symbiotic gut bacteria to recycle nitrogen into amino acids^{107,194} whereas the Camponotini tribe benefit from the production of amino

acids by obligate intracellular symbiont *Blochmannia*^{1,195}. In addition, many studies have investigated the bacterial communities of specific ant clades, like *Daceton*¹⁶⁵, *Paraponera*¹⁶¹, ponerine ants¹⁶⁰, *Pseudomyrmex*¹⁶⁴, *Solenopsis*¹⁵⁹, and spiny ants¹⁶². Several bacteria symbionts are associated across these multiple ant clades. For example, Acetobacterales, Entomoplasmatales and Rhizobiales are bacterial orders commonly found in ant guts. Their role in ants is not fully understood, but Acetobacterales might be involved in larval immune function¹⁹⁶ and development¹⁹⁷, while Entomoplasmatales might be involved in processing of chitin of insect prey⁶⁵, and Rhizobiales might be involved in degrading proteins¹⁹⁸ and urea recycling pathways¹⁹⁹. The symbiont *Wolbachia* is found across multiple ant clades and may induce reproductive changes²⁰⁰ and vitamin B supplementation²⁰¹.

We aimed to better understand which factors are structuring the bacterial communities associated with a diverse group of ants. To this aim, we collected 38 ant species in French Guiana from a wide phylogenetic range spanning several ant subfamilies, and we studied the bacterial communities associated with each sample both in terms of abundance and diversity, using 16S rRNA amplicon sequencing and qPCR. Our ant sampling strategy had three aims. First, the wide phylogenetic range chosen allows us to control for the host's evolutionary history. Second, the species collected possess different ecological traits (diet and nesting mode), which will permit us to test these factors for their correlation with the host-associated bacterial communities. And third, among the 38 collected ant species, nine species were collected both in the rainforest and in the city, which will allow us to test the possible impact of the environment on the same ant species. The list of collected samples and different phylogenetic (subfamily, genera) and ecological factors (habitat, nesting mode, and diet) chosen are presented in Supplementary File 1. This study gives a broader understanding on the different structuring factors of bacterial communities associated with ants in a tropical environment.

C. Material and Methods

Sample collections

Samples were collected in two sites in French Guiana, at the Nouragues Reserve and in the city of Cayenne (Supplementary Figure 1). A total of 49 ant nests were collected, representing 38 ant species from 18 genera spanning eight subfamilies. The complete list of collected species is detailed in Supplementary File 1. For each nest we also collected a sample of the nest material or foraging area.

DNA extractions

DNA extractions from ant and nest samples were performed using the DNeasy PowerSoil Kit (Qiagen, USA) following the manufacturer's protocol with a few modifications before the first step. For the DNA analyses, a single worker from each collected nest was used for the ants, and between 0.1 and 0.25 g of nest material were used for the nest. First, each ant samples and nest samples were introduced in a single 1.5 mL Eppendorf tube. Then, 1 mL of liquid nitrogen was added to each Eppendorf tube, and each sample was immediately manually crushed in the tube with a sterile micropestel. To avoid any contamination between samples during this step, a new sterile micropestel was used for each different sample tube. The crushed samples were then transferred to tubes of 500 μ L PowerSoil bead solution. Following the procedure suggested in Rubin et al.⁸⁷, 60 μ L of solution C1 and 100 μ g of proteinase K were added to each PowerSoil tube, and tubes were incubated at 56°C overnight. The extraction then proceeded following the DNeasy PowerSoil Kit protocol. The exact same protocol was followed with 4 blank tubes containing no ant or nest sample, to serve as negative controls. Filtered pipette tips and sterile techniques were used in every step to avoid contamination⁸⁶. All DNA samples were quantified via Qubit to verify the success of the DNA extraction. The Qubit quantification was performed with the High Sensitivity Assay Kit (Life Technologies Corp., USA).

DNA amplification

Amplifications of the V4 region of 16S rRNA and Miseq sequencing of each DNA dataset were performed by the Argonne National Laboratory (Lemont, Illinois, USA). Amplifications were performed using 515F (5'-GTGCCAGCMGCCGCGGTAA) and 806R (5'-GGACTACHVGGGTWTCTAAT) primers as described in Caporaso et al.¹⁰¹, following the protocol suggested in the Earth Microbiome Project (EMP) (<http://www.earthmicrobiome.org/protocols-and-standards/16S/>). Each PCR reaction contained 12µL of DNA-free PCR water, 10 µL of 5 Prime HotMasterMix 1X, 1 µL of 5 mM forward primer, 1 µL of 5 mM Golay barcode tagged reverse primer, and 1 µL of extracted DNA. The amplification conditions were as follow: 94°C for 3 min, with 35 cycles at 94°C for 45 s, 50°C for 60 s, and 72°C for 90 s, with a final cycle of 10 min at 72°C. Each PCR reaction was performed 3 times. 1% agarose gel electrophoresis was performed to confirm the efficiency of the amplification. Amplification samples from each dataset were pooled separately, each pool containing 100 µL, and were cleaned using the QIAquick PCR Purification Kit (Qiagen, USA), following the manufacturer's instructions. The molarity of the pool was determined and diluted down to 4 nM, denatured, and then diluted to a final concentration of 6.75pM with 10% PhiX for sequencing. The pooled samples from each dataset were sequenced separately in two runs with the MiSeq Illumina V3 Reagent Kit 600 Cycles (300x300) using the custom sequencing primers and procedures described in the supplementary methods in Caporaso et al.¹⁰¹.

Bacterial quantification

The bacterial quantification of each sample was performed through qPCR quantification (Thermo Fisher Scientific) on real-time CFX Connect equipment (Bio-Rad). SYBRAdvanced 2X (Bio-Rad) SYBR green supermix and 2 µL of extracted DNA were used to verify the total amount of bacteria present in each sample. Amplification of the V4 region of 16S rRNA in the qPCR were performed using 515F and 806R primers as described in Caporaso et al.¹⁰¹, following the protocol suggested in the Earth Microbiome Project (EMP) (<http://www.earthmicrobiome.org/protocols-and-standards/16S/>). Each

qPCR reaction was performed in triplicate. Standard curves were generated from serial dilutions of linearized plasmids containing *E. coli* 16S rRNA inserts, following the same parameters of Rubin et al.⁸⁷. All triplicate qPCRs values were satisfactory and had R2 from 70% and 100%. Then, the mean of the triplicate values for each sample were used in the analysis. Negative controls were also analyzed following the same protocol.

Bacterial diversity

Demultiplexing of sequences and taxonomic assignments were performed separately for the ant and the nest sequences, but following the same protocol. Demultiplexed sequences were analyzed using the Qiime2-2019.1¹⁰⁹ with the plugin demux (<https://github.com/qiime2/q2-demux>). Sequence quality control and feature table construction were performed through the dada2 plugin¹¹³. Taxonomic assignment was conducted with the SILVA_132_QIIME database²⁰² and the ASV (amplicon sequence variants) were selected with 99% identity. To generate the taxonomy table, paired-end sequence reads were trimmed in the v4 region of 16S rRNA with the 515F/806R primers. Thereby, our own classifier was created using the "feature-classifier fit-classifier-naive-bayes" command. Once the classifier was obtained, the reads were classified by taxon using the "feature-classifier classify-sklearn" command²⁰³.

The filtration of contaminants (blank samples) in the datasets were performed with the Decontam package¹²⁴ of the R software version 4.02²⁰⁴. In this package, the prevalence method was used to remove the contaminant sequences. The decontaminated datasets were inserted back into Qiime2¹⁰⁹ to filter all mitochondria and chloroplasts sequences from both datasets. The alignment was performed using the align-to-tree-mafft-fasttree command²⁰⁵, to reconstruct the bacterial phylogeny.

The alpha and beta diversity analyses were performed by using the "qiime diversity core-metrics-phylogenetic" command. Alpha and beta diversities were visualized using the visualization interface <https://view.qiime2.org/>. Alpha diversity metrics computed were Shannon index, Faith's phylogenetic index, Pielou's evenness and ASV richness. The Shannon index is a measure of species richness that

incorporates community abundance and evenness¹⁴⁸. The Faith's phylogenetic index is a measure of biodiversity that incorporates phylogenetic difference between samples¹⁴⁹. The Pielou's evenness is a measure of relative evenness of species richness²⁶. The ASV richness is the mean number of ASVs per sample¹¹⁸. Beta diversity metrics computed were Jaccard similarity index, Bray-Curtis dissimilarity, unweighted unifrac distance and weighted unnormalized unifrac distance. The Jaccard similarity index is a measure of unique taxa without accounting for abundance¹⁵¹. The Bray-Curtis dissimilarity is a measure of overabundant taxa¹⁵². The unweighted unifrac distance is a measure of unique branch length¹⁵³. The weighted unnormalized unifrac distance is a measure of abundance but does not correct for different evolutionary rates between taxa¹⁵⁴.

Statistical analysis

Statistical analyses of bacterial quantification were performed using one-way ANOVAs (Analysis of Variance) with the PAST software version 4.02²⁰⁶. Statistical analyses of alpha diversity and beta diversity were performed on Qiime2-2019.1¹⁰⁹, using the "diversity alpha-group-significance" command with the pairwise Kruskal-Wallis methods, and the "diversity beta-group-significance" command with the pairwise permanova method and 999 permutations respectively. The results of bacterial quantification and alpha diversity were visualized using PCoAs (Principal Coordinates Analysis) with the R packages "ggpubr" version 0.3.0²⁰⁷ and "PMCMR" version 4.3²⁰⁸. The results of beta diversity were visualized using PCoAs with the R packages "ggplot2" version 3.3.0²⁰⁹ and "ggfortify" version 0.4.10²¹⁰. The correlation between ant evolutionary history and bacterial composition was performed using a Mantel test using the R package phytools²¹¹. The contribution of each ASV at the order level for each sample was obtained using the "qiime taxa barplot" and "qiime taxa collapse" commands on Qiime2-2019.1¹⁰⁹. The SIMPER (Similarity Percentage) analyses were performed using the PAST software version 4.02²⁰⁶ and visualized using boxplots with the R packages "ggpubr" version 0.3.0²⁰⁷ and "PMCMR" version 4.3²⁰⁸.

D. Results

Assessing sequencing quality

A total of 49 ant samples and 44 nest samples were sequenced with four control samples (two for each type of sample) (Supplementary File 1). For the ant samples a total of 1,879,666 reads were sequenced. After rarefaction at a sampling depth of 4,500 reads and three samples removed, a total of 5,908 ASVs were retrieved in 46 ant samples ranging from 7,548 to 80,226 reads, with a mean frequency of 40,842 reads per sample. The rarefaction curve shows that sequences from every samples reached a plateau, indicating that the majority of the bacterial diversity has been recovered (Supplementary Figure 2A). However, three samples (CSM3695a, CSM3695b and PJF10) were excluded from the dataset because they did not reach the minimum sampling depth of 4,500 reads. Nest samples were rarefied to 9,000 reads (Supplementary Figure 2B) and in total 19,934 ASVs were obtained from 2,350,607 reads sequenced in the 44 nest samples, ranging from 9,130 to 76,765 reads, with a mean frequency of 53,442 reads per sample.

Alpha diversity analyses

The bacterial alpha diversity associated with Amazonian ants was tested with four metrics: the Shannon diversity, the Pielou's evenness, the Faith's phylogenetic diversity and the ASV richness. All metrics were calculated with Kruskal-Wallis ANOVAs, and all statistics for each alpha diversity metrics are in Supplementary File 2. For all these metrics, there were no statistical difference between rainforest ants and city ants (Supplementary Figure 3 ; Shannon diversity: $H=0.017$, $p\text{-value}=0.895$; Pielou's evenness: $H=0.002$, $p\text{-value}=0.965$; Faith's PD: $H=0.213$, $p\text{-value}=0.644$; ASV richness: $H=0.148$, $p\text{-value}=0.700$).

The Shannon diversity and Pielou's evenness were statistically higher in herbivorous ants than in carnivorous and omnivorous, even though the Shannon diversity was not overall statistically different between diets (Shannon diversity: Supplementary Figure 4A, $H=4.316$, $p\text{-value}=0.116$; Pielou's evenness: Supplementary Figure 4B, $H=6.431$, $p\text{-value}=0.040$). Additionally no difference was

detected between the diets in the other metrics (Faith's PD: Supplementary Figure 4C, $H=0.424$, $p\text{-value}=0.809$; ASV richness: Supplementary Figure 4D, $H=0.456$, $p\text{-value}=0.796$).

The Shannon diversity was higher in ground nesting ants compare to tree nesting and rotten log nesting ants and was also higher in tree nesting ants compare to rotten log nesting ants (Figure 11A, $H=15.563$, $p\text{-value}=4.173\text{E-}04$). The Pielou's evenness was higher in ground and tree nesting ants than in rotten log nesting ants (Figure 11B, $H=13.742$, $p\text{-value}=0.001$). The Faith' PD and the ASV richness were also higher in ground nesting ants than in tree and rotten log nesting ants (Faith' PD: Figure 11C, $H=12.159$, $p\text{-value}=0.002$; ASV richness: Figure 11D, $H=16.350$, $p\text{-value}=2.817\text{E-}04$).

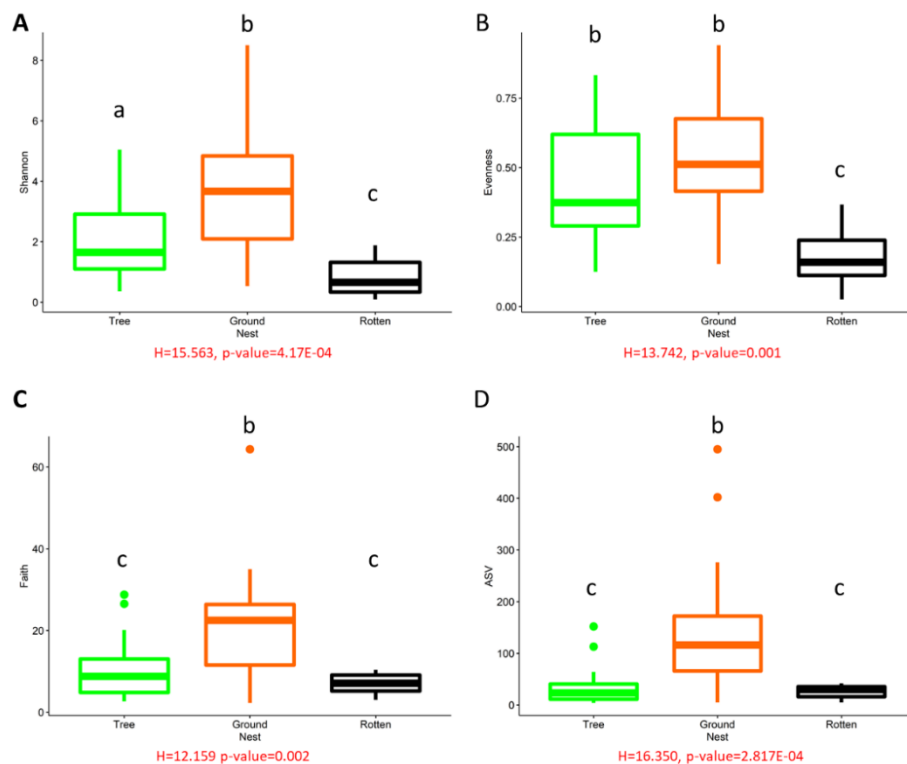


Figure 11. Differences in bacterial alpha diversity by ant nesting mode. (A) Mean (\pm standard error) Shannon diversity. (B) Mean (\pm standard error) Pielou's evenness. (C) Mean (\pm standard error) Faith's phylogenetic diversity. (D) mean (\pm standard error) ASV count. Different letters at the top of the figure illustrate nesting mode with significant differences ($P<0.05$) in this alpha diversity metric. The statistical p-values were obtained with pairwise Kruskal-Wallis. P-values for this figure are found in Supplementary File 2.

The Shannon diversity (Supplementary Figure 5A, $H=18.096$, $p\text{-value}=0.012$), Pielou's evenness (Supplementary Figure 5B, $H=17.280$, $p\text{-value}=0.016$) and ASV richness (Supplementary Figure 5D, $H=17.834$, $p\text{-value}=0.013$) were statistically different across the ant subfamilies. Overall, all three alpha diversity metrics were higher in Ectatomminae and Paraponerinae than in most of the other subfamilies. The Faith's PD was not statistically different across the subfamilies (Supplementary Figure 5C, $H=12.113$, $p\text{-value}=0.097$). Still, the Faith's PD was also higher in Ectatomminae and Paraponerinae than the other subfamilies.

The Shannon diversity (Figure 12A, $H=31.546$, $p\text{-value}=0.017$), Pielou's evenness (Figure 12B, $H=32.781$, $p\text{-value}=0.012$) and ASV richness (Figure 12D, $H=26.805$, $p\text{-value}=0.049$) were also statistically different across the ant genera. Overall, all three alpha diversity metrics were higher in *Atta*, *Cephalotes*, *Ectatomma*, *Gigantiops* and *Paraponera* than in the other genera. The Pielou's evenness were also higher in Dolichoderinae (*Azteca* and *Dolichoderus*), Dorylinae (*Eciton* and *Labidus*), and Formicinae (*Camponotus* and *Gigantiops*) than in Ponerinae (*Anochetus*, *Neoponera* and *Odontomachus*) and Pseudomyrmicinae (*Pseudomyrmex*). The Faith's PD was not statistically different across the ant genera (Figure 12C, $H=26.395$, $p\text{-value}=0.068$). Still, the Faith's PD was higher in *Anochetus*, *Atta*, *Ectatomma*, *Gigantiops*, *Labidus* and *Paraponera* than in the other genera.

Overall, the forest/city dataset, in which we kept only the samples from the nine species collected both in the forest and in the city, followed the same pattern as the whole dataset discussed above for every factor and all the tested metrics (Shannon diversity habitat: $H=0.045$, $p\text{-value}=0.833$; Pielou's evenness habitat: $H=0.005$, $p\text{-value}=0.944$; Faith's PD habitat: $H=0.079$, $p\text{-value}=0.778$; ASV richness habitat: $H=0.011$, $p\text{-value}=0.916$; Shannon diversity diet: $H=4.701$, $p\text{-value}=0.095$; Pielou's evenness diet: $H=3.964$, $p\text{-value}=0.138$; Faith's PD diet: $H=1.019$, $p\text{-value}=0.601$; ASV richness diet: $H=2.072$, $p\text{-value}=0.355$; Shannon diversity nesting mode: $H=5.834$, $p\text{-value}=0.045$; Pielou's evenness nesting mode: $H=6.390$, $p\text{-value}=0.041$; Faith's PD nesting mode: $H=10.139$, $p\text{-value}=0.007$; ASV richness nesting mode: $H=8.350$, $p\text{-value}=0.015$; Shannon diversity subfamily: $H=12.620$, $p\text{-value}=0.027$; Pielou's evenness subfamily: $H=11.378$, $p\text{-value}=0.044$; Faith's PD subfamily: $H=10.655$,

p-value = 0.059; ASV richness subfamily: $H=11.380$, p-value=0.044). However, in the forest/city dataset, contrary to the whole dataset, none of the four alpha diversity metrics tested were statistically different across ant genera (Shannon diversity: $H=16.494$, p-value=0.057; Pielou's evenness: $H=17.026$, p-value=0.058; Faith' PD: $H=15.788$, p-value=0.071; ASV richness: $H=15.236$, p-value=0.085).

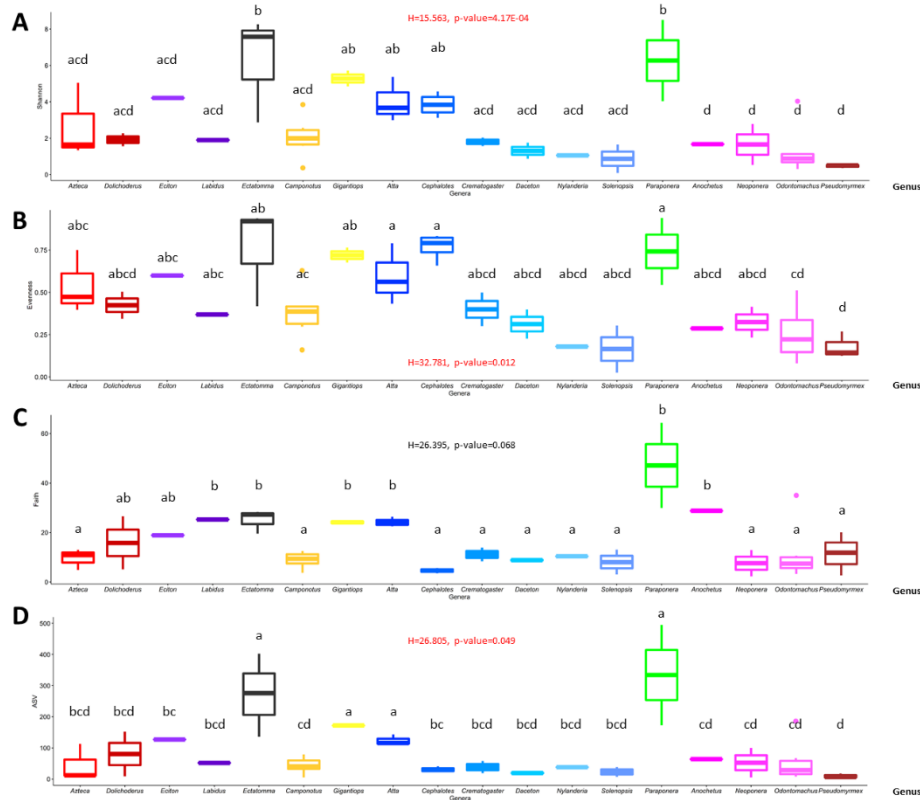


Figure 12 Differences in bacterial alpha diversity by ant genus. (A) Mean (\pm standard error) Shannon diversity. (B) Mean (\pm standard error) Pielou's evenness. (C) Mean (\pm standard error) Faith's phylogenetic diversity. (D) mean (\pm standard error) ASV count. Different letters at the top of the figure illustrate genera with significant differences ($P < 0.05$) in this alpha diversity metric. The statistical p-values were obtained with pairwise Kruskal-Wallis. P-values for this figure are found in Supplementary File 2.

Bacterial abundance analyses

To investigate which bacterial orders are the most abundant in bacteria associated with Amazonian ants, we used SIMPER analyses. The taxa barplot of bacterial relative abundance exhibited very diverse patterns across ant genera, habitat, nesting mode and diet (Figure 13, Supplementary File 3). The 15 most abundant bacterial orders found in samples were in order Rickettsiales, Enterobacteriales, Acetobacteriales, Lactobacillales, Burkholderiales, Xanthomonadales, Erysipelotrichales, Flavobacteriales, Pseudomonadales, Corynebacteriales, Entomoplasmatales, Opiritales, Spingomonadales and Micrococcales.

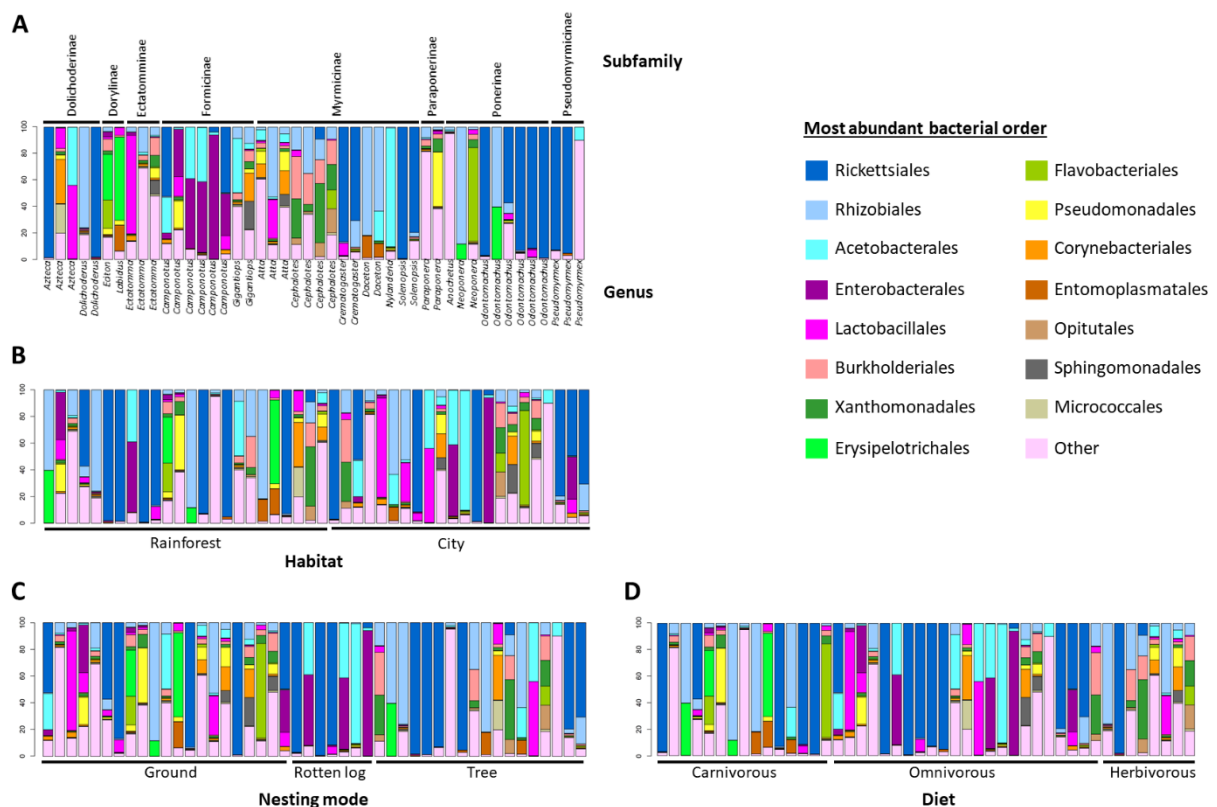


Figure 13. Taxa barplot based on relative abundance ordered by ant subfamily and genus (A), habitat (B), nesting mode (C) and diet (D). The relative abundance bars are colored by percentage of bacterial order of ASV. The 15 most abundant bacterial order associated with Amazonian ants are listed. The less abundant bacterial order are grouped under the term “Other”.

Rickettsiales were preponderant (at least 10% of relative abundance) in every habitat, diet and nesting mode tested (Figure 13B, C and D). Rhizobiales were also very common, being found across carnivorous and herbivorous ants (carnivorous: 21.38%; herbivorous: 23.63%), as well as in ants nesting in trees (21.36%) and in the ground (10.25%), and in city ants (10.93%). Acetobacterales and Enterobacteriales were common in omnivorous (respectively 13.76% and 12.71%) and ants nesting in rotten logs (respectively 24.3% and 28.69%). Acetobacterales were also common in city ants (10.93%). Burkholderiales and Xanthomonadales were common in herbivorous (respectively 11.00% and 10.98%), while Erysipelotrichales were common in carnivorous (9.29%).

The bacterial relative abundance had different patterns across ant genera, but were relatively consistent within the same genus, especially for *Camponotus*, *Cephalotes*, *Crematogaster*, *Daceton*, *Solenopsis* and *Odontomachus* (Figure 13A, Supplementary File 3). *Crematogaster* and *Odontomachus* were very conserved, consisting primarily of Rickettsiales (respectively 78.58% and 72.87%) and Rhizobiales (respectively 10.85% and 11.60%). Rickettsiales was also the most abundant bacterial order in *Solenopsis* (59.59%). *Camponotus* were dominated by Enterobacteriales (45.19%), but also contained Acetobacterales (18.47%) and Rickettsiales (17.60%). *Daceton* were dominated by Rhizobiales (72.49%), but also contained Entomoplasmatales (13.20%) and Acetobacterales (11.31%). *Cephalotes* are associated with Xanthomonadales (24.65%), Burkholderiales (22.71%) and Rhizobiales (19.19%). In addition, some of the most abundant bacterial orders were only found in a few ant genera. Erysipelotrichales (Firmicutes) were only found in *Eciton* (34.47%) and *Labidus* (62.80%). *Eciton* also contained Flavobacteriales (21.37%), while *Labidus* also contained Entomoplasmatales (19.24%).

The bacterial qPCR quantification analysis revealed no statistical difference in the bacterial quantification associated with ants across ant subfamily, genera, diet, nesting mode and habitat (Supplementary Figure 6; Habitat: t-test, $t=1.438$, $p\text{-value}=0.157$; Diet: ANOVA, $F=0.193$, $p\text{-value}=0.956$; Nesting mode: ANOVA, $F=0.128$, $p\text{-value}=0.957$; Subfamily: ANOVA, $F=0.400$, $p\text{-value}=0.813$; Genus: ANOVA, $F=0.310$, $p\text{-value}=0.968$). There was no statistical difference either in the forest/city dataset across ant habitat, diet, nesting mode subfamily and genus (Habitat: t-test, $t=1.345$,

p-value=0.194; Diet: ANOVA, $F=1.761$, p-value=0.106; Nesting mode: ANOVA, $F=1.823$, p-value=0.141; Subfamily: ANOVA, $F=0.331$, p-value=0.941; Genus: ANOVA, $F=1.043$, p-value=0.466).

Beta diversity analyses

The bacterial diversity of ants was compared with the bacterial diversity of their nest samples and both the Jaccard and unUniFrac distances showed statistical differences (Supplementary Figure 7A and B; Jaccard: pseudo- $F=2.967$, p-value=0.001; unUniFrac: pseudo- $F=8.162$, p-value=0.001). The forest/city dataset showed the same statistical differences (Jaccard: pseudo- $F=1.919$, p-value=0.001; unUniFrac: pseudo- $F=3.535$, p-value=0.001).

Next, we compared the bacterial diversity associated with the nests from the rainforest and the nests from the city, using PERMANOVA with Bray-Curtis and weighted UniFrac (wUniFrac) distances, because we were interested in the abundance and diversity of the bacteria species associated with each nest. The PERMANOVAs revealed statistical differences between rainforest nests and city nests (Supplementary Figure 7C and D; Bray-Curtis: pseudo- $F=1.463$, p-value=0.005; wUniFrac: pseudo- $F=1.237$, p-value=0.023). However, rainforest nests and city nests were not statistically different in the forest/city dataset (Bray-Curtis: pseudo- $F=1.061$, p-value=0.307; wUniFrac: pseudo- $F=0.655$, p-value=0.798).

Then we tested for dissimilarities in bacterial diversity across ant habitat, diet, nesting mode, subfamily and genus. All statistics for each beta diversity metric are in Supplementary File 4. There was no difference between rainforest ants or city ants (Bray-Curtis: Figure 14A, pseudo- $F=0.920$, p-value=0.619; wUniFrac: Figure 14B, pseudo- $F=0.945$, p-value=0.502)

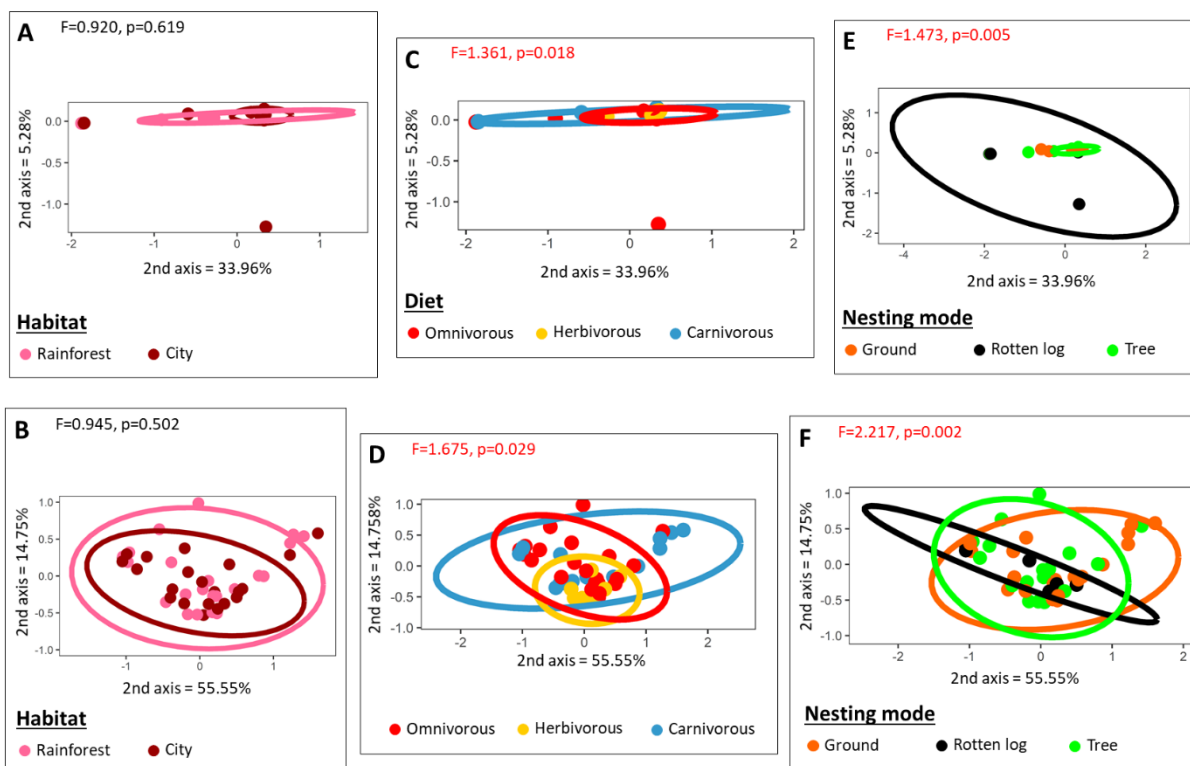


Figure 14. Differences in bacterial communities across ant habitat (A and B), diet (C and D) and nesting mode (E and F). The distances used were Bray-Curtis (A, C and E) or wUniFrac (B, D and F). The statistical p-values were obtained with pairwise PERMANOVAs. The circles represent a confidence interval of 0.95. P-values for this figure are found in Supplementary File 3.

However, the PERMANOVAs demonstrated statistical differences across ant diets and nesting modes for both the Bray-Curtis and wUniFrac distances (Bray-Curtis diet: Figure 14C, pseudo- $F=1.361$, p -value=0.018; wUniFrac diet: Figure 14D, pseudo- $F=1.675$, p -value=0.029; Bray-Curtis nesting mode: Figure 14E, pseudo- $F=1.473$, p -value=0.005; wUniFrac nesting mode: Figure 14F, pseudo- $F=2.217$, p -value=0.002). Across the diet, carnivorous were statistically different from herbivorous for both the Bray-Curtis and wUniFrac distances (Bray-Curtis diet: pseudo- $F=1.618$, p -value=0.012; wUniFrac diet: pseudo- $F=1.418$, p -value=0.046). Omnivorous were also statistically different from herbivorous but not from carnivorous with the Bray-Curtis distance (omnivorous-herbivorous: pseudo- $F=1.482$, p -value=0.007; omnivorous-carnivorous: pseudo- $F=1.096$, p -value=0.255), however there was no

difference in bacterial diversity between omnivorous and the other diets with the wUniFrac distance (omnivorous-herbivorous: pseudo-F=1.797, p-value=0.059; omnivorous-carnivorous: pseudo-F=1.751, p-value=0.062). Across the nesting mode, ants nesting in the ground were statistically different from ants nesting in rotten logs and ants nesting in trees for both the Bray-Curtis and wUniFrac distances (Bray-Curtis ground-rotten log: pseudo-F=1.659, p-value=0.016; Bray-Curtis ground-tree: pseudo-F=1.268, p-value=0.049; wUniFrac ground-rotten log: pseudo-F=2.533, p-value=0.006; wUniFrac ground-tree: pseudo-F=2.122, p-value=0.026). Ant nesting in rotten logs were also statistically different from ants nesting in trees with the Bray-Curtis distance (pseudo-F=1.601, p-value=0.028), but not with the wUniFrac distance (pseudo-F=1.996, p-value=0.057)

PERMANOVAs showed statistical differences across ant subfamilies and genera (Bray-Curtis subfamily: Figure 15A, pseudo-F=1.368, p-value=0.018; wUniFrac subfamily: Figure 15B, pseudo-F=1.715, p-value=0.025; Bray-Curtis genus: Figure 15C, pseudo-F=1.473, p-value=0.005; wUniFrac genus: Figure 15D, pseudo-F=2.217, p-value=0.002). Overall, our results showed that ants from the genera *Camponotus* (subfamily: Formicinae), *Cephalotes* (subfamily: Myrmicinae) and *Odontomachus* (subfamily: Ponerinae) are statistically different from the other ant genera. To test, whether these dissimilarities come from the ant evolutionary history, we performed Mantel test. However, the Mantel tests showed no correlation between bacterial diversity and the ant phylogeny for both the Bray-Curtis and wUniFrac distances (Figure 15E; Bray-Curtis: p-value=0.516; wUniFrac: p-value=0.498).

As we have seen previously in the alpha diversity analyses, the forest/city dataset followed the same pattern as the whole dataset for every factor for the two beta diversity metrics tested (Bray-Curtis habitat: pseudo-F=0.646, p-value=0.994; wUniFrac habitat: pseudo-F=0.515, p-value=0.962; Bray-Curtis diet: pseudo-F=1.578, p-value=0.008; wUniFrac diet: pseudo-F=1.299, p-value=0.145; Bray-Curtis nesting mode: pseudo-F=1.688, p-value=0.004; wUniFrac nesting mode: pseudo-F=1.641, p-value=0.021; Bray-Curtis subfamily: pseudo-F=1.368, p-value=0.018; wUniFrac subfamily: pseudo-F=1.715, p-value=0.025; Bray-Curtis genus: pseudo-F=1.473, p-value=0.005; wUniFrac genus: pseudo-F=2.217, p-value=0.002; Bray-Curtis phylogeny: p-value=0.681; wUniFrac phylogeny: p-value=0.732).

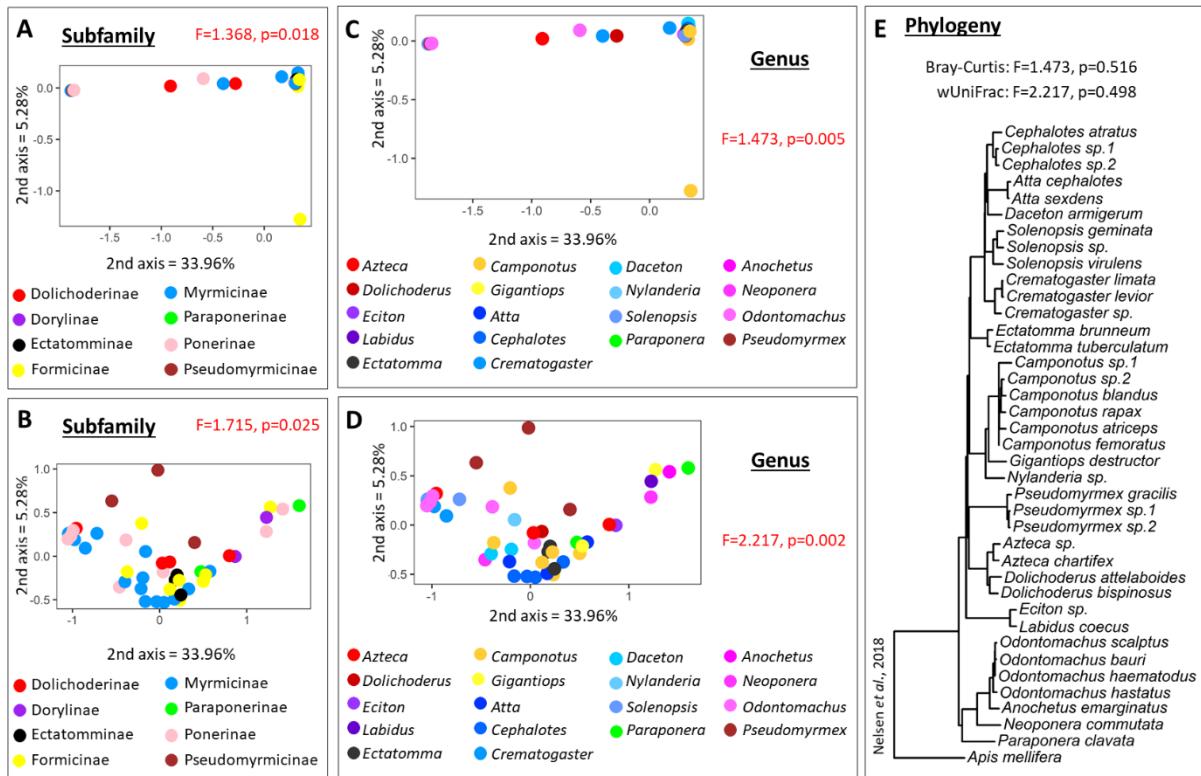


Figure 15. Differences in bacterial communities across ant subfamily (A and B), genus (C and D) and phylogeny (E). The distances used were Bray-Curtis (A, C and E) or wUniFrac (B, D and E). The statistical p-values were obtained with pairwise PERMANOVAs. P-values for this figure are found in Supplementary File 3.

ASV similarity percentage analyses

We used a SIMPER (Similarity Percentage) analysis to identify the most abundant bacterial orders responsible for the dissimilarities observed previously. This analysis identified 15 bacterial orders responsible for the differences. All statistics for SIMPER analyses are in Supplementary File 5 (whole dataset) and Supplementary File 6 (forest/city dataset).

Erysipelotrichales had an increased abundance in carnivorous and ground nesting ants compared to the other diets and nesting modes (Supplementary Figure 8A and B).

Rhizobiales had a higher abundance in herbivorous than in omnivorous ants, however there was no statistical difference in Rhizobiales abundance between omnivorous and carnivorous, and

between carnivorous and herbivorous (Supplementary Figure 8C). Rhizobiales had also a higher abundance in ground nesting than in rotten log nesting ants, however there was no statistical difference in Rhizobiales abundance between tree nesting and ground nesting, and between tree nesting and rotten log nesting (Supplementary Figure 8D).

Lactobacillales had an increased abundance in ground nesting ants than in rotten log nesting ants, but there was no statistical difference in Lactobacillales abundance between tree nesting and ground nesting, and between tree nesting and rotten log nesting (Supplementary Figure 9A). In the opposite, Rickettsiales had a higher abundance in rotten log nesting than in tree nesting ants, but there was no statistical difference in Rickettsiales abundance between ground nesting and tree nesting, and between ground nesting and rotten log nesting (Supplementary Figure 9B).

Entomoplasmatales had an increased abundance in carnivorous than in herbivorous ants, however, once again there was no statistical difference in Entomoplasmatales abundance between omnivorous and carnivorous, and between omnivorous and herbivorous (Supplementary Figure 10).

Finally, Acetobacterales had an increased abundance in city ants than in rainforest ants (Supplementary Figure 11).

There was no statistical difference in any bacterial orders between ant subfamilies and genera. Pairwise Kruskal-Wallis comparisons for all the discussed bacterial orders are in Supplementary File 7.

E. Discussion

Bacterial communities are omnipresent in nature and play major roles in every ecosystems on our planet, and many environmental factors affect their community structure¹⁷². Understanding which of these factors influence bacterial community structure is one of the key goals of microbial ecology²¹². Different studies have shown the impact of temperature¹⁷⁶, humidity¹⁸⁸, urbanization²¹³ and pH¹⁷⁵ on bacterial community structure. In host-associated bacteria, the host's biological factors influence bacterial community structure.

DNA amplicon sequencing of the 16S rRNA bacterial barcoding gene have been very useful to study the diversity and which ecological and evolutionary factors structure bacterial communities associated with various organisms^{71,194,214,215}. For example, in ants many studies have focused on the impact of host biology on their bacterial community^{164,165,201,216}. However, these studies usually focus on studying the bacterial community within one host genus or subfamily^{18,160,162,163}. Here, we studied the impact of the host habitat, diet, nesting mode and taxonomy on structuring bacterial communities on a wide phylogenetic diversity of ants from 18 ant genera in eight subfamilies.

Dietary niche structures bacterial communities associated with Amazonian ants

Symbiotic bacteria are known to play an important role in their host's diet. For example, they have been shown to complement different kinds of hosts nutritionally insufficient diet in ruminants⁵⁸, wood-feeding insects¹⁸³, blood-sucking insects and vertebrates^{214,217}, and sap-sucking insects¹⁸⁴. In ants, several nutritional symbioses have already been identified, in which specialized gut symbionts provide their host with essential amino acids synthesized via nitrogen recycling^{1,107,218}. Nutritional-provisioning bacteria have also been strongly associated with their host throughout evolution. Molecular phylogenetic studies have shown that the evolution of herbivory in ants is correlated to their symbiotic bacteria²¹⁹. On the other hand, phylogenetic analyses have demonstrated that multiple lineages of carnivorous army ants are associated with closely related specialized symbionts²²⁰.

Our results showed differences in bacterial diversity, but not in bacterial abundance, between carnivorous and herbivorous ants for both the Bray-Curtis and wUniFrac distances (Figure 14C and D ; Supplementary Figure 6B ; Supplementary File 4). Several comparative studies have also highlighted the major role of host dietary niches in structuring the gut bacterial community of their host, with specific diet specializations^{17,215,221,222}. However, we obtained contradictory results on the bacterial diversity associated omnivorous ants compared to other diets. Indeed, the Bray-Curtis distance indicated a difference between omnivorous and herbivorous, but this difference does not appear with the wUniFrac distance (Supplementary File 4). These differences can be explained by the fact that

when analyzing differences between two ASVs, the Bray-Curtis distance takes only into account if these two ASVs are different, while the wUniFrac distance takes into account if these two ASVs are different and phylogenetically closely related to each other. Our results could mean that the ASVs responsible for the difference between omnivorous and herbivorous have a high level of shared branch lengths, which would generate a difference with the Bray-Curtis distance but not with the wUniFrac distance.

The structure of the gut bacterial community is hypothesized to be determined by the abundance and diversity of nutrients extracted from the host diet during digestion. This theory, called the nutrient niche theory, was proposed by Rolf Freter in 1983 and posits that gut ecological niches are determined by available nutrients in the gut^{56,57,223}. This means that a specific bacteria species can only assert itself in the host gut if it is able to use a limiting nutrient. Since then, the nutrient niche theory has been supported by numerous diet supplementation studies, which have shown that the presence and abundance of specific bacterial organisms can be altered by experimentally modifying the type and abundance of nutrients present in the gut^{58,224–227}. In our study, we identified three bacterial orders which seem to have an impact on the differences in bacterial diversity associated with carnivorous and herbivorous ants: Entomoplasmatales, Erysipelotrichales and Rhizobiales (Supplementary Figure 8A and C ; Supplementary Figure 10). Our SIMPER analyses showed that Entomoplasmatales and Erysipelotrichales have a higher abundance in carnivorous than in omnivorous and herbivorous ants. Indeed, when we looked specifically at the bacterial relative abundance associated with each ant diet, we noticed that Erysipelotrichales and Entomoplasmatales are mostly present in carnivorous ants and punctually present in omnivorous ants, but never present in herbivorous ants. These observations are in accordance with the literature, as Erysipelotrichales have been found to be associated with carnivorous organisms in numerous studies and have been found to be increased by high-protein and high-fat diets^{228–230}. Moreover, Entomoplasmatales are known to be a major symbiont in the predatory army ants^{192,220} and have also recently been found in the predatory ant *Daceton armigerum*¹⁶⁵. Our SIMPER analyses also showed that Rhizobiales have a higher abundance in herbivorous than in omnivorous ants. When we looked specifically at the bacterial

relative abundance associated with each ant diet, we noticed that Rhizobiales are largely present in carnivorous and herbivorous ants, but present at a lower abundance in omnivorous ants. Some bacterial species within the Rhizobiales are nitrogen-fixing although this function has not been found in the bacterial species within ants, and their presence has been found to be correlated to herbivory in ants²¹⁹. Still, Rhizobiales have also been found in some carnivorous ants^{160,165,199}.

Overall, our results demonstrate strong differences in structuring bacterial communities associated with ant diet, which could suggest that ants with extreme diets, like strictly herbivorous and strictly predatory ants, may need to rely on symbiotic bacteria to complete their nutritionally insufficient diets²³¹. In return, nutrient availability derived from these extreme diets would allow only specific bacteria to colonize their host gut^{56,57}.

Bacterial community structure associated with Amazonian ants are also influenced by abiotic factors and nesting modes

Community structures can be influenced by multiple abiotic factors, resulting in different habitat selection. Studies of abiotic factors in rainforests have shown that environmental parameters vary between the canopy and the forest floor²³². For example, a three-year observation study of abiotic factors in a rainforest has shown that the ambient temperature is always a few degrees higher at the canopy level than at the ground level. In contrast, the ground benefits from a higher humidity level than the canopy²³³. Arthropods, as well as free-living bacteria, vertically structure their communities according to their different tolerance to these abiotic factors²³⁴.

Some symbiotic bacteria can confer their host an improved tolerance to environmental perturbations. For example in *Drosophila melanogaster*, some strains of *Wolbachia* have been shown to increase the heat tolerance of their host²³⁵. In pea aphids, the facultative symbionts *Serratia symbiotica* also protect their hosts from heat stress²³⁶. On the other hand, experimental studies on crickets have shown that the entire microbiome of the insect shifts to adapt to cold temperatures⁶⁷.

Symbiotic bacteria also allow their host to withstand other environmental perturbations than fluctuating temperatures, such as increase humidity²³⁷, or on the contrary increase dryness⁶⁸.

Overall, our results indicated differences in bacterial alpha and beta diversity, but not in bacterial abundance, between ants nesting in the ground and ants nesting in trees and in rotten logs (Figure 11; Figure 14E and F; Supplementary Figure 6C; Supplementary Files 2 and 4). As stated previously, in the rainforest, dryness level is higher in the canopy than on the ground. As such, to withstand this dryness organisms need increased desiccation resistance. In insects, the incorporation of hydrophobic aromatic compounds to their cuticle allow them to influence water evaporation by increasing the thickness and hydrophobicity of their cuticle²³⁸. These aromatic compounds incorporated are usually derived from amino acids in nutrients. Therefore, many herbivorous insects rely on bacteria to help producing these compounds, and increase the insect cuticle thickness and desiccation resistance^{189,239,240}. Moreover, studies have shown that desiccation resistance is higher in arboreal arthropods than in ground-dwelling arthropods^{241,242}. Desiccation resistance conferred by specific bacteria could therefore help explain the difference in bacterial diversities observed in our results between ants nesting on the ground and ants nesting in trees.

Apart from ground nests and arboreal nests, we also collected ants nesting in rotten logs. Rotten log nests are peculiar compared to arboreal nests and ground nests, because they are a very damp habitat and shelter mostly decomposers which feed on the log. In our results, ants nesting in rotten logs had a bacterial diversity different from ants nesting in the ground and ants nesting in trees with the Bray-Curtis distance (Supplementary File 4). However, with the wUniFrac distance, the bacterial diversity of ants nesting in rotten logs and ants nesting in trees was not statistically different (Supplementary File 4). This indicates that bacteria from ants living in rotten logs are different but phylogenetically closely related to bacteria from ants living in trees. When we looked specifically at which bacterial orders were found in ants nesting in rotten logs, we noticed that these ants had a bacterial diversity composed mainly of three bacterial orders adding up to almost 95% of bacterial

relative frequency in these ants. These three bacterial orders were Rickettsiales (41.62%), Enterobacteriales (28.69%) and Acetobacterales (24.53%) (Figure 13C; Supplementary File 3).

In our study, we identified four bacterial orders which seem to have an impact on the differences in bacterial diversity associated with ant nesting mode: Erysipelotrichales, Rhizobiales, Lactobacillales and Rickettsiales (Supplementary Figure 8B and D; Supplementary Figure 9A and B). Our SIMPER analyses showed that Erysipelotrichales have a higher abundance in ground nesting ants than in tree nesting and rotten log nesting ants. Interestingly, Erysipelotrichales are associated with carnivorous organisms and predatory ants like army ants and Ponerinae, which could explain the prevalence of Erysipelotrichales in these ground nesting ants. Our SIMPER analyses also showed that Rhizobiales and Lactobacillales have a higher abundance in ground nesting ants than in rotten log nesting ants, while Rickettsiales have a higher abundance in rotten logs nesting ants than in tree nesting ants. Rhizobiales can be associated to herbivorous and carnivorous ants. But Rhizobiales can also be found as free-living soil bacteria, possessing the ability to infect roots of leguminous plants²⁴³. All these reasons may explain their presence in ground nesting ants more frequently than in rotten log nesting ants. Lactobacillales are also commonly found in soil^{244,245}.

Overall, we show that different bacterial communities are associated with ground-dwelling ants and arboreal ants. This variation in diversity could be explained by abiotic factors between the soil and the canopy in rainforests, such as temperature and humidity. The bacterial communities is structured vertically²³⁴ and some association with bacteria could possibly allow arboreal ants to adapt to the canopy by enhancing their desiccation resistance²⁴².

Bacterial communities in long term association with specific ant hosts are conserved

Several ant genera have well-established long-term relationships with specific bacterial communities. These bacterial communities are usually conserved within the different species of their host genus and often provide benefits to the host. For example, *Camponotus* ants are well known to be associated to the obligatory endosymbiont *Blochmannia*, which has been shown to play roles in

Camponotus diet supplementation¹. On the other hand *Cephalotes* ants are associated with a core microbiome composed of five main bacterial orders, which recycle nitrogen from urea to synthesize essential and non-essential amino acids for their host¹⁰⁷.

Our results indicate differences in bacterial beta diversity, but not in bacterial abundance, between ant subfamilies and genera (Figure 15; Supplementary Figure 6D; Supplementary File 4). However, our Mantel test demonstrated that there is no correlation between the beta diversity distance matrices and the ant phylogeny (Figure 15E). Focusing on the differences in the beta diversity, we noticed that overall, three genera were different from the other ant genera: *Camponotus* (subfamily: Formicinae), *Cephalotes* (subfamily: Myrmicinae) and *Odontomachus* (subfamily: Ponerinae) (Supplementary File 4).

Looking deeper, we noticed that *Camponotus* are mainly composed of three bacterial orders, together adding up to more than 80% of relative bacterial abundance: Enterobacterales (45.19%), Acetobacterales (18.47%) and Rickettsiales (17.60%) (Figure 13A; Supplementary File 3). *Blochmannia* (order: Enterobacterales) are known to be the main symbiont of *Camponotus*. *Blochmannia* gives *Camponotus* nutritional supplementation, especially during host development^{246,247}. In our study, half of the Enterobacterales ASVs were identified as *Blochmannia* ASVs, and the other half were identified as undetermined Enterobacterales. However, *Blochmannia* are known to possess a fast rate of molecular evolution, therefore most of the undetermined Enterobacterales ASVs identified are also probably *Blochmannia* ASVs^{248,249}. *Blochmannia* has been shown to represent up to 95-98% of reads in *Camponotus chromaiodes*²⁵⁰. *Camponotus chromaiodes* have also been shown to be strongly associated with Acetobacterales²⁵⁰. In our study, one undetermined species of *Camponotus* also contained more than 90% reads of *Blochmannia*, and *Camponotus atriceps* (collected in two nests in our study) contained more than 90% reads of Acetobacterales and Enterobacterales added together (Figure 13A). Another common *Camponotus* symbiont is *Wolbachia* (order: Rickettsiales). *Wolbachia* is a common insect infectious symbiont and has beneficiary effect, like vitamin B supplementation in

bedbugs^{201,217}, as well as negative effects when manipulating host reproduction²⁵¹. In our study, all Rickettsiales ASVs associated with *Camponotus* were identified as *Wolbachia*.

Focusing on the *Cephalotes* samples, we found that they are mainly composed of 4 bacterial orders: Xanthomonadales (24.65% of bacterial relative abundance), Burkholderiales (22.71%), Rhizobiales (19.19%) and Opitutales (8.72%) (Figure 13A; Supplementary File 3). The *Cephalotes* microbiome has been extensively studied, and a core microbiome has been identified and shown to be very stable across *Cephalotes* species^{17,194,216,219}. The five bacterial orders composing the *Cephalotes* core microbiome are Burkholderiales, Opitutales, Pseudomonadales, Rhizobiales and Xanthomonadales, and together these symbionts synthesize amino acids via nitrogen-recycling for their host¹⁰⁷. In our study, we detected four of these bacterial orders in all *Cephalotes* samples. We detected very few Pseudomonadales ASVs in our *Cephalotes* samples, however Pseudomonadales are part of the Gammaproteobacteria, and we did detect several undetermined Gammaproteobacteria in all *Cephalotes* samples. Therefore, some of these Gammaproteobacteria could correspond to Pseudomonadales ASVs.

The last ant genus for which we found difference in beta diversity is *Odontomachus*. The main bacterial orders composing *Odontomachus* are Rickettsiales (72.87% of bacterial relative abundance, all corresponding to *Wolbachia* ASVs), Rhizobiales (11.60%) and Erysipelotrichales (6.66%) (Figure 13A; Supplementary File 3). Few studies have been done on the bacterial diversity associated with *Odontomachus* and one in particular focuses on *Wolbachia* symbionts²⁵². In our study we collected *Odontomachus haematodus* in three different nests, and in all of them we found a very high prevalence of *Wolbachia* (more than 90% of bacterial relative abundance). Rhizobiales in *Odontomachus* have also been identified in a study focusing on bacteria associated with the midgut of Ponerinae¹⁶⁰. Finally, as Erysipelotrichales is a common symbiont of carnivorous species^{228–230}, its presence in predatory *Odontomachus* samples was not surprising.

Even though *Camponotus*, *Cephalotes*, and *Odontomachus* were the only ant genera to have overall statistical differences in bacterial beta diversity compared to the other ant genera, we also

identified some ant genera which displayed a conserved bacterial diversity: *Crematogaster*, *Daceton*, *Eciton*, *Labidus*, *Nylanderia*, *Pseudomyrmex* and *Solenopsis* (Figure 13A; Supplementary File 3). However, these ant genera statistically have a lower alpha diversity compared to other ant genera (Figure 12; Supplementary File 2). In *Crematogaster*, *Pseudomyrmex* and *Solenopsis* the main bacterial symbiont found was *Wolbachia* (order: Rickettsiales) (Figure 13A; Supplementary File 3). Common associations between *Wolbachia* and these three ant genera have already been demonstrated in various studies^{159,252–254}, highlighting once again the high prevalence of *Wolbachia* in ants. In *Daceton*, the main bacterial order found were Rhizobiales (72.49% of bacterial relative abundance), Entomoplasmatales (13.20%) and Acetobacterales (11.31%). These findings support results from a recent study which also identified Rhizobiales and Entomoplasmatales as the main bacterial orders found in *Daceton*¹⁶⁵. In army ants (*Eciton* and *Labidus*), the main bacterial order found was Erysipelotrichales, which is as stated earlier a common symbiont in carnivorous species. Two studies identified Firmicutes and Entomoplasmatales as being the main symbionts in army ants^{192,220}. However, Erysipelotrichales belong to the phylum Firmicutes, meaning that some of the Firmicutes found in their studies may be Erysipelotrichales. In our study, Entomoplasmatales was also a major bacterial order in *Labidus*, but not in *Eciton*. Finally, in *Nylanderia* the main bacterial order was Acetobacterales (89.50%). However, as we collected only one species of *Nylanderia*, more studies on this ant genus are needed to determine if Acetobacterales is prevalent in all *Nylanderia* species.

In contrast, we also identified two ant genera which displayed very high species richness: *Ectatomma* and *Paraponera* (Figure 13A; Supplementary File 3). *Ectatomma* and *Paraponera* have statistically higher alpha diversity compared to other ant genera (Figure 12; Supplementary File 2). These two ant genera are characterized by not being associated with a core microbiome, but instead have transient bacterial communities. A few papers have already investigated the bacterial communities associated with *Paraponera clavata*^{255,256}. They found no clear pattern of a core microbiome, but they mention finding a small bacterial community which could be impacted by the ant diet. In our study, we did not find any of the bacterial genera mentioned in these papers, however

we did find the same bacterial phyla. These results confirm the hypothesis that ants hosting transient bacterial communities do not have any pattern of a core microbiome. Instead, they are associated with a high diversity of different bacteria at small abundance, and these communities seem to be impacted by the host environment, particularly their diet.

Overall, our results demonstrated that strong associations between ants and bacteria, sometimes from evolutionary long-term relationships, are common in various ant genera and very conserved within these specific genera. Additionally, ants involved in these associations often have low species richness, being composed of only a few bacterial orders at high relative abundance.

Habitat does not participate in structuring bacterial communities associated with Amazonian ants

As we have shown previously, bacterial communities associated with ants are structured by host diet and nesting mode. Yet, these ecological factors may vary in different habitats. Ants may need to adapt their feeding habits in particular, as different prey or plants may live in contrasting habitats. This process is known as environmental filtering, a process where the environment selects against certain species. For omnivorous ants or ants whose diet does not rely on a specific species, they could potentially be able to adapt faster and easier to the resources they will find in their new environment. Urbanization is known to be associated with a variety of effects on arthropods, like pollution, habitat fragmentation and decrease in species richness²⁵⁷. In her review, McIntyre regrouped arthropods in three categories: rural taxa which can only be found at high abundance in a rural environment, urban taxa which can only be found in an urban environment, and rural-urban taxa which can be found both in rural and urban environments²⁵⁸.

Environment filtering has been shown to also affects microbial communities. For example, a survey of airborne bacteria in different environments reported differences in bacterial densities between urban, rural and forest habitats²¹³. In another study, Belt *et al.* identified different bacteria in streams between urban and forested watershed²⁵⁹. In our study, we also shown that the bacterial

communities associated with ant nests are different in the rainforest and in the city (Supplementary Figure 7C and D).

For all these reasons we investigated the potential differences in bacterial communities associated with ants collected in the rainforest or in the city. In our study, we found nests of 10 ant species both in the primary forest and in an urban environment, allowing us to test for a potential habitat effect on bacterial communities associated with ants. However, we did not find any correlation between the bacterial communities from these ants in these contrasting habitats in term of diversity or abundance (Supplementary Figure 3; Supplementary Figure 6A; Figure 14A and B; Supplementary File 2 and 4). Our results corroborate previous work investigating the differences in microbial diversity of insects collected in different habitat which did not find any differences in the bacterial diversity between insects collected in urban environments and insects collected in rural environments^{260,261}.

We did find a statistical difference in the abundance of one bacterial order (Acetobacterales) between ants collected in the rainforest and ants collected in the city (Supplementary Figure 11; Supplementary File 3). However this result represent an exception in our study which support that environmental filtering does not seem to be affect the structure of ant bacterial communities.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

A.C., C.S.M., and C.D. conceived the study and collected the samples. A.C. performed the analyses. A.C., C.S.M., and C.D. interpreted the results. A.C. drafted the manuscript. A.C., C.S.M., and C.D. revised the manuscript.

Funding

This work received financial support from an “Investissement d’Avenir” grant managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-25-01) to C.D. and the National Science Foundation (NSF DEB 1900357) to C.S.M. We are grateful to the University of French Guiana for Anaïs Chanson’s Ph.D. fellowship and traveling grants.

Acknowledgments

The authors are grateful to Axel Touchard and Peter Flynn for their help in the collection and taxonomic identification of the ant samples. The authors are grateful to the Argonne National Laboratory (Lemont, Illinois, USA) for the 16S rRNA amplification and sequencing. The authors are thankful to the Cornell Institute of Biotechnology and the Cornell Information Technology support for their help in the setting up of the softwares used in this study. The authors are also thankful to Manuela Ramalho for all her assistance with statistical analysis.

Supplementary Informations

L’ensemble des figures et tableaux des *Supplementary Informations* sont en Annexe 1.

III. Conclusion

L’objectif du chapitre 1 était d’étudier les communautés bactériennes associées aux fourmis amazoniennes, et de déterminer quels facteurs taxonomiques et écologiques jouaient un rôle dans la structuration de ces communautés bactériennes. En utilisant la technique du métabarcoding, nous avons montré que les communautés bactériennes associées aux fourmis amazoniennes sont différentes en fonction du régime alimentaire de la fourmi, les fourmis herbivores ont des communautés bactériennes différentes des fourmis carnivores ou omnivores. Ces communautés bactériennes ont permis aux fourmis de devenir herbivores avec régime pauvre en azote et sont donc

généralement des symbiotes associés à ces fourmis. Les fourmis carnivores sont associées aux bactéries Erysipelotrichales qui sont communément associées à des régimes alimentaires riches en protéines²²⁸⁻²³⁰, et Entomoplasmatales qui pourraient intervenir dans la digestion de la chitine des insectes consommés⁶⁵.

Nous avons aussi montré que les communautés bactériennes sont différentes en fonction du type de nid de la fourmi ; les fourmis nichant dans le sol ayant des communautés bactériennes différentes des fourmis nichant dans les arbres. Ces communautés bactériennes pourraient être différentes car elles permettraient aux fourmis auxquelles elles sont associées de s'adapter aux conditions environnementales qui les entourent. Par exemple, certaines bactéries associées aux fourmis arboricoles participent au développement de la résistance de leur cuticule à la sécheresse^{189,242}. Les communautés bactériennes associées aux fourmis pourraient aussi être différentes en fonction du type de nid parce qu'elles pourraient être en partie acquise par l'environnement. Par exemple, les fourmis nichant au sol sont associées aux Lactobacillales et Rhizobiales qui sont des bactéries pouvant communément être trouvées dans le sol^{243,245}. En revanche, les communautés bactériennes n'étaient pas différentes en fonction de l'habitat dans lequel la fourmi a été collectée.

Nous avons également montré que les communautés bactériennes associées aux fourmis amazoniennes sont différentes en fonction de la taxonomie de la fourmi, mais pas en fonction de son histoire évolutive. En effet, certains genres de fourmis ont des communautés bactériennes très conservées et composées de seulement quelques ordres bactériens différents. C'est par exemple le cas des fourmis du genre *Camponotus*, *Cephalotes* et *Odontomachus*. Ces fourmis sont généralement associées à des communautés bactériennes ayant un rôle important dans la survie de leur hôte, ce qui pourrait expliquer leur conservation au sein de ces genres. C'est par exemple le cas de la bactérie *Blochmannia* qui intervient dans la production d'acides aminés chez la fourmi *Camponotus*^{1,195}, et le cas de bactéries intestinales spécifiques qui jouent un rôle dans le recyclage de l'azote en acides aminés chez la fourmi *Cephalotes*^{107,194}. À l'inverse, d'autres genres de fourmis ont des communautés bactériennes très diverses et composées de nombreux ordres bactériens différents, comme les fourmis

du genre *Azteca*, *Ectatomma* et *Paraponera*. Les communautés bactériennes associées à ces fourmis pourraient être principalement transitoires et leur présence dépendraient de l'environnement.

Chapitre II : Etude des communautés eucaryotes associées aux fourmis amazoniennes

I. Introduction

Au cours du chapitre 1, nous avons étudié les communautés bactériennes associées aux fourmis amazoniennes, et montré que ces communautés étaient structurées par la taxonomie de la fourmi, son régime alimentaire et son type de nid. Cependant, les fourmis peuvent également être associées à d'autres organismes que les bactéries.

En effet, plusieurs études ont révélé des associations entre fourmis et champignons. Un des exemples les plus étudiés est l'association mutualiste entre certaines fourmis Attines, principalement celles du genre *Atta* et *Acromyrmex*, et leur champignon basidiomycète (genre : *Leucoagaricus*)³. Les fourmis cultivent leur champignon mutualiste directement dans leur nid en le fertilisant avec des morceaux de plantes, puis utilise le champignon fertilisé comme nutriment. Il existe également des associations parasites entre fourmis et champignons. La classe de champignons Sordariomycetes contient notamment plusieurs espèces de champignons parasites des fourmis, comme le champignon *Ophiocordyceps* qui infecte de nombreuses espèces de fourmis et prend en partie contrôle de leur cerveau les transformant en ce qu'on appelle les « fourmis zombies »⁹. Outre leur association avec des champignons, certaines espèces de fourmis sont également associées à plusieurs familles de nématodes, comme les Mermithidae, les Rhabditidae, ou encore les Seuratidae²⁶².

De plus, il existe certaines associations tripartites entre insectes, bactéries et champignons/nématodes. Par exemple, reprenant l'exemple des fourmis Attines et de leur champignon mutualiste présenté dans le paragraphe précédent, ce champignon mutualiste peut être contaminé par un champignon parasite du genre *Escovopsis* (classe : Sordariomycete). Afin de limiter la contamination de leur champignon mutualiste, les fourmis Attines abritent sur leur cuticule une bactérie actinomycète qui est capable de produire un métabolite secondaire spécialisé pour lutter contre le champignon parasite *Escovopsis*²⁶³. Dans un autre exemple, certains nématodes

entomopathogènes forment une relation mutualiste avec les bactéries du genre *Xenorhabdus* et *Photorhabdus* pour tuer plus efficacement de nombreuses familles d'insectes²⁶⁴.

Cependant, malgré ces quelques exemples, les associations entre fourmis et les communautés eucaryotes d'une manière générale ne sont pas très répandues. En utilisant la méthode du métabarcoding présentée dans l'état de l'art, il est possible d'étudier la diversité des communautés eucaryotes associées à différents échantillons de fourmis grâce au marqueur taxonomique spécifique des eucaryotes : l'ARN ribosomique 18S. Cette méthode a été utilisée récemment pour étudier les communautés eucaryotes associées aux fourmis *Daceton armigerum*¹⁶⁵.

Pour ces différentes raisons, de la même façon que nous l'avons fait dans le chapitre 1 pour les communautés bactériennes, nous avons analysé la diversité et la structuration des communautés eucaryotes associées aux fourmis amazoniennes. Cette étude a deux objectifs principaux. D'une part, nous avons étudié si les communautés eucaryotes sont différentes en fonction de l'espèce de la fourmi, de son type de nid ou de l'habitat dans lequel elle a été collectée. D'autre part, l'analyse des communautés bactériennes du chapitre 1 ayant été réalisées sur les mêmes échantillons de fourmis que ceux utilisés dans ce chapitre, nous avons recherché de possible patrons de co-occurrence entre communautés bactériennes et communautés eucaryotes associées aux fourmis amazoniennes. Ce chapitre sera associé au chapitre 1 de cette thèse dans un article qui sera soumis fin juin 2021 pour être publié dans une édition spéciale « Microbiota in insects » dans le journal « Microorganisms » (ISSN 2076-2607).. C'est pourquoi ce chapitre 2 est rédigé en anglais dans ce manuscrit de thèse.

II. Potential impact of host nesting mode in structuring their associated eukaryote communities

Anaïs Chanson¹, Corrie S. Moreau², Christophe Duplais^{3*}

¹Université de Guyane, UMR8172 EcoFoG, AgroParisTech, CNRS, Cirad, INRAE, Université des Antilles, Kourou, France

²Departments of Entomology and Ecology & Evolutionary Biology, Cornell University, Ithaca, NY, USA

³CNRS, UMR8172 EcoFoG, AgroParisTech, Cirad, INRAE, Université des Antilles, Université de Guyane, Kourou, France

***Corresponding author**

Christophe Duplais

CNRS-UMR EcoFoG at the Pasteur Institute of French Guiana

23 avenue Pasteur, 97300 Cayenne, France

christophe.duplais@cirs.fr

Phone: +594 594 293 134

Keywords: Formicidae, insect, 18S rRNA, metabarcoding, nesting mode

A. Abstract

We previously shown that ant-associated bacterial communities are impacted by host diet, nesting mode and taxonomy. However, the functions and structure of eukaryote communities associated with hosts, and especially with ants, are still poorly understood. Here, we studied the diversity of eukaryote communities associated with Amazonian ants collected in the rainforest and in the city. We used the Amazonian ants collected previously and we used 18S rRNA amplicon sequencing to examine the impact of the host's ecology and phylogenetic factors in structuring their eukaryote communities. We investigated the roles of ant host habitat, nesting mode, diet and taxonomy on structuring ant eukaryote communities. Our results reveal that (1) only nesting mode structure eukaryote communities, (2) habitat, diet and host taxonomy does not structure eukaryote communities, and (3) there does not seem to be any co-occurrence patterns between bacterial and eukaryote communities. This work helps understanding how the host ecology structure its eukaryote communities.

B. Introduction

The global molecular analysis of eukaryote diversity based on amplicon sequencing has been studied in terrestrial habitats, mostly on soil^{265 266–268 269}, and in freshwater²⁷⁰ and in the ocean²⁷¹ for aquatic habitats. Yet, like bacteria, eukaryotes can be associated with various other organisms on the spectrum of symbiosis. In insects, many microbial eukaryotes including pathogenic fungi, nematodes and protists have been identified. Ascomycota fungi infect a wide range of insect hosts and after killing the insect often feed on its cadaver⁵⁹. Basidiomycota and Chytridiomycota fungi are known parasites of scale insects^{272,273} and insect eggs^{274,275}, whereas Zygomycetes occurs on various insects^{276–279}. Eukaryotes are known to be pathogens of ants. The fungal class Sordariomycetes contains several species infecting ants, like the pathogen fungus infecting fungus-growing ants²⁶³, or the fungus causing zombie ants⁹. Among the nematodes, many families have been shown to infect different ant species²⁶².

Concerning nematodes, several *Heterorhabditis* and *Steinernema* species have been reported to be pathogens of coleopterans²⁸⁰, dipterans²⁸¹, lepidopterans²⁸² and orthopterans²⁸³. Other nematodes also use insects as vector but does not necessarily kill it. It is the case for several *Brugia*^{284,285}, *Dirofilaria*²⁸⁶, *Onchocerca*²⁸⁷, and *Wuchereria*²⁸⁸ species, which use mosquitoes or midges as vectors. Finally, among the stramenophiles, oomycota are known to infect insects^{12,289}.

Mutualistic associations between eukaryotes and insects are also widespread especially between insects and fungus²⁹⁰. One of the most commonly found example is the cultivation of fungus for nutrition, which exists in several species of ants³, beetles⁵ and termites⁴. Another important aspect of mutualism between insects and fungus is the dispersal of fungi. Many insects even evolved specific organs to carry fungal spores^{60,291} or carry them in their gut²⁹². Finally, insects also benefit from antimicrobial molecules produced by some fungi⁷. Mutualistic associations between insects and nematodes are scarce but a few examples are known²⁹³, and some nematodes may mutualistic associations with bacteria within insect hosts²⁶⁴.

However, unlike for the study of insect-associated bacteria, the study of insect-associated eukaryote diversity is not as widespread. Indeed, researchers tend to study eukaryotes associated with insects only when the insect shows outward signs of infection. For example, a common sign of insect infection is when the nematode cut its way out of its host body. Another sign of fungal infection is behavioral and morphological changes such as climbing on the top of canopy or a change of color and size. However relying on the observational methods limits the study of the eukaryote diversity in insects which is more recently achieved by performing 18S rRNA amplicon sequencing.

In this study we examine the microbial eukaryote diversity associated with a wide phylogenetic range of Amazonian ant species. As a similar approach as the first chapter we recovered the eukaryote diversity associated with ant samples to determine whether an host factor is structuring eukaryote communities. Additionally, we analyzed the bacterial and eukaryote communities within the same ant

host for possible co-occurrence patterns between the bacterial and eukaryote communities. This work addresses the assembly patterns of eukaryote communities associated with ants in Amazonia.

C. Material and Methods

Sample collection

For this chapter, we used the 49 ant samples and blank controls collected during the first chapter. The list of collected species is detailed in Supplementary Table 1.

DNA extraction and amplification

We used the same DNA extraction protocol as the first chapter and we amplified the 18S rRNA gene specific to eukaryotes. Amplifications of the V1-V2 region of 18S rRNA and Miseq sequencing were performed by the Argonne National Laboratory (Lemont, Illinois, USA). Amplifications were performed using F04 and R22 primers as described in Creer et al.¹⁰², following the protocol suggested in the Earth Microbiome Project (EMP) (<http://www.earthmicrobiome.org/protocols-and-standards/16S/>). Each PCR reaction contained 12µL of DNA-free PCR water, 10 µL of 5 Prime HotMasterMix 1X, 1 µL of 5 mM forward primer, 1 µL of 5 mM Golay barcode tagged reverse primer, and 1 µL of extracted DNA. The amplification condition was as follows: 94°C for 3 min, with 35 cycles at 94°C for 45 s, 50°C for 60 s, and 72°C for 90 s, with a final cycle of 10 min at 72°C. Each PCR reaction was performed 3 times. 1% agarose gel electrophoresis was performed to confirm the efficiency of the amplification. Amplification samples from each dataset were pooled separately. Each pool containing 100 µL and were cleaned using the QIAquick PCR Purification Kit (Qiagen, USA) following the manufacturer's instructions. The molarity of the pool was determined and diluted down to 4 nM, denatured, and then diluted to a final concentration of 6.75 pM with a 10% PhiX for sequencing. The pooled were sequenced with the MiSeq Illumina V3 Reagent Kit 600 Cycles (300x300) using the custom sequencing primers and procedures described in the supplementary methods in Creer et al.¹⁰².

Eukaryote diversity

Demultiplexed sequences were analyzed using the Qiime2-2019.1¹⁰⁹ with the plugin demux (<https://github.com/qiime2/q2-demux>). Sequence quality control and feature table construction were performed using the dada2 plugin¹¹³. Taxonomy was assigned with the SILVA_132_QIIME database²⁰² and the ASV (amplicon sequence variants) were selected with 99% identity. To generate the taxonomy table, paired-end sequence reads were trimmed in the V1-V2 region of 18S rRNA with the F04/R22 primers. Our classifier was created using the "feature-classifier fit-classifier-naive-bayes" command, and the reads (rep-seqs) were classified by taxon using the "feature-classifier classify-sklearn" command²⁰³.

The filtration of contaminants was performed with the Decontam package¹²⁴ of the R software version 4.02²⁰⁴. In this package, the prevalence method was used to remove the contaminant sequences from our samples. The decontaminated table was inserted back into Qiime2¹⁰⁹ to filter all mitochondria and chloroplasts sequences. Hymenopteran sequences were also excluded from the table so that sequences coming from the ant own DNA would not appear in our analyses. The alignment was performed using the align-to-tree-mafft-fasttree command²⁰⁵ to reconstruct the microbial phylogeny.

The alpha and beta diversity analyses were performed using the "qiime diversity core-metrics-phylogenetic" command. Alpha and beta diversities were visualized using the visualization interface <https://view.qiime2.org/>. Alpha diversity metrics computed were the Shannon index¹⁴⁸, Faith's phylogenetic index¹⁴⁹, Pielou's evenness²⁶ and ASV richness¹¹⁸. Beta diversity metrics computed were Jaccard similarity index¹⁵¹, Bray-Curtis dissimilarity¹⁵², unweighted unifrac distance¹⁵³ and weighted unnormalized unifrac distance¹⁵⁴.

Bacteria-eukaryote composition correlations

In the previous chapter, we analyzed the bacterial composition associated with Neotropical ants. The DNA amplification of bacteria was performed on the same individuals as the DNA amplification of eukaryotes in this chapter. As such, we looked for specific correlations between bacterial and eukaryote ASVs. To this aim, we used the SCNIC package²⁹⁴ with the SparCC method²⁹⁵. The SparCC method log-transforms the quantity of the ASV in each sample of a specific dataset, then infer linear Pearson correlations between the different log-transformed ASVs. We first used the "qiime feature-table filter-features" command on Qiime2-2019.1¹⁰⁹ to remove all the bacterial and eukaryote ASVs which were found associated with only one of our sample. Next, we used the "between" function of the SCNIC package to test for correlations between our 16S rRNA dataset and our 18S rRNA dataset, and to create a correlation network. We choose a maximum p-value of 0.05, meaning that all the correlations with a p-value higher than 0.05 will not be taken into account in the resulting network. The network was visualized using the Cytoscape software version 3.7.2²⁹⁶.

Statistical analyses

Statistical analyses of alpha diversity were performed using the "diversity alpha-group-significance" command on Qiime2-2019.1¹⁰⁹, with the pairwise Kruskal-Wallis method. The results were visualized using PCoAs with the R packages "ggpubr" version 0.3.0²⁰⁷ and "PMCMR" version 4.3²⁰⁸. Statistical analyses of beta diversity were performed using the "diversity beta-group-significance" command on Qiime2-2019.1¹⁰⁹, with the pairwise permanova method and 999 permutations. The results were visualized using PCoAs with the R packages "ggplot2" version 3.3.0²⁰⁹ and "ggfortify" version 0.4.10²¹⁰. The correlation between ant evolutionary history and eukaryote composition was performed with a Mantel test using the R package phytools²¹¹. The contribution of each ASV at the phyla level for each sample was obtained using the "qiime taxa barplot" and "qiime

taxa collapse” commands on Qiime2-2019.1¹⁰⁹. The SIMPER (Similarity Percentage) analyses were performed using the PAST software version 4.02²⁰⁶ and visualized using boxplots with the R packages “ggpubr” version 0.3.0²⁰⁷ and “PMCMR” version 4.3²⁰⁸.

D. Results

Assessing the sequencing quality

To assess the eukaryote diversity associated with Amazonian ants, a total of 51 samples were sequenced: 49 ant samples and 2 negative controls (Supplementary File 8). In the raw dataset, there were a total of 21,719 reads. After rarefaction of the samples to 199 reads, 33 samples were excluded from the dataset because they did not reach the minimum sampling depth (Supplementary Figure 12A). In total, in the 16 remaining samples we obtained 118 ASVs from 20,185 reads ranging from 199 to 7,815 reads, with a mean frequency of 1,261 reads per sample.

Alpha diversity analyses

The eukaryote alpha diversity associated with Amazonian ants was tested with the following metrics: Shannon diversity, Pielou’s evenness, Faith’s phylogenetic diversity and ASV richness. Each metric was calculated with Kruskal-Wallis ANOVAs, and statistics are in the Supplementary File 9.

There was no statistical difference between rainforest ants and city ants for any metric tested (Supplementary Figure 13; Shannon diversity: $H=0.540$, $p\text{-value}=0.462$; ASV richness: $H=1.773$, $p\text{-value}=0.183$; Faith’s PD: $H=1.103$, $p\text{-value}=0.294$; Pielou’s evenness: $H=0.011$, $p\text{-value}=0.916$). No statistical difference was found in any of the metrics tested between ant subfamilies (Supplementary Figure 14; Shannon diversity: $H=4.352$, $p\text{-value}=0.360$; ASV: $H=2.581$, $p\text{-value}=0.630$; Faith’s PD: $H=3.643$, $p\text{-value}=0.456$; Pielou’s evenness: $H=5.993$, $p\text{-value}=0.200$).

The Shannon diversity, ASV richness and Faith’s PD were higher in ground nesting ants than in rotten log nesting ants, even though there was no overall statistical difference between nesting modes

for the Shannon diversity (Shannon diversity: Figure 16A, $H=3.824$, $p\text{-value}=0.148$; ASV richness: Figure 16B, $H=4.634$, $p\text{-value}=0.099$; Faith's PD: Figure 16C, $H=3.107$, $p\text{-value}=0.212$). There was also no difference between nesting modes for the Pielou's evenness (Figure 16D, $H=1.283$, $p\text{-value}=0.526$).

Overall no difference was found in any of the metrics between ant genera (Shannon diversity: $H=12.353$, $p\text{-value}=0.136$; ASV richness: $H=10.424$, $p\text{-value}=0.237$; Faith's PD: $H=9.213$, $p\text{-value}=0.325$; Pielou's evenness: $H=12.090$, $p\text{-value}=0.147$). Yet, the Shannon diversity and Pielou's evenness were statistically higher in *Cephalotes* than in *Camponotus* (Supplementary Figure 15A and 15D).

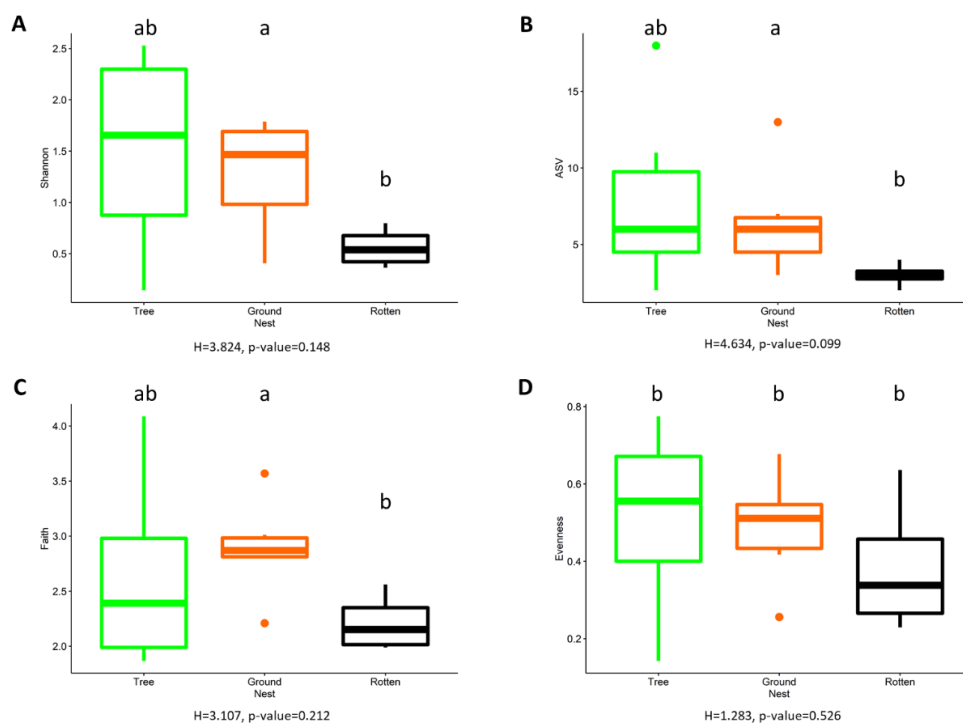


Figure 16. Differences in eukaryote alpha diversity by ant nesting mode. (A) Mean (\pm standard error) Shannon diversity. (B) Mean (\pm standard error) ASV count. (C) Mean (\pm standard error) Faith's phylogenetic diversity. (D) mean (\pm standard error) Pielou's evenness. Different letters at the top of the figure illustrate nesting mode with significant differences ($P < 0.05$) in this alpha diversity metric. The statistical p -values were obtained with pairwise Kruskal-Wallis. P -values for this figure are found in Supplementary File 9.

Beta diversity analyses

The beta diversity analyses were also performed using the software Qiime2 to test for dissimilarities in eukaryote diversity in our dataset. Statistics for the different beta diversity metrics tested (Bray-Curtis distance and wUniFrac distance) are in Supplementary File 10.

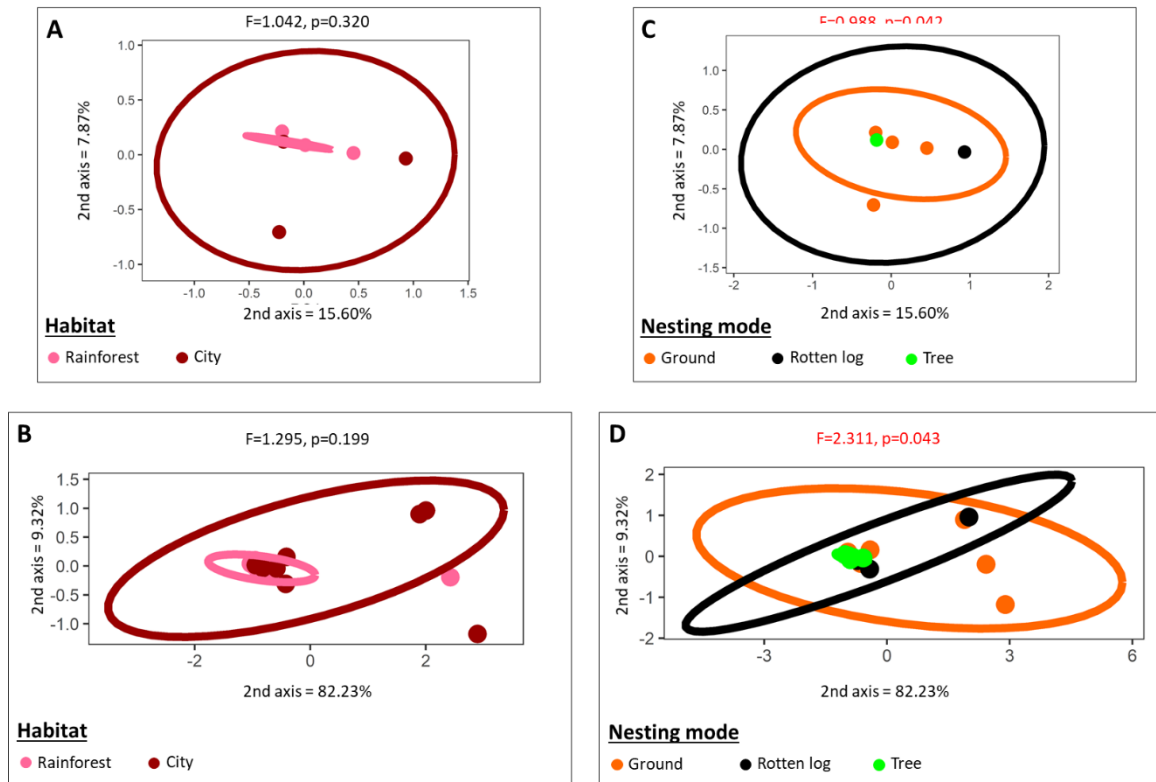


Figure 17. Differences in eukaryote communities across ant habitat (A and B) and nesting mode (C and D). The distances used were Bray-Curtis (A and C) or wUniFrac (B and D). The statistical p-values were obtained with pairwise PERMANOVAs. P-values for this figure are found in Supplementary File 10.

For the two metrics tested, there was no difference between the ant habitats (Bray-Curtis: Figure 17A, pseudo-F=1.042, p-value=0.320; wUniFrac: Figure 17B, pseudo-F=1.295, p-value=0.199), subfamilies (Bray-Curtis: Figure 18A, pseudo-F=1.043, p-value=0.245; wUniFrac: Figure 18B, pseudo-F=0.718, p-value=0.652), and genera (Bray-Curtis: Figure 18C, pseudo-F=1.121, p-value=0.103; wUniFrac: Figure 18D, pseudo-F=1.686, p-value=0.097). There was also no correlation between the eukaryote diversity and the ant evolutionary history, as shown by the Mantel test (Figure 18E ; Bray-

Curtis: $F=0.074$, $p\text{-value}=0.227$; wUniFrac: $F=-0.064$, $p\text{-value}=0.603$). However, there was a statistical difference between the nesting modes for both metric tested (Bray-Curtis: Figure 17C, pseudo- $F=0.988$, $p\text{-value}=0.042$; wUniFrac: Figure 17D, pseudo- $F=2.311$, $p\text{-value}=0.043$). Indeed, for these metrics, ants nesting in the ground are statistically different than ants nesting in trees. However, ants nesting in rotten logs are not statistically different from ants nesting in the ground or in trees.

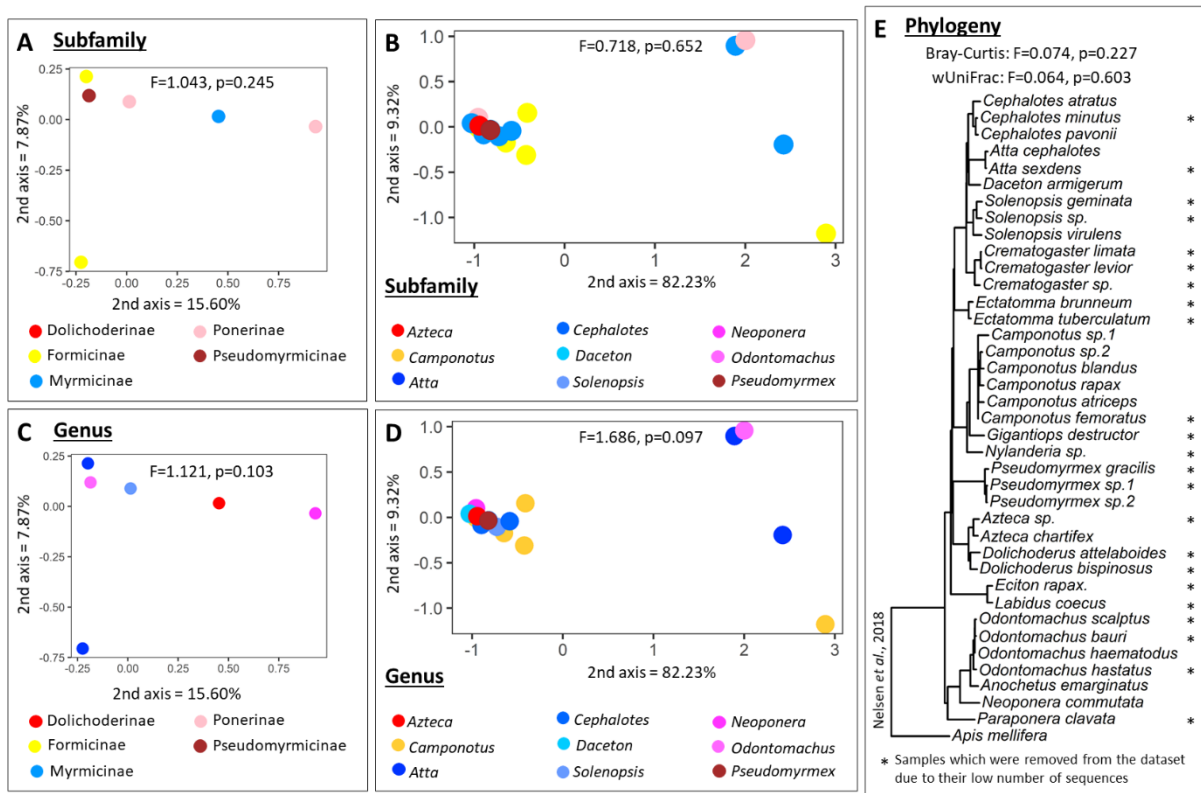


Figure 18. Differences in eukaryote communities across ant subfamily (A and B), genus (C and D) and phylogeny (E). The distances used were Bray-Curtis (A, C and E) or wUniFrac (B, D and E). The statistical p-values were obtained with pairwise PERMANOVAs. P-values for this figure are found in Supplementary File 10.

ASV similarity percentage analyses

To investigate which eukaryote subphyla are the most associated with Amazonian ants, we used SIMPER analyses. The taxa barplot of total eukaryote relative abundance was largely dominated by undetermined eukaryotes in every factor tested (Figure 19, Supplementary File 11). Overall, the 11 most abundant eukaryote subphyla were undetermined eukaryotes, Nematoda, Eugregarinorida, Chytridiomycota, Mortierellomycotina, Saccharomycotina, Ustilaginomycotina, Pezizomycotina, Basidiobolomycetes, Mucoromycotina and Arthropoda.

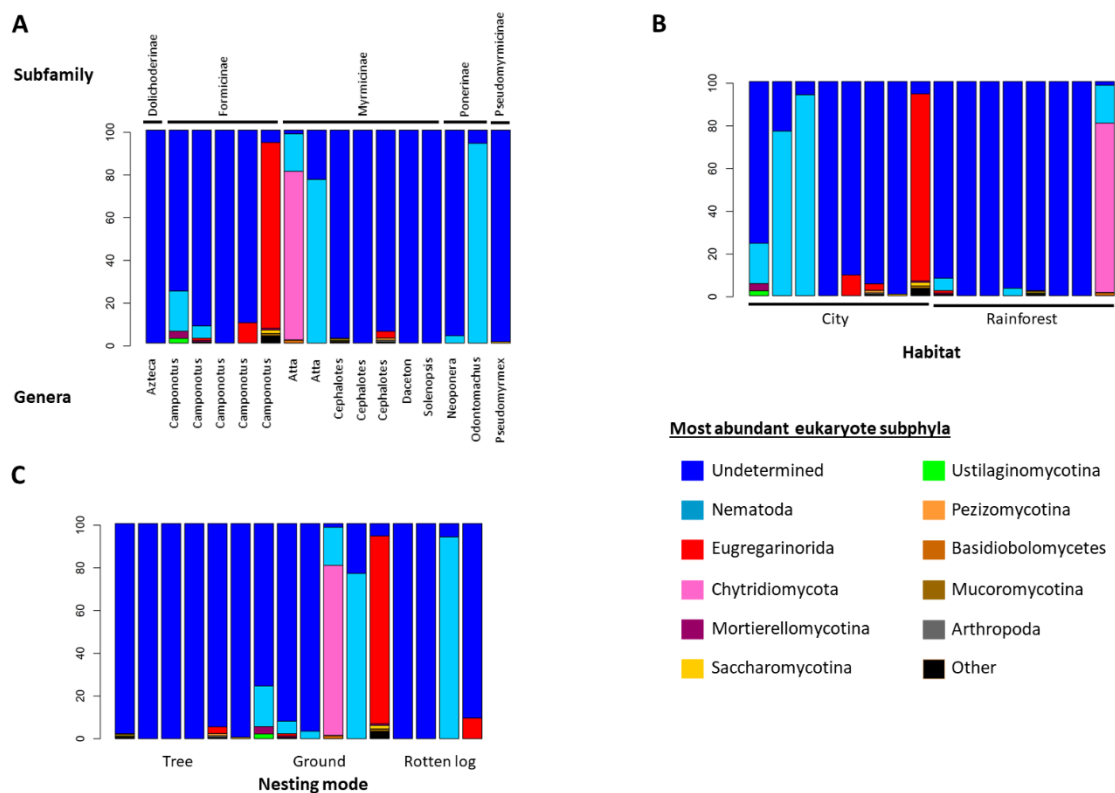


Figure 19. Taxa barplot based on relative abundance ordered by ant subfamily and genus (A), habitat (B) and nesting mode (C). The relative abundance bars are colored by percentage of eukaryote subphyla. The 11 most abundant eukaryote subphyla associated with Amazonian ants are listed. The less abundant eukaryote subphyla are grouped under the term “Other”. The term “Undetermined” represent eukaryote ASVs which could not be identified with the SILVA taxonomy.

Focusing on the different ant habitats and nesting modes, undetermined eukaryotes represent half of the total eukaryote relative abundance in city ants and ground nesting ants (49.10% and 61.90% respectively), while they represent almost 75% of the relative abundance in ants nesting in rotten logs, and more than 98% in rainforest ants and ants nesting in trees. Nematoda are also very common in city ants and ground nesting or rotten logs nesting ants (23.70%, 20.40% and 23.40% respectively). Eugregarinorida are present in city ants and in ground nesting ants (14.70% and 12.50% respectively), while Chytridiomycota are only present in ground nesting ants.

The subfamilies Dolichoderinae and Pseudomyrmicinae are entirely associated with undetermined eukaryotes. Formicinae and Myrmicinae also contain mainly undetermined eukaryotes (72.70% and 73.90% respectively), but Formicinae also contain Eugregarinorida (19.60%) while Myrmicinae contain Nematoda and Chytridiomycota (13.50% and 11.30% respectively). Ponerinae are composed of undetermined eukaryotes and Nematoda (51.40% and 48.60% respectively). At the host genus level, in most of the tested genera (*Azteca*, *Cephalotes*, *Daceton*, *Solenopsis*, *Neoponera* and *Pseudomyrmex*), undetermined eukaryotes represent more than 95% of total eukaryote relative abundance. *Camponotus* are also largely dominated by undetermined eukaryotes (72.70%), but contain also 20% of Eugregarinorida. In contrast, *Atta* and *Odontomachus* contain very few undetermined eukaryotes (respectively 12.50% and 6.220%). Instead, *Atta* consist primarily of Nematoda and Chytridiomycota (respectively 47.20% and 39.50%), while *Odontomachus* are largely composed of Nematoda (93.80%).

We used a SIMPER (Similarity Percentage) analysis to obtain the contribution of the 11 most abundant eukaryote subphyla to the dissimilarities observed previously. All statistics for SIMPER analyses are in Supplementary File 11. There was no statistical difference in any eukaryote subphyla between ant habitats, subfamilies and genera. However, there was statistical difference in some eukaryote subphyla between nesting modes. Indeed, undetermined eukaryotes had an increased relative abundance in tree nesting ants compared to ground nesting ants. In contrast, Nematoda had

an increased relative abundance in ground nesting ants compared to tree nesting ants (Supplementary Figure 16A and B). Pairwise Kruskal-Wallis comparisons for these are in Supplementary File 13.

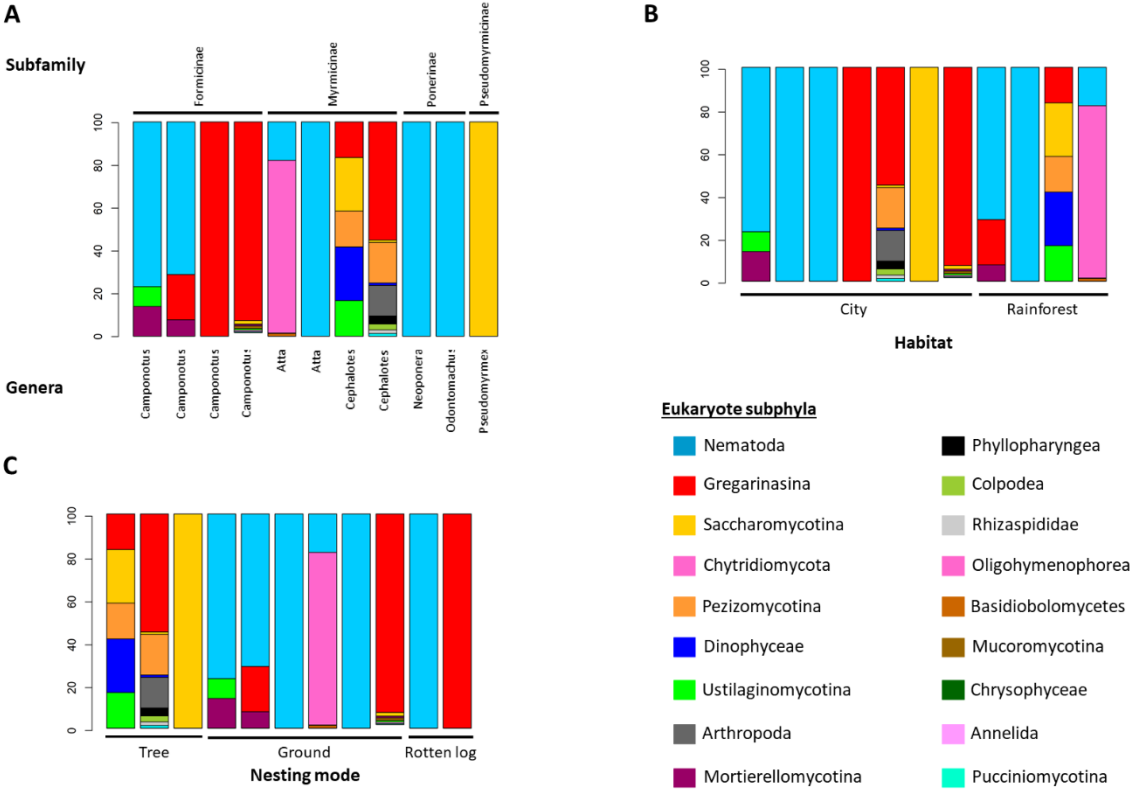


Figure 20. Taxa barplot based on relative abundance ordered by ant subfamily and genus (A), habitat (B) and nesting mode (C), after removing all the undetermined ASVs. The relative abundance bars are colored by percentage of eukaryote subphyla. A description of each eukaryote subphylum and examples of genera associated to each subphylum can be found in Supplementary File 14.

Due to the high amount of undetermined ASVs in our samples, we decided to remove these sequences to have a better representation of identified eukaryotes subphyla results (Figure 20, Supplementary File 12). The most abundant eukaryote subphyla are Nematoda and Gregarinasina. As shown before, Nematoda are more abundant in ground nesting ants than in tree nesting ants (Figure 20C) and they also are abundant in *Camponotus*, *Atta*, *Neoponera* and *Odontomachus* (Figure 20A). Gregarinasina are common in *Camponotus* and *Cephalotes* (Figure 20A). Saccharomycotina and

Chytridiomycota are present in only one sample each (*Pseudomyrmex* and *Atta* respectively) (Figure 20A). The majority of genera are associated with only a few different eukaryote subphyla, with one subphylum being predominant in the sample and the other subphyla having a small abundance (Figure 20A). However, *Cephalotes* is the only genera where the samples are associated with at least five different eukaryote subphyla, with no subphylum being dominant but instead the different eukaryote subphyla are present in similar relative abundances. A description of each eukaryote subphylum and examples of genera associated to each subphylum can be found in Supplementary File 14.

Correlation network between 16S rRNA and 18S rRNA ASVs

To determine if some co-occurrences are found between bacterial and eukaryote sequences associated with Amazonian ants, we used the package SCNIC. In the network, only correlations between two ASVs with a p-value lower than 0.05 are taken into account. In the correlation network generated three groups appeared (Supplementary Figure 17). In the first group, one eukaryote ASV (Nematoda) co-occurs with five bacterial ASVs (two Corynebacteriales, one Cytophagales, one Rhizobiales and one undetermined). In the second group, two eukaryote ASVs (Eugregarinorida) co-occur with four bacterial ASVs (one Acetobacteriales, one Enterobacteriales, one Rhizobiales and one undetermined). Finally, in the third group, one eukaryote ASV (Nematoda) co-occurs with one bacterial ASV (Pseudomonadales). In the first and second groups the co-occurrences were found in four samples of *Camponotus* ants (two samples for each group), while in the third group the co-occurrence is associated with two samples of *Atta cephalotes* and one sample of *Odontomachus haematodus*. Statistics for the correlation network with a p-value lower than 0.05 are in Supplementary File 15.

E. Discussion

Soils are among the most taxon-rich of any microbial habitat²⁹⁷, and global eukaryote soil diversity studies have been made on a wide range of different habitats^{265–269}. Eukaryotic

microorganisms can also be associated with various arthropods, and entomopathogenic eukaryotes studies are abundantly used for pest managements^{10,298}. However, the factors structuring the presence of eukaryotes associated with insect hosts are still mostly unknown. Here, we used 18S rRNA amplicon sequencing to assess the diversity of eukaryote communities associated with Amazonian ants, and to determine which host ecological and evolutionary factors are structuring the eukaryote communities.

Prevalence of undetermined eukaryotes associated with Amazonian ants

The low numbers of currently available 18S rRNA sequences in public databases currently limit the study of eukaryotes. In our study the majority of the retrieved 18S rRNA sequences, after excluding any sequence identified as “Hymenoptera”, could not be identified further than being eukaryotes (Figure 19). And among the determined eukaryote ASVs the identification accuracy was variable with only a few ASVs assigned at the species level, while other AVSs could only be identified at the order or kingdom level. This issue has already been discussed in previous molecular studies²⁶⁶.

One of the possible ways to address this issue is to use multiple PCR-primers. Different studies compared the use of universal eukaryotic primers versus the use of multiple PCR-primers targeting specific eukaryote groups. The results show that between 50% and 75% of the global eukaryote diversity in soil is lost when using only universal eukaryotic primers^{103,104}. This multiple PCR-primers approach would be helpful in particular to retrieve sequences from less abundant eukaryote groups^{104,299}. Moreover, all public databases are not equal in terms of numbers of sequences identified, and the use of the multiple PCR-primers would allow us to use different databases, increasing the possibility to obtain more precise taxonomic assignment for the different ASVs retrieved. A study by Bridge et al. has also shown that even some of the publicly available sequences may not be reliable, demonstrating again the need to strengthen the number and quality of sequences in the databases by the multiplication of eukaryote diversity studies and the use of multiple PCR-primers approach³⁰⁰.

In our study, we retrieved more undetermined eukaryotes in tree nesting ants than in ground or rotten log nesting ants (Figure 19C; Supplementary Figure 16A) and we retrieved more undetermined eukaryotes in rainforest ants than in city ants (Figure 19B). The fact that we seem to have more undetermined eukaryotes in tree nesting ants and in rainforest ants could come from the fact that there is a greater number of studies made on eukaryote diversity from soils^{266–268,297} and from urban areas^{104,299,301,302}, which could cause a lack of sequences from eukaryotes found primarily in the canopy or in the rainforest. If this is the case, more studies on global eukaryote diversity from trees and arboreal species would be necessary in order to add more sequences in eukaryote database.

Detecting known ant-associated eukaryotes

Currently the associations between ants and eukaryotes remain mostly unknown. Yet some studies have shown the associations (mutualistic or parasitic) between a few ant genera and specific eukaryotes from different phyla.

For example, a few fungi are known to be associated with ants. The black yeast Chaetothyriales (Ascomycota) can be found in the nests of carton nesting and domatia nesting ants³⁰³, or on the cuticle of fungus-growing ants²³¹. In our data, sequences from Chaetothyriales were found only in one sample of *Cephalotes atratus* collected in the city. Our dataset contains one carton nesting species (*Azteca chartifex*), but as only undetermined sequences were retrieved in this sample, the presence/absence of Chaetothyriales remains unknown. Fungus-growing ants also harbor mutualistic fungi in their nest (class: Agaricomycetes) and may be associated to different pathogenic fungi (class: Sordariomycetes), which can infect the ant or their mutualistic fungus^{3,263}. Several ants can also be infected by the parasitic zombie ant fungus *Ophiocordyceps* (class: Sordariomycetes)⁹. However, no Sordariomycetes or Agaricomycetes sequence were retrieved in any of our samples.

In the Protostomia phylum, some associations with ants have also been determined and a notorious example is the nematode *Myrmeconema* (class: Enoplea) which infects the ant *Cephalotes*

atratus^{304,305} and induces fruit mimicry in the tropical ant. No Enoplea sequences were retrieved in any of our samples. However the absence of *Myrmeconema* infection in our *Cephalotes atratus* samples could have been predicted since none of the collected workers showed the usual morphological characteristics of the infection (red and swollen gaster). The trematode *Dicrocoelium dendriticum* is also associated with several ant species as intermediate hosts before being consumed by ruminants³⁰⁶ however we did not detected sequences of this eukaryote³⁰⁷.

Recently, a study by Ramalho et al. investigated the eukaryote diversity associated with the Amazonian trap-jaw ant *Daceton armigerum*¹⁶⁵. They also retrieved a majority of undetermined eukaryotes along with few taxonomically assigned sequences. By comparing results we found some overlap of the eukaryote phyla retrieved such as algae, arachnids, Colpodea, fungi, and nematodes and differences as in our study we did not recover Coccidia, birds, fishes or insects sequences.

Diverse groups of eukaryotes are associated with Amazonian ants

The majority of determined eukaryotes found associated with Amazonian ants are from three different eukaryote groups: the Alveolata (protists), the Nucleomyces (fungi) and the Holozoa (animals) (Supplementary File 12). The total abundance of alveolates or fungi associated with ants is higher than the total abundance of microscopic animals. Additionally these three major eukaryote groups are more abundant in ground nesting ants than in tree nesting ants (Supplementary File 12). However these results might be a misrepresentation as the majority of the sequences associated with tree nesting ants remained undetermined (Figure 19C).

Alveolate protists are the most abundant eukaryote group associated with our Amazonian ants samples. Other molecular analysis studies on eukaryote diversity in diverse environments have also shown alveolates to be one of the most diverse and abundant groups retrieved^{265,267,268,270,271}. Among the alveolates, the most abundant phylum in our samples was by far the apicomplexans. These results are not surprising since ground nesting invertebrates are known to be associated with

apicomplexans^{265,308}. In our results, apicomplexans were only found associated to *Camponotus* and *Cephalotes* (Figure 20A)

Among the fungi found the most abundant was Chytridiomycota, but it was only found in one *Atta* sample (Supplementary File 12). Some members of this phylum are known to occur on the scales of insects¹⁰ and are of high concern since they infect frogs causing a global threat to species diversity. Apart from Chytridiomycota, the other fungi phyla detected were Ascomycota (Pezizomycotina and Saccharomycotina), Basidiomycota (Pucciniomycotina and Ustilaginomycotina), Mucoromycota (Mortierellomycotina and Mucoromycotina) and Zoopagomycota (Basidiobolomycetes). These different fungi phyla are commonly found in soil studies^{266,267,301,302}, and play an important role in the nutrient cycles soil ecosystems^{301,302}. But some members of these different phyla are also known insect-associated fungi. Pezizomycotina and Saccharomycotina are closely associated with insects^{10,309,310}. In particular, Saccharomycotina provide many nutritional resources for their insect host, and rely on their hosts for dispersion³¹¹. Pucciniomycotina and Ustilaginomycotina contain also a large number of members closely associated with insects¹⁰. Mortierellomycotina and Mucoromycotina are primarily found in soil, but some of their members have been reported as being weak insect parasites¹⁰. Basidiobolomycetes contains also many entomopathogens¹⁰. In our results, Ascomycota were found primarily associated with tree nesting species, and in particular with the *Cephalotes* genus (Figure 20A and C; Supplementary File 12). In contrast, Basidiomycota, Mucoromycota and Zoopagomycota were found almost exclusively associated with ground nesting *Camponotus* ants (Figure 20A and C).

In our results, nematodes and annelids are exclusively associated with ground nesting ants, while parasitic arthropods are exclusively associated with tree nesting *Cephalotes* ants (Figure 20A and C; Supplementary File 12). The nematodes associated with Amazonian ants are from the class Chromadorea, and about half are from the order Rhabditida, which contains mostly parasitic nematodes^{312–314}. The arthropods associated with Amazonian ants are from two species of Acari: *Achiptera coleoprata* and *Sarcoptes scabiei*.

Bacteria and eukaryotes co-occurrence network

In the previous chapter, we demonstrated that bacterial communities associated with Amazonian ants were structured by the host diet, nesting mode and taxonomy. In this chapter, our results show that eukaryote communities associated with the Amazonian ants were structured by the host nesting mode. Since both studies were conducted on the same ant samples, we used a network correlation analysis to test for co-occurrence between the bacterial and eukaryote communities.

Co-occurrence correlation networks can be implemented for different questions and are often used in bacterial communities' studies. For example, they are used in medical studies to compare the bacterial co-occurrence of healthy people and people subject to specific diseases^{315–317}. They have also been used in ecological studies to compare the bacterial composition between different biomes³¹⁸. But they can also be used by combining different datasets to compare the co-occurrence of bacteria with specific metabolites or diet³¹⁹, or the co-occurrence of bacteria with archaea³²⁰.

Here we chose to use the SCNIC software²⁹⁴ to construct our network, because it allows us to directly use the biome table retrieved in Qiime2, and we can use it to test for co-occurrence between one or between two datasets. However, to construct a co-occurrence network, we need microbial ASVs which occur in more than one sample. In our study, only four eukaryote ASVs were found in at least two different samples, making the search for co-occurrence patterns difficult. In the co-occurrence network, bacterial ASVs present seem to be structured by the host taxonomy and not by the presence of a specific eukaryote ASV. Indeed, among the four eukaryote ASVs present, three were found only in *Camponotus* ants and are co-occurring with bacterial orders which have been demonstrated in the first chapter to be common in ground nesting *Camponotus* ants (Acetobacterales, Enterobacterales and Rhizobiales). As such, to determine co-occurrence patterns between bacterial and eukaryote communities, future studies need to retrieve more 18S rRNA ASVs sequences, which is currently the main limitations to such studies.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

A.C., C.S.M., and C.D. conceived the study and collected the samples. A.C. performed the analyses. A.C., C.S.M., and C.D. interpreted the results. A.C. drafted the manuscript. A.C., C.S.M., and C.D. revised the manuscript.

Funding

This work received financial support from an “Investissement d’Avenir” grant managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-25-01) to C.D. and the National Science Foundation (NSF DEB 1900357) to C.S.M. We are grateful to the University of French Guiana for Anaïs Chanson’s Ph.D. fellowship and traveling grants.

Acknowledgments

The authors are grateful to Axel Touchard and Peter Flynn for their help in the collection and taxonomic identification of the ant samples. The authors are grateful to the Argonne National Laboratory (Lemont, Illinois, USA) for the 16S rRNA amplification and sequencing. The authors are thankful to the Cornell Institute of Biotechnology and the Cornell Information Technology support for their help in the setting up of the softwares used in this study. The authors are also thankful to Manuela Ramalho for all her assistance with statistical analysis.

Supplementary Informations

L’ensemble des figures et tableaux des *Supplementary Informations* sont en Annexe 2.

III. Conclusion

Les objectifs du chapitre 2 étaient d'une part d'étudier les communautés eucaryotes associées aux fourmis amazoniennes et de déterminer quels facteurs taxonomiques et écologiques jouaient un rôle dans la structuration des communautés eucaryotes, et d'autre part d'analyser les potentiels patrons de co-occurrence entre communautés bactériennes et eucaryotes associés aux mêmes échantillons de fourmis. En utilisant la technique du métabarcoding, nous avons montré que les communautés eucaryotes associées aux fourmis amazoniennes sont différentes en fonction du type de nid, car les fourmis nichant dans le sol ont des communautés eucaryotes différentes des fourmis nichant dans les arbres. Tout comme les communautés bactériennes, les communautés eucaryotes associées aux fourmis pourraient aussi être différentes en fonction du type de nid parce qu'elles pourraient être en partie acquise par l'environnement. En revanche, les communautés bactériennes n'étaient pas différentes en fonction de l'habitat dans lequel la fourmi a été collectée, ni en fonction de la taxonomie ou de l'histoire évolutive des fourmis. Concernant l'étude des co-occurrences entre communautés bactériennes et eucaryotes, nos résultats semblent montrer que les communautés bactériennes sont structurées par la taxonomie de leur hôte et non par la présence d'eucaryotes spécifiques.

Cependant, ces résultats doivent être considérés avec prudence, car après avoir exclu toutes les séquences des hyménoptères, la majorité des séquences eucaryotes obtenues ne pouvaient être identifiées. De plus, parmi celles qui pouvaient être identifiées, quelques-unes ont pu être identifiées jusqu'à l'espèce, mais d'autres n'ont pu être identifiées qu'au phylum ou à l'ordre. Ce problème pourrait être limité en utilisant l'approche du multi-marqueur métabarcoding présenté dans l'état de l'art. En effet, différentes études ont montré que l'utilisation de cette approche permettait d'obtenir une plus grande diversité qu'en utilisant uniquement l'ARN ribosomique 18S^{103,104}. Cette approche du multi-marqueur métabarcoding pourrait nous aider à obtenir à la fois plus de séquences eucaryotes à partir de nos échantillons, et en même temps une meilleure identification taxonomique de ces séquences, qui sont les deux principales limites pour étudier la diversité des communautés eucaryotes et leur co-occurrence avec les communautés bactériennes au sein d'un même hôte. Malgré ces limites, nous

avons tout de même identifié de nombreux eucaryotes connus pour être associés aux insectes, comme les apicomplexans²⁶⁵ (appartenant au groupe des protistes), plusieurs phylum de champignon (comme les Saccharomycotina³¹¹), et les Rhabditida³¹² (ordre des nematodes).

Chapitre III : Etude des clusters de gènes de biosynthèse des bactéries associées aux fourmis du genre *Cephalotes*

I. Introduction

Au cours des chapitres 1 et 2, nous avons étudié la diversité et la structuration des communautés bactériennes et eucaryotes associées aux fourmis amazoniennes. Nous avons montré que les communautés bactériennes en particulier étaient structurées par la taxonomie de la fourmi, son régime alimentaire et son type de nid. Un autre facteur important dans la structure des communautés est la capacité de chaque organisme à pouvoir se maintenir dans cette communauté. Pour cela, les organismes doivent être capables de s'installer et de survivre dans cet environnement compétitif.

Dans le cas des bactéries, la production de métabolites secondaires pourrait être un des moyens utilisés pour s'établir dans ces communautés. En effet, les bactéries sont capables de produire une grande diversité de métabolites secondaires, dont certains ont des fonctions connues, d'autres seulement supposées et d'autres encore complètement inconnues. Parmi ceux ayant des fonctions connues, on peut citer les peptides non ribosomiques et les polycétides qui sont couramment utilisés dans le domaine pharmaceutique car ils ont de nombreuses actions antimicrobiennes très étudiées, ou encore les sidérophores⁷⁹ qui sont des molécules chélatrices du fer permettant à la bactérie de se nourrir en captant le fer dans leur environnement. Ainsi, dans un contexte de vie en communauté, la production de nombreux et divers métabolites secondaires permettrait aux bactéries de prospérer.

L'étude de la production de métabolites secondaires dans une communauté bactérienne est compliquée. Certaines méthodes prometteuses utilisées aujourd'hui sont la co-culture de bactéries ou l'utilisation de biofilms bactériens. Mais ces méthodes ne permettent d'étudier que les métabolites produits à un moment donné et non tous les métabolites qui pourraient être produits. Une autre méthode, plus indirecte, est l'étude des clusters de gènes de biosynthèse qui sont localisés dans le génome bactérien et qui codent pour des enzymes spécifiques utilisés dans les différentes voies de biosynthèse pour produire des métabolites secondaires⁷⁶.

C'est cette dernière méthode que nous avons décidé d'utiliser dans le chapitre 3 de cette thèse. L'objectif de ce dernier chapitre est d'étudier la diversité des gènes de biosynthèse des communautés bactériennes associées aux fourmis amazoniennes. Pour cela, nous nous sommes concentrés sur les communautés bactériennes associées aux fourmis du genre *Cephalotes*, et ce pour deux raisons. Tout d'abord, comme nous l'avons montré dans le premier chapitre de cette thèse, les communautés bactériennes associées aux fourmis *Cephalotes* sont très conservées et composées en grande majorité de cinq ordres bactériens : Burkholderiales, Opitutales, Pseudomonadales, Rhizobiales et Xanthomonadales. Ensuite, plusieurs génomes et métagénomes de bactéries intestinales associées à différentes espèces de fourmis *Cephalotes* ont été publiés récemment dans une étude démontrant que les cinq ordres bactériens cités précédemment jouent un rôle fondamental dans le recyclage de l'azote en acides aminés essentiels pour la fourmi hôte¹⁸. Pour ces deux raisons, nous avons analysé la diversité des gènes de biosynthèse des génomes et métagénomes des communautés bactériennes associées aux fourmis *Cephalotes*. L'article correspondant à ce chapitre a été soumis en mars 2021 à la revue scientifique « *Frontiers in Microbiology* » tel qu'il est présenté ici. C'est pourquoi ce chapitre 3 est rédigé en anglais dans ce manuscrit de thèse.

II. Assessing biosynthetic gene cluster diversity of specialized metabolites in the conserved gut symbionts of herbivorous turtle ants

Anaïs Chanson¹, Corrie S. Moreau², Christophe Duplais^{3*}

¹Université de Guyane, UMR8172 EcoFoG, AgroParisTech, CNRS, Cirad, INRAE, Université des Antilles, Kourou, France

²Departments of Entomology and Ecology & Evolutionary Biology, Cornell University, Ithaca, NY, USA

³CNRS, UMR8172 EcoFoG, AgroParisTech, Cirad, INRAE, Université des Antilles, Université de Guyane, Kourou, France

***Corresponding author**

Christophe Duplais

CNRS-UMR EcoFoG at the Pasteur Institute of French Guiana

23 avenue Pasteur, 97300 Cayenne, France

christophe.duplais@cirs.fr

Phone: +594 594 293 134

Keywords: Insect-microbe mutualism; symbiosis; metagenomics; gut bacteria; Cephalotes

A. Abstract

Cephalotes are herbivorous ants (>115 species) feeding on low-nitrogen food sources and they rely on gut symbionts to supplement their diet by recycling nitrogen food waste into amino acids. These conserved gut symbionts, composed of five bacterial orders, have been studied previously for their primary nitrogen metabolism, however little is known about their ability to biosynthesize specialized metabolites which can play a role in bacterial interactions between communities living in close proximity in the gut. We investigated the diversity of biosynthetic gene clusters (BGCs) producing specialized metabolites in the genomes and metagenomes of conserved gut symbionts by studying 17 *Cephalotes* species collected across several geographical areas. Our results reveal that (i) mining metagenomes and genomes show complementary results to retrieve BGCs especially when bacterial isolates are difficult to culture, (ii) the conserved gut symbionts involved in the nutritional symbiosis have a large diversity of BGCs of different chemical families, (iii) the phylogenetic analysis of BGCs encoding the production of arylpolyenes, non-ribosomal peptides (NRP), polyketides (PK), and siderophores shows high similarity between BGCs of a single symbiont across different ant host species, and between BGCs originated from different bacterial orders within a single host species. Additionally, the diversity of BGCs was found in four of the five conserved symbionts co-occurring in the hindgut except for one major player (Opitutales) localized alone in the midgut and lacking BGCs. This spatial isolation prevents direct interaction of Opitutales with other symbionts and suggests that BGCs have an essential role for symbionts living in close proximity. These findings together pave the way for studying the mechanisms of BGCs conservation and evolution in gut symbiont genomes and the role of bacterial specialized metabolites involved in multipartite mutualism between symbionts and *Cephalotes* ants.

B. Introduction

Bacterial symbionts are widespread among insects and have shaped the evolution of their hosts^{181,182}. The symbiotic interactions between bacterial communities inside the host, and between bacteria and hosts, relies on metabolites, metabolic pathways and enzymes regulating metabolic flux.

Among mutualistic bacteria the nutritional symbionts supply a large diversity of nutrients to the diet of their insect hosts which often feed on nutritionally limited food sources. In the case of strictly blood-sucking ticks, intracellular bacterial symbionts provide vitamin B to the host^{2,321}. Herbivorous aphids and ants feeding on nitrogen-poor diets depend on gut bacteria to recycle nitrogen-rich food waste to biosynthesize essential and non-essential amino acids^{18,322}. Gut bacteria are also essential to degrade various substrates to enrich the host's diet as is the case for xylophagous termites that rely on gut bacteria to degrade the lignocellulose³²³ whereas the gut bacterial community of bees degrade polysaccharides³²⁴.

Another type of microbial symbiosis involves defensive mechanisms as the bacterial symbionts can enhance insect resistance to a variety of natural enemies including bacteria, fungi or nematodes^{325,326}. For example, in different insect families, including beetles, wasps, bees and ants, the bacterial symbionts produce a cocktail of antimicrobial compounds to protect the nest⁷³, the brood^{75,327} or a fungal strain in case of multipartite mutualism^{72,328}.

Bacterial secondary metabolites are produced via specific groups of genes called biosynthetic gene clusters (BGCs). In a biosynthetic gene cluster, genes are located in close proximity to each other in the bacterial genome, and encode for specific enzymatic steps in the biosynthetic pathway of metabolites³²⁹. Bioinformatics tools have been developed to mine bacterial genomes^{329,330} and retrieve the BGCs of different chemical families including arylpolyene, lanthipeptide, non-ribosomal peptide, polyketide and terpene. Metabolites play several roles in bacterial communities for colonizing tissue, communication through quorum sensing, competition using antimicrobial molecules, and nutrient acquisition with siderophore. BGC studies often focus on the genomes of culturable strains from environmental bacteria^{331,332} or host-associated bacteria^{333,334} and contribute to the discovery of bioactive compounds for medicine. Mining of bacterial metagenomes has also been investigated from environmental samples³³⁵, as well as bacterial communities associated within hosts³³⁶⁻³³⁹.

The study of bacterial BGCs in symbiosis anticipates the chemical structure of metabolites encoded in symbiont genomes and provide a framework to identify candidate genes which are maintained to

support bacterial colonization and maintenance within a host. In addition the comparison between bacterial BGCs across different hosts³⁴⁰ in different geographic areas³⁴¹ allows to test what environmental or phylogenetic factors shape BGC diversity in host-associated bacterial communities.

To investigate the BGC diversity of insect symbionts, we studied the association between gut bacteria and herbivorous *Cephalotes* turtle ants. In this nutritional symbiosis the nitrogen-poor diet of ants is supplemented by a core microbiome which recycle urea food waste into amino acids beneficial to the ant^{219,342}. Genomic approaches have revealed the redundancy in functions related to the nitrogen flux in the genome of conserved symbionts belonging to five bacterial orders¹⁸. Unlike most intra- or extracellular single nutritional symbionts, the *Cephalotes* multipartite gut bacteria mutualism may retain metabolic functions selected not solely for direct benefit to hosts, but for maintaining diverse bacterial community members. To assess if the gut bacteria associated with *Cephalotes* possess BGCs we focused on bacterial genomes and metagenomes from the gut of 17 *Cephalotes* species collected in distinct geographic areas of North and South America¹⁸. Here we mined both genomes and metagenomes of gut bacteria associated with *Cephalotes* turtle ants to predict the specialized metabolites of the symbionts involved in multipartite mutualism. We analyze the bacterial origin of each BGC to understand their occurrence in the conserved gut symbionts involved in a nutritional symbiosis. We focused in on different core genes of BGCs (arylpolyene, non-ribosomal peptide (NRP), polyketide (PK) and siderophore) to identify architectural similarity patterns across different symbionts and host species. This work shows the relevance of genome and metagenome mining in an insect-microbe symbiosis and pave the way for future study on the evolution of BGCs in symbiont genomes across the host phylogeny.

C. Material and Methods

1. Genomes and metagenomes analysis

The 14 genomes of cultured gut bacteria and 18 metagenomes of *Cephalotes* gut bacteria were obtained from JGI-IMG version 5.0³⁴³ (Supplementary Tables 1 and 2) from the previous projects Gs0085494 (“*Cephalotes varians* microbial communities from the Florida Keys, USA”), Gs0114286 (“Symbiotic bacteria isolated from *Cephalotes varians*”), Gs0117930 (“*Cephalotes* ants gut microbiomes”) and Gs0118097 (“Symbiotic bacteria isolated from *Cephalotes rohweri*”)¹⁸.

The 14 cultured isolate genomes are all part of the *Cephalotes* core microbiome: two *Cephaloticoccus* genomes (order: Opitutales, class: Opitutae; one genome from each ant host), four *Ventosimonas* genomes (order: Pseudomonadales, class: Gammaproteobacteria; one genome from *C. varians* and three genomes from *C. rohweri*), six Burkholderiales genomes (class: Betaproteobacteria; four genomes from *C. varians* and two genomes from *C. rohweri*), one Rhizobiales genome (class: Alphaproteobacteria; from *C. varians*), and one Xanthomonadales genome (class: Gammaproteobacteria; from *C. rohweri*). All these bacteria belong to the Proteobacteria phylum, except Opitutales which are part of the Verrucomicrobia phylum.

The metagenomic data used in this study have two metagenomes from the same *C. varians* species. The metagenomes were analyzed via the software Anvi'o version 5.5¹⁴² to sort the different bacterial families composing each metagenome into distinct bins. In this analysis, the fasta sequence of a metagenome is used to create a contig database and a profile database. Then, this contig database is visualized and bins are manually created to maximize completeness while minimizing contamination. Finally, the software CheckM version 1.1¹⁴⁴ was used to verify the completeness, contamination and strain homogeneity of each bin, and to identify the taxonomic lineage of each bin (Supplementary Table 3).

The published *Cephalotes* phylogeny was retrieved³⁴⁴ and the packages ape³⁴⁵ and phytools²¹¹ of the R software version 3.6.1²⁰⁴ were used to exclude from this phylogeny the *Cephalotes* species from which no symbiont genomes or metagenomes were available.

2. Bacterial biosynthetic gene clusters analysis

The bacterial biosynthetic gene clusters (BGCs) of each genome and each metagenomic bin were analyzed with antiSMASH version 5.0³⁴⁶ (Supplementary Tables 4 and 5) with the following analysis options: strict detection, and activation of search for KnownClusterBlast, ClusterBlast, SubClusterBlast and Active site finder. BGCs smaller than 5kb were then filtered out of the data and were not included. The taxonomic classification of each cluster was verified to the genus level using the software Blast+¹⁴⁷ (Supplementary Table 5). There was no identification conflict between CheckM and Blast+. NaPDos (<https://npdomainseeker.sdsc.edu/>, accessed July 2020)³⁴⁷ was used to classify the ketosynthase (KS) domain and condensation (C) domain sequences of NRP and PK retrieved from the genome and metagenome mining analysis and to infer the KS and C phylogenies.

3. Genomic similarity assessment

The genomic similarity between the cultured isolate genomes and the metagenomes from the *C. rohweri* and *C. varians* were compared using two independent methods (Supplementary Tables 6 and 7). First, each metagenomic bin was compared to each cultured isolate genome from the corresponding host species using the combined values of the average nucleotide identity (gANI) and alignment fraction (AF)^{348,349}. In this method, the nucleotide of the protein-coding genes of each metagenomic bin (genome 1) and each cultured isolate genome (genome 2) are compared using NSimScan³⁵⁰, and only the bidirectional best hits (BBHs) displaying at least 70% sequence identity over at least 70% of the length of the shorter sequence are kept. The gANI is calculated by dividing the sum of the percent of identity times the alignment length for all BBHs by the sum of the lengths of the BBH genes. A gANI higher than 0.95 indicates a threshold for same species level³⁵¹. The AF is calculated by dividing the sum of the lengths of all BBH genes by the sum of the length of all the genes in genome 1. For AF values higher than 0.5 are considered conspecifics³³⁷.

Next, the contigs of each metagenomic bin were mapped to each cultured isolate genome from the corresponding host species using the `bbmap.sh` function in the BBMerge software version 38.22³⁵² with the parameters: `kfilter=22 subfilter=15 maxindel=80`. These parameters were used in a similar study published recently³³⁷. BBMap aligns contig sequences from each metagenomic bin (genome 1) and each cultured isolate genome (genome 2) and return the percentage of contigs in genome 2 which were found in genome 1.

4. Genetic networking construction

The genetic networking of the genomic and metagenomic bacterial BGCs was generated using the the 'biosynthetic gene similarity clustering and prospecting engine' BiG-SCAPE version 1.0.0^{40,41}. The network was constructed with the three following options: "`--include_singletons`", "`--mix`", and "`--cutoffs 1.0`". Briefly, the Pfam-domains of each BGC were identified using HMMER version 3.1b2¹⁴⁶ with the respective Hidden Markov Models (HMM) obtained from the Pfam database¹³⁷. The distance metric is created by calculating the distance between every pair of BGC. Pairwise distance calculation is divided between three values: the percentage of shared domain types (Jaccard Index), the similarity of aligned domain sequences (Domain Sequence Similarity index) and the similarity of domain pair types (Adjacency Index). Specific details for each index and the method are available in Navarro-Munoz *et al.*³⁵⁴. The resulting similarity matrix was filtered with different thresholds between 0.4 and 0.2^{78,354-356}, and the thresholds 0.45 was chosen because it filtered out enough to form different groups while maximizing the number of bacterial BGCs in each network^{78,356}. The filtered similarity matrix was visualized with Cytoscape version 3.7.2²⁹⁶, using the MCL clusterization algorithm from clusterMaker2³⁵⁷ version 1.3.1 with an index value $I = 2.0$.

5. CORASON phylogenetic reconstruction of BGCs

The phylogenetic reconstruction of four types of BGCs (arylpolyene, NRP, siderophore and T1PK) were analyzed with CORASON version 1.0³⁵⁴ and a query gene. First, a database for each type of BGCs was created from all the BGCs of the same type recovered in the genomes and metagenomes, which were obtained from our previous AntiSMASH version 5.0 analysis. Then a BGC from each of our databases was chosen as the reference BGC for its database, and a gene from the biosynthetic core of this reference BGC was chosen as the query protein. The choice of the reference BGC and query protein for a database were made taking different variables into account: (1) the reference BGC must be one of the longest BGCs in the database, (2) the query protein must come from a biosynthetic core gene or an additional biosynthetic gene close to the core, and (3) the query protein must be present in at least half of the BGCs in the database.

Once a reference BGC and a protein query were selected for a database, the CORASON version 1.0 was used to determine the core of this type of BGCs, and to infer the phylogenies. For the phylogenetic reconstruction, the core BGCs are concatenated and aligned using Muscle version 3.8.31³⁵⁸, and the phylogeny is inferred using the Maximum likelihood method in FastTree version 2.1.10³⁵⁹. Specific details for the method are available in Navarro-Munoz *et al.*³⁵⁴. If less than half of the BGCs present in a database appear in the phylogeny, or if many BGCs appear more than once in the phylogeny (which may happen if the selected query protein belongs to a biosynthetic gene that can be found multiple times in the same BGC), the reference BGC and query protein previously selected were deemed to be inadequate for the analysis. In this case, the process to choose a reference BGC and query protein were done once more, following the same procedure as before. The reference BGC and query protein selected for the four types of BGCs studied are presented in Supplementary Table 8. The strength of this method is to combine four important steps into one tool: conserved core gene calculation, orthologue genes blasting, sequence alignment and phylogenetic reconstruction.

D. Results

1. Mining bacterial genomes

The bacterial genomes (N=14) studied herein originated from cultured gut bacteria associated with *Cephalotes varians* and *Cephalotes rohweri*. Among these 14 bacterial genomes, only three genomes, all from *C. varians*, did not possess any BGC: Burkholderiales sp. Cv33a, *Cephaloticoccus capnophilus* Cv41 and Rhizobiales sp. JR021-5. In the other bacterial genomes, a total of 31 biosynthetic gene clusters (BGCs) have been found and the type and length of the 31 BGCs are listed in Supplementary Table 4. In bacteria associated with *C. varians*, 10 BGCs were found representing seven types of BGCs: arylpolyene, beta-lactone, ectoine, nonribosomal peptide (NRP), resorcinol, polyketide type I (T1PK) and terpene. In bacteria associated with *C. rohweri*, 21 BGCs were found representing six types of BGCs: arylpolyene, beta-lactone, ladderane, NRP, siderophore and T1PK. In *C. varians* the majority of BGCs originate from Burkholderiales symbionts whereas in *C. rohweri* they originate from different bacterial orders (Burkholderiales, Pseudomonadales and Xanthomonadales) (Supplementary Figure 18).

2. Assessing the quality of metagenomic data

The analysis of the 17 *Cephalotes* metagenomes resulted in 168 bins. Almost half the bins (72 out of 168) have a predicted completeness higher than 90%, 48 bins have a completeness higher than 70%, 31 bins have a completeness higher than 50%, and only 17 bins have a completeness lower than 50% (Supplementary Table 3). Concerning the contamination, half the bins (85 out of 168) have a predicted low contamination level (<5%). The 168 bins were identified from 31 different bacterial orders (Supplementary Figure 19A) and 54% of the bins were not contaminated with different bacterial orders (Supplementary Figure 19B). Mining these bins using AntiSMASH version 5³⁴⁶ resulted in a total of 233 BGCs found in 111 bins (Supplementary Table 5). The BGCs were identified from 21 different bacterial

orders. The bacterial orders having the highest number of BGCs are Burkholderiales (94 BGCs), Pseudomonadales (42 BGCs), Xanthomonadales (24 BGCs) and Rhizobiales (19 BGCs).

To further evaluate the quality of the metagenomic data we investigate the genomic similarity of each metagenomic bin to the cultured isolate genomes from *C. rohweri* and *C. varians* (seven strains from each ant species). We created a heatmap representing the average nucleotide identity (gANI) and the alignment function (AF) values as a proxy of the sequence similarity (Supplementary Figure 20, Supplementary Table 6)³³⁷. The mapping of bins to cultured isolate genomes was considered successful if it met the $gANI \times AF > 80$ threshold. In *C. rohweri*, six out of eleven metagenomic bins were successfully mapped at the species level ($gANI > 0.95$) to seven cultured isolate genomes (Supplementary Figure 20A). The metagenomic bins 5 and 12 were mapped to three *Ventrosimonas* sp. genomes, bin 9 was mapped to the genome of *Xanthomonadales* sp., bins 14 and 15 were mapped to three *Burkholderiales* sp. genomes, and bin 16 were mapped the genome of *Cephaloticoccus primus*. In *C. varians* PL005W, four out of eight metagenomic bins were successfully mapped at the species level ($gANI > 0.95$) to four genomes (Supplementary Figure 20B). The metagenomic bin 4 was mapped to the genome of *Ventrosimonas gracilis*, bins 7 and 8 were mapped to *Burkholderiales* sp. genomes, and finally bin 10 was mapped to the genome of *Cephaloticoccus capnophilus*.

We compare the number of shared BGCs between metagenomic bins and genomes (Supplementary Figure 21, Supplementary Table 7). In *C. rohweri*, 11 BGCs were shared including six from the bacterial order Burkholderiales (two NRPs, two beta-lactones and two T1PKs), four from Pseudomonadales (two arylpolyenes and two NRPs) and one from Xanthomonadales (one ladderane). The analysis shows that 10 BGCs were unique to the genomes and 23 BGCs were unique to the metagenomes. In *C. varians* PL005W two BGCs were shared, one T1PK from Burkholderiales and one arylpolyene from Pseudomonadales. Additionally, six BGCs were found only in the genomes and 16 BGCs were found only in the metagenomes.

In *C. varians* PL010W metagenome only one bin was mapped to a *Burkholderiales* sp. genome with the $gANI \times AF > 80$ threshold (Supplementary Figure 20B). Four BGCs were identified (Supplementary

Table 5) and none of them match any BGCs found in *C. varians* genome (Supplementary Table 7). Because of the low numbers of BGCs recovered in the *C. varians* PL010W metagenome, we removed this metagenome for further analysis and use only *C. varians* PL005W metagenome for this species.

3. Diversity of BGCs across the *Cephalotes* phylogeny

From the 233 BGCs retrieved in the genomes and metagenomes we identified 19 different types of BGCs: acyl amino acids, arylpolyene, bacteriocin, beta-lactone, butyrolactone, cyclodipeptide (CDP), ectoine, furan, hserlactone, ladderane, lanthipeptide, linear azol(in)e-containing peptides (LAP), non-ribosomal peptide (NRP), phenazine, resorcinol, siderophore, polyketide type I (T1PK), terpene, and thiopeptide (Supplementary Table 5). The arylpolyene and NRP BGCs were the most numerous with 66 and 57 respectively retrieved in the genomes and metagenomes. The type and length of all the BGCs are presented in Supplementary Tables 4 and 5. Few bacterial BGCs found in the metagenomic analysis are hybrids made of different types of BGCs: arylpolyene-beta-lactone-resorcinol, arylpolyene-ladderane, arylpolyene-resorcinol, and NRP-T1PK.

The gut metagenomes of 17 species of *Cephalotes* studied here were collected in three different countries, ten species from Brazil, five species from Peru and two species from the USA. The lowest number of BGCs (N=1) was found in *C. angustus*, the highest number of BGCs (N=31) was recorded for *C. rohweri* and a mean of ~14 BGCs per species (Supplementary Table 5). In Figure 21, the most abundant types of BGCs, those having at least five BGCs found in metagenomes, are shown across the different *Cephalotes* species representing a total of 202 BGCs originating from 19 different bacterial orders. The vast majority of BGCs were identified in Burkholderiales, Pseudomonadales and Xanthomonadales, and Rhizobiales.

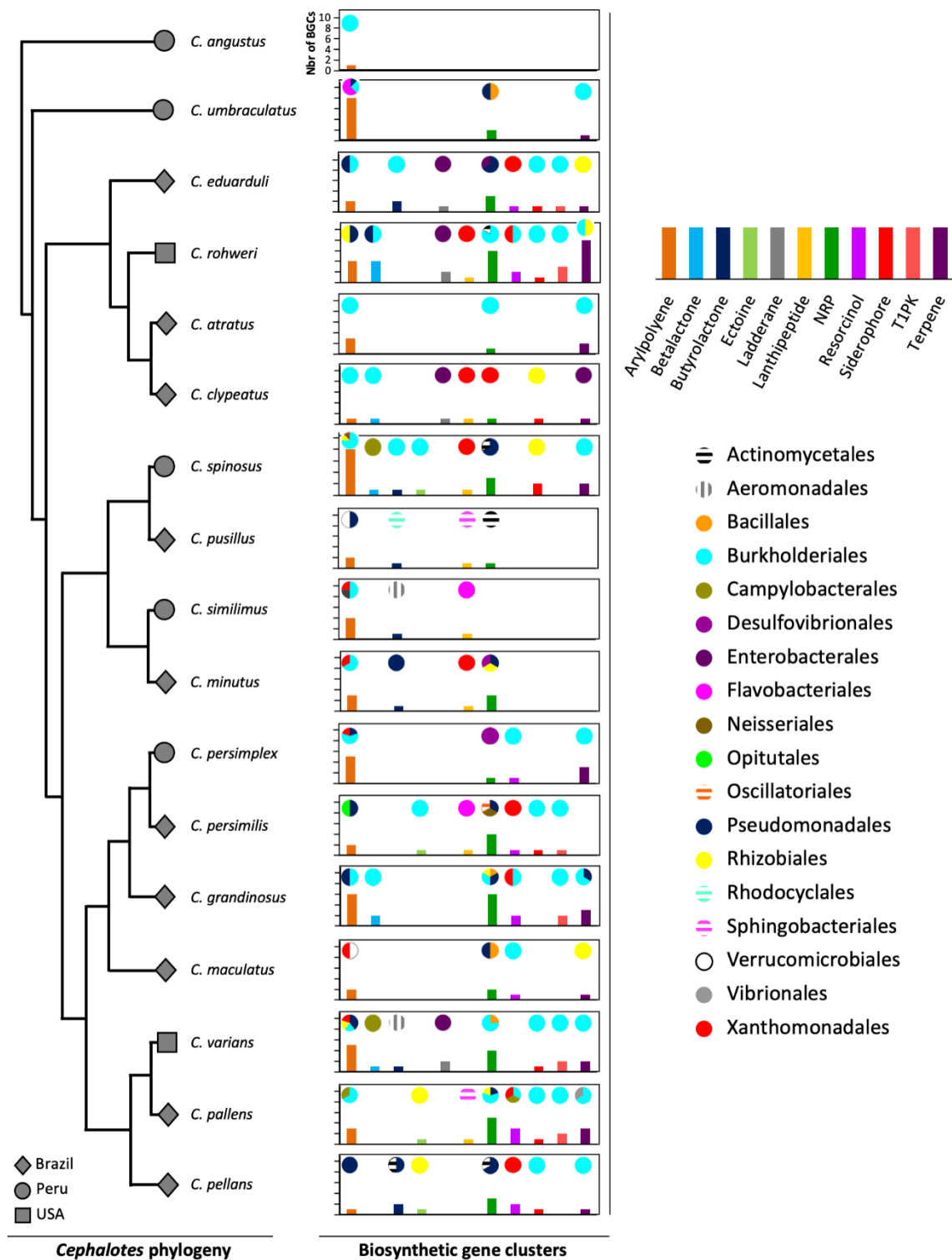


Figure 21. Diversity of biosynthetic gene clusters of bacteria associated with *Cephalotes* turtle ants. The grey symbols represent the country from which an ant originated. The graphs represent the bacterial BGCs found in the metagenomes of each *Cephalotes* species in this study. The bars of the graphs represent the number of BGC of each type found in the metagenomes. The pies represent the proportion of bacterial orders from which each BGC type originated.

4. Genetic networking and BGCs phylogenies

To assess the genetic architectural diversity of BGCs (number and type of gene, gene order, gene domain, and BGC length) we used BIG-SCAPE version 1.0.0³⁵⁴ to create a genetic network of the BGCs found in the genomes and metagenomes (Figure 22). We used the type of BGCs and the bacterial order to color the outer and inner circle of nodes respectively. In this analysis, 31 networks of BGCs are formed, each containing between 2 to 24 nodes, and 90 nodes are singletons. Each network is mainly composed of the same type of bacterial BGCs. A third of the networks contain one or two BGC nodes of a different type than rest of the nodes composing this network. Additionally, in half of the networks, BGCs originated from the same bacterial order, while in the other half the BGCs originated from two or three different bacterial orders. Notably the 26 out of 31 BGCs from cultured isolate genomes are clustered in networks shared with BGCs originated in metagenomes.

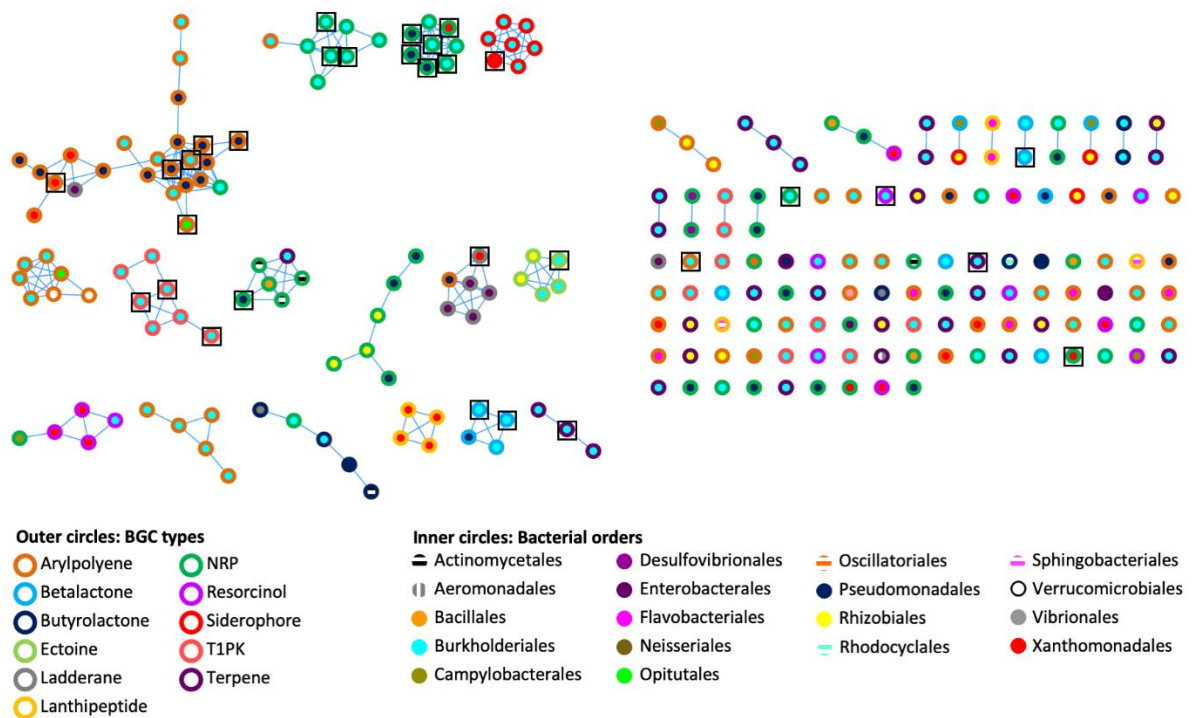


Figure 22. Genetic networks of the BGCs identified in bacterial genomes and gut *Cephalotes* metagenomes. A distance threshold of 0.45 and an MCL index value of 2.0 was applied to generate the network. Color codes are respective for bacterial BGCs type and bacterial order. Black squares represent the 31 BGCs found in bacterial genomes.

The phylogenetic reconstruction of the BGCs using a query gene were inferred using CORASON version 1.0³⁵⁴ (Figure 23, Supplementary Figures 23 and 24, Supplementary Table 8). The four types of BGCs, arylpolyene, NRP, siderophore, and T1PK, were chosen for this analysis since they are abundant in the genomic data and also because these chemical families are thought to play an important role in bacteria colonization, interaction and competitiveness^{360–363}. Using an arylpolyene KS gene as a query gene 38 different BGCs were retrieved in the arylpolyene BGC phylogeny out of the 66 detected in the *Cephalotes* bacterial genomes and metagenomes (Supplementary Figure 22). These 38 arylpolyene BGCs form several distinct clades. One clade is composed of nine BGCs which are grouped in the largest arylpolyene network cluster (Figure 23A). These BGCs originated from the genomes and metagenomes of Burkholderiales and Pseudomonadales isolated from six different ant species (Figure 23A). In this clade, the arylpolyene BGCs have a length of between 24,572 to 43,580 nucleotides (nt) and 26 to 44 genes. In the NRP BGC phylogeny based on an AMP-binding gene (adenylation domain), 33 different NRP were retrieved out of the 57 detected in the *Cephalotes* bacterial genomes and metagenomes (Supplementary Figure 23). These 33 NRP BGCs form several distinct clades including a polytomy composed of seven BGCs found in the network cluster (Figure 23B). In this clade, the NRP BGCs were identified in the genomes and the metagenomes from three bacterial orders (Burkholderiales, Pseudomonadales, and Xanthomonadales) isolated from a single ant host (*C. rohweri*). NRP BGCs have a length of between 23,818 to 53,610 nt and 25 to 53 genes in this clade. The PKS-KS gene results in the T1PK BGC phylogeny of 10 BGCs out of the 15 detected in the *Cephalotes* bacterial genomes and metagenomes (Figure 23C). All BGCs originated from the genomes and metagenomes of Burkholderiales isolated from six different ant species. T1PK BGCs have a length of between 10,882 to 47,950 nt and 5 to 42 genes. Seven T1PK BGCs were found in a single network cluster and the three T1PK BGCs remaining all fall as singletons in the genetic network analysis (Figure 22). The *lucA-lucC* gene was used to infer the siderophore BGC phylogeny. Eight siderophores were retrieved out of the 10 detected in the *Cephalotes* bacterial genomes and metagenomes (Figure 23D). Siderophore BGCs have a length of between 7,469 to 10,804 nt and 4 to 11 genes. A network matches seven siderophore

BGCs from the genomes and metagenomes of Burkholderiales and Xanthomonadales each associated with a different ant species.

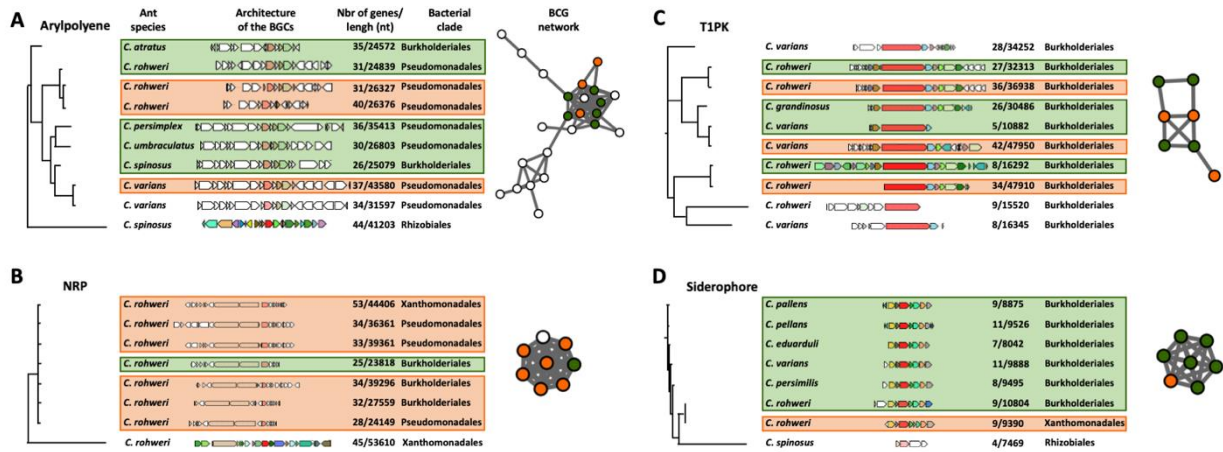


Figure 23. CORASON phylogenetic reconstruction of arylpolyene (A), NRP (B), T1PK (C) and siderophore (D) biosynthetic gene clusters. Because of their large size, only a selected clade of the arylpolyene and NRP phylogenies are presented here. The whole phylogenies of these clusters can be found in Supplementary Figures 23 and 24 respectively. These phylogenies show the genomic composition of each BGC, the size and number of genes of each BGC, the bacterial origin of the BGC (on the right of the BGC) and the host ant species (on the left of the BGC). Each colored BGCs in the phylogeny are represented as node in the networks which were generating in the previous genetic network analysis (Figure 22). The nodes in the network are colored orange for BGC originated from a bacterial genome, or green for BGC originated from metagenome. The white nodes represent BGCs present in the network analysis but were not found in the phylogenetic clade presented here.

The elongation domain phylogenies of NRP synthase and PK synthase, respectively C and KS domains, were inferred using NaPDos³⁴⁷ to check if the sequences of known metabolites matched our data (Figure 24). The majority of the identified C domain sequences (66 out of 89) from genomes and metagenomes belong to the cyclization domain class catalyzing both peptide bond formation and subsequent cyclization. The retrieved KS domain sequences (N=13) from genomes and metagenomes fall into a single clade corresponding to *Cis*-AT modular class or the iterative domain class. It has been shown that the C and KS domains are the most informative to predict pathway associations of NRP and

PK clusters respectively^{364,365}. Unfortunately, no close similarity with the identified C and KS sequences of known non-ribosomal peptide and polyketide were found when comparing with NaPDos and the NCBI database, thus no prediction could be made about the chemical structure of bacterial secondary metabolites.

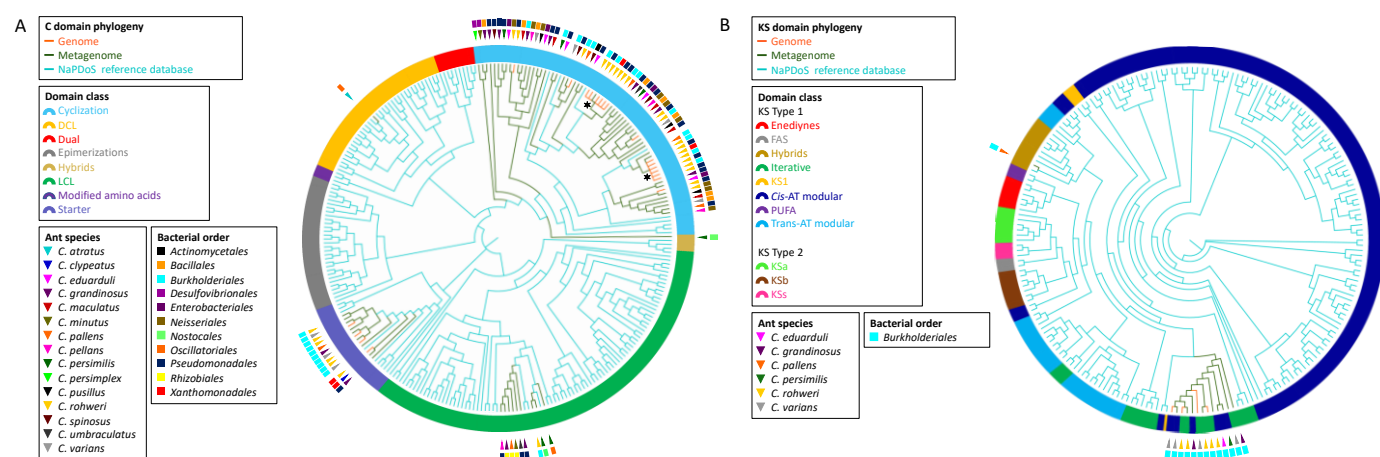


Figure 24. Phylogeny of retrieved C and KS elongation domain sequences in the NRP and PK assemblies respectively. The phylogenies were inferred from the C domain sequences (A) and the KS domain sequences (B) from the *Cephalotes* bacterial genomes (orange branches) and metagenomes (green branches) with the NaPDos reference database (blue branches). The ring surrounding the tips represent C domain class or KS domain class. The bacterial order from which each domain sequence was retrieved is indicated by the colored rectangle. The ant species associated with each sequence is indicated by the colored triangle. The symbol * designates the clade of the two C domain sequences found in the NRP phylogenetic reconstruction (see Figure 23).

E. Discussion

Since the expansion of computational tools for mining microbial genomes, the discovery of BGCs of interest for pharmaceutical applications has become a growing field³²⁹. This approach is complementary to natural products chemistry and metabolomics in helping with the prioritization of

strains producing potential drugs with original chemical structures. Genome mining studies on large scale prokaryote genomic data have revealed that BGCs are distributed in large gene cluster families, with the vast majority of them still uncharacterized⁷⁸. Some of these BGCs are distributed widely across the entire bacterial domain, with the most prominent being arylpolyenes, polyketides, saccharides and siderophores, having known and putative functions in microbe-host and microbe-microbe interactions⁷⁸. The rise of advanced genomic tools to mine genomes and predict the chemical structure of metabolites from BGCs will certainly benefit the field of host-microbe symbiosis. Herein we present the first genome and metagenome mining study of gut bacteria across different species of herbivorous turtle ants to assess the diversity of BGCs in the genome of nutritional symbionts. When applied across the host phylogeny a genome mining approach provides crucial information about how the evolutionary history of the host may impact the diversity of BGCs in the symbiont genome. For instance BGCs can undergo a vast array of evolutionary process through *de novo* assembly, gene duplication/deletion, and horizontal gene transfer which drive the microbial chemodiversity³⁶⁶. The systematic study of BGCs diversity of symbionts across host species from different geographic areas may provide a list of BGCs likely to play a crucial functional role for the symbiont and for the host.

Recently studies mining metagenomes for BGCs include the defensive symbionts associated with the eggs of the beetle *Lagria villosa*³³⁶, with the marine sponge *Mycale hentscheli*³⁶⁷ and with the gill of shipworm³³⁷. In the latter study the quality of the metagenomic data was evaluated by mapping the metagenomic bins to the symbiont genomes to assess sequence similarity. In our case the main limitation is to culture the gut symbionts of *Cephalotes* which are very difficult to grow *in vitro*. To date 14 strains have been isolated in *Cephalotes varians* and *Cephalotes rohweri* and 10 out of 19 metagenomic bins from these two species were mapped to cultured isolate genomes (Supplementary Figure 23, Supplementary Table 6). Although the gANI×AF values (from 0.76 to 0.91) indicate moderate to good sequence similarity between genomes and bins, the percentages of contigs mapped for each bin vary from 11% to 83%. Despite the fact that having more cultured isolate genomes could increase the mapping of bins to genomes we cannot rule out partial misassembly of metagenomic bins. Notably

several BGCs detected in metagenomic bins present high sequence similarity and high percentage of contigs mapped with those retrieved from genomes suggesting good confidence in the metagenome-assembled genome analysis (Supplementary Figure 21, Supplementary Tables 7 and 8). Although the contamination of bins can be high in some cases, we are looking for the highest bin completeness in this study to assess the diversity of BGCs.

In the network analysis using the entire metagenomic data 26 out of 31 BGCs from cultured isolate genomes were found in networks shared with BGCs retrieved from metagenomes (Figure 22). Additionally some core gene sequences from genomic and metagenomic data of the BGCs of arylpolyene, NRP, T1PK and siderophore fall into the same clade in the respective reconstructed phylogenies (Figure 23). In parallel the C and KS domain sequences from genomes and metagenomes are also found in the same phylogenetic clade (Figure 24). We do not know how much of the underlying BGC diversity is captured by these metagenomes but taking together the strategy of mining genomes and metagenomes provide a complementary approach to explore the BGCs diversity in insect-microbe mutualism especially when gut symbionts are difficult to culture *in vivo*.

The low number of metagenomes (N=17) collected in a large geographic range (Brazil, Peru and the USA) prevents to test whether the host geography impacts the symbionts' BGCs. More genomes and metagenomes covering smaller geographic area are needed to test the geographic mosaic theory of coevolution which stipulates that in strongly interacting species, geography and community ecology can shape coevolution through local adaptation^{368,369}. For instance in the symbiosis between antibiotic-producing cuticular symbionts and fungus-growing ants it was demonstrated that the geographical isolation of ant populations in Central America has an impact on the antibiotic potency of the locally adapted symbiont but not on the BGC architecture³⁷⁰. Further effort is required to test if this pattern applies to the conserved gut symbionts of *Cephalotes* turtle ants.

Four bacterial orders possess more than 80% of all the 233 BGCs detected in the *Cephalotes* bacterial genomes and metagenomes: Burkholderiales (44%), Pseudomonadales (15%), Xanthomonadales (15%) and Rhizobiales (9%) (Figure 21, Supplementary Tables 4 and 5). Interestingly,

these four bacterial orders belong to the *Cephalotes* core bacterial community¹⁸. The *Cephalotes* core bacterial community is maintained across the *Cephalotes* phylogeny and benefit these herbivorous ants with a low-nitrogen diet by recycling nitrogen from urea into essential and non-essential amino acids. In many insect nutritional symbioses the endosymbiont resides in host cells or bacteriocytes and have a reduced genome with several copies of the functional genes responsible for the biosynthesis of amino acids and vitamins to enrich in nutrients the host diet. The genome reduction of gut symbionts in *Cephalotes* turtle ants has not been investigated yet however the mechanism for maintaining different communities originating from five bacterial orders is more complex compared to other symbioses involving a single gut symbiont (bean bugs³⁷¹), or bacteriocyte dwelling endosymbionts (*Camponotus* ants³⁷², beetles²³⁹). The conserved bacterial symbionts in *Cephalotes* have redundancy in N-metabolism and our results show these bacterial symbionts possess the vast majority of the BGCs detected. These findings could suggest that metabolic functions of a complex microbiome are often selected not solely for direct benefit of hosts, but for sustaining the gut community³⁷³.

This hypothesis is in accordance with recent findings in a bean bug which needs to acquire a specific *Burkholderia* gut symbiont from the soil environment in every new generation³⁷¹. They demonstrated that a large number of *Burkholderia* species were able to access the bean bug gut, but in experiments of co-inoculations, the specific *Burkholderia* symbiont always outcompetes all the other *Burkholderia* species. These findings highlight that *Burkholderia* gut symbionts colonize with high efficiency the insect gut and this trait is associated with the production of amino acids and siderophores³⁷³. The *Cephalotes* core symbionts are transmitted through oral-anal trophallaxis and this behavior is prevalent in *Cephalotes*³⁷⁴. Newly eclosed workers from *Cephalotes rohweri* performed oral-anal trophallaxis with mature adults prior the development of the proventricular filter which later blocks the entry of transient bacteria³⁷⁵. This remarkable transmission mechanism ensures the maintenance of the core gut symbionts over evolutionary time suggesting the maintenance of genes promoting gut colonization and bacterial interaction are likely highly selected to be maintained. The gut symbiont communities of *Cephalotes* originating from five bacterial orders need to colonize gut

tissue, tolerate each other, and possibly participate to cross-feeding. These interactions between co-occurring gut symbionts may be through the production of specialized metabolites.

To identify the structural similarities of conserved bacterial BGCs across different hosts, we performed CORASON version 1.0 gene analysis on four types of BGCs: arylpolyene, NRP, T1PK and siderophore (Figure 23, Supplementary Figures 23 and 24). The selected clades presented in Figures 23A and B show similarities in the sequence of the core genes for arylpolyene BGCs originating from Burkholderiales and Pseudomonadales, and for NRP BGCs originating from Burkholderiales, Pseudomonadales, and Xanthomonadales. These arylpolyene and NRP BGCs are each clustered in the networking analysis (Figures 22 and 23) demonstrating the architectural similarities. In the NRP BGC clade of Figure 23B the query gene AMP-binding (adenylation domain) and the two genes coding for a C domain have 100% similarity and 100% coverage across the BGCs retrieved from the six genomes and metagenomes of Burkholderiales, Pseudomonadales, and Xanthomonadales from the same host *C. rohweri*. This high similarity between the six C domain sequences from three bacterial orders is also confirmed in the phylogenetic analysis (Figure 24). This pattern suggests a possible convergence of the core genes in the gut of *C. rohweri* possibly via horizontal gene transfer. In insect-bacteria symbiosis horizontal gene transfer has been shown to occur even in bacteria going through genome reduction³³⁶. However the three sequences did not match any known sequence in the database limiting the identification of the bacterial origin of these genes. On the other hand, the T1PK and siderophore core gene phylogenies (Figures 23B and C) show that BGCs from Burkholderiales across different host species share a strong likely genetic conservation. Several identical genes of T1PK and siderophore BGCs are present across the core gene phylogenies supporting a conserved architecture which suggests a common origin or convergent evolution of Burkholderiales BGCs across *Cephalotes* species. Symbiotic convergent evolution has been reported in insect-bacteria symbiosis^{376,377}, showing insect symbionts retain the ability to produce a certain number of proteins and functions which have an importance for both the bacteria and the host survival. Further work focusing on the occurrence of

structurally similar BCGs in the genomes of environmental bacteria of the same order could reveal the origin and evolution of the BCGs in multipartite mutualisms.

A final remark concerns the division of labor in a multipartite mutualism. In honey bees the gut microbiota is dominated by five coevolved bacterial clusters. However the gut bacterial species have different abilities in digesting polysaccharides which can be explained by a divergence into different ecological niches inside the gut of their hosts³²⁴. In *Cephalotes* gut four of the five conserved symbionts possess the majority of the retrieved BGCs but Opitutales a major player in the *Cephalotes* nutritional symbiosis is mostly lacking BGCs. Although multiple bins of Opitutales were retrieved only two BGCs (arylpolyyene) were detected (Supplementary Figure 24). Opitutales are localized in the midgut whereas Burkholderiales, Pseudomonadales, Rhizobiales, and Xanthomonadales are found in the ileum of the hindgut³⁷⁸. This spatial isolation prevents direct interaction of Opitutales with other symbionts and supports the idea that BGCs have an essential role for symbionts living in close proximity.

Taken together our results address the diversity and maintenance of bacterial BGC in multipartite nutritional symbioses. In contrast to symbionts undergoing genome reduction, which keeps only the gene involved in supplementing the host nutrition, this work suggests a broader gene selection in the genomes of gut bacteria involved in multipartite mutualism. We uncover high similarities of BGCs encoding for biosynthesis of arylpolyenes, NRP, T1PK and siderophores in bacteria found across and within *Cephalotes* species. These metabolites potentially contribute to the symbionts' maintenance in complex bacterial communities and future works will hopefully confirm the functional role of specific metabolites in ant-gut bacteria symbioses.

Data availability

The metagenomic bins and the biosynthetic gene clusters detected in genomes and metagenomes are available in the Dryad Digital Repository (<https://datadryad.org/>) under <https://doi.org/10.5061/dryad.s7h44j168>.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

A.C., C.S.M., and C.D. conceived the study. A.C. performed the analyses. A.C., C.S.M., and C.D. interpreted the results. A.C. drafted the manuscript. A.C., C.S.M., and C.D. revised the manuscript.

Funding

This work received financial support from an “Investissement d’Avenir” grant managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-25-01) to C.D. and the National Science Foundation (NSF DEB 1900357) to C.S.M. We are grateful to the University of French Guiana for Anaïs Chanson’s Ph.D. fellowship and traveling grants.

Acknowledgments

The authors are grateful to Yi Hu and Jon Sanders for their knowledge and advice on bacterial metagenomes analysis and to Manuela Ramalho for her unfailing assistance with statistical analysis. The authors are thankful to Simon Shaw for his help with the detection of metagenomic biosynthetic gene clusters, and to Nelly Selem for her assistance with the CORASON analysis. We are grateful to reviewers’ comments of different journal rejections which have greatly improved the manuscript.

Supplementary Informations

L’ensemble des figures et tableaux des *Supplementary Informations* sont en Annexe 3.

III. Conclusion

L'objectif de ce chapitre 3 était d'analyser la diversité des clusters de gènes de biosynthèse des communautés bactériennes associées aux fourmis du genre *Cephalotes*. Nos résultats révèlent une grande diversité de types et de tailles de clusters de gènes de biosynthèse présents.

La grande majorité de ces clusters de gènes de biosynthèse provient de quatre ordres bactériens : Burkholderiales, Pseudomonadales, Xanthomonadales and Rhizobiales. Ces quatre ordres bactériens font partie des cinq ordres bactériens composant le noyau des communautés bactériennes associées aux fourmis *Cephalotes*¹⁸. Ces cinq ordres bactériens sont extrêmement conservés dans la phylogénie des fourmis *Cephalotes*, et ils ont un rôle nutritionnel fondamental en recyclant l'azote de l'urée en acides aminés essentiels pour leur hôte. Cela suggère qu'il est nécessaire que ces symbiotes puissent se maintenir dans les communautés bactériennes et être transmis entre générations.

La transmission des symbiotes entre générations chez les *Cephalotes* est réalisée par trophallaxie orale-anale³⁷⁴. Parmi les différents types de clusters de gènes de biosynthèse retrouvés dans ces bactéries, nous avons mis en évidence quatre types de clusters qui ont des fonctions connues et supposées pour la colonisation et la maintenance des symbiotes dans une communauté. Ces quatre types de clusters sont les arylpolyènes, qui pourraient jouer un rôle dans la colonisation des tissus⁷⁸, les peptides non ribosomiques et les polycétides, qui produisent des molécules à activité antimicrobiennes³⁶¹, ainsi que les sidérophores, qui permettent aux symbiotes de capter le fer de leur environnement⁷⁹. Nos résultats suggèrent donc que les symbiotes composant le noyau des communautés bactériennes associées aux fourmis *Cephalotes* possèdent des clusters de gènes de biosynthèse leur permettant de produire des métabolites secondaires jouant un rôle dans la colonisation et la maintenance des bactéries dans les communautés, ce qui permettrait à ce noyau de communautés bactériennes de se maintenir dans leur forte association avec les fourmis *Cephalotes* et effectuer leur rôle de supplémentation nutritionnelle pour ces fourmis.

Un autre aspect important révélé par nos résultats est la quasi-totale absence de clusters de gènes de biosynthèse chez les bactéries de l'ordre des Opitutales, malgré la prépondérance de ces bactéries

dans les génomes et métagénomes étudiés. Les bactéries Opituales appartiennent également au noyau des communautés bactériennes associées aux fourmis *Cephalotes* et participent à la supplémentation nutritionnelle des fourmis. Il serait donc logique que ces bactéries possèdent également une grande diversité de clusters de gènes de biosynthèse. Cependant, les Opituales sont situés dans l'intestin moyen des fourmis, alors que les quatre autres ordres bactériens composant le noyau des communautés bactériennes associées aux *Cephalotes* sont situés dans l'intestin postérieur³⁷⁸. Cette isolation spatiale des Opituales pourrait limiter leurs interactions avec d'autres communautés bactériennes, et donc diminuer leur besoin de produire des métabolites spécialisés pour se maintenir dans une communauté. Cette idée renforce notre hypothèse que certains clusters de gènes de biosynthèse trouvés dans les génomes et métagénomes du noyau des communautés bactériennes associées aux *Cephalotes* leur permet de produire des métabolites spécialisés jouant un rôle dans le maintien de ces communautés bactériennes dans leur association avec les *Cephalotes*.

Conclusion générale et perspectives

Ce travail de thèse visait à étudier les différents mécanismes impliqués dans la structuration et la maintenance des associations entre fourmis amazoniennes et leurs communautés microbiennes. Pour cela, nous avons utilisé des méthodologies différentes mais complémentaires dans notre étude.

Les objectifs des chapitres 1 et 2 étaient (1) d'étudier la diversité des communautés microbiennes associées aux fourmis amazoniennes, (2) de déterminer quels facteurs taxonomiques, écologiques et environnementaux de l'hôte jouent un rôle dans la structuration de ces communautés bactériennes, et (3) de rechercher d'éventuels patrons de co-occurrence entre communautés bactériennes et eucaryotes associées aux mêmes échantillons de fourmis. Pour cela, nous avons collecté 38 espèces différentes de fourmis amazoniennes dans la forêt et dans un environnement urbain, et nous avons classé ces fourmis dans différents groupes selon leur habitat, leur régime alimentaire, leur type de nid et leur taxonomie. Ensuite, nous avons utilisé la technique du métabarcoding en ciblant les gènes 16S rRNA et 18S rRNA qui sont des gènes marqueurs des procaryotes et des eucaryotes respectivement.

Nos résultats révèlent que (1) les communautés microbiennes sont différentes en fonction du type de nid de la fourmi, (2) les communautés bactériennes sont également différentes en fonction du régime alimentaire de la fourmi, (3) certains genres de fourmis ont des communautés bactériennes très conservées, particulièrement quand ces communautés bactériennes jouent un rôle dans l'écologie de la fourmi hôte, (4) les communautés microbiennes ne sont pas différentes en fonction de l'habitat ou de l'histoire évolutive de la fourmi, et (5) les motifs de co-occurrence entre communautés bactériennes et eucaryotes observées paraissent être structurés par la taxonomie de la fourmi plutôt que par un spécifique eucaryote.

Les résultats obtenus dans l'analyse des communautés bactériennes sont soutenus par la littérature, où d'autres études sur différents hôtes ont déjà montré que certaines communautés bactériennes sont spécifiques à un certain régime alimentaire^{65,230}, ou à un besoin écologique

spécifique¹⁸⁹. Ainsi, cette méthodologie pourrait être étendue à d'autres systèmes d'étude de communautés bactériennes associées à un hôte. En revanche, les résultats obtenus dans l'analyse des communautés eucaryotes sont à considérer avec beaucoup plus de prudence, car la majorité des séquences obtenues dans l'analyse n'ont pu être déterminées avec la base de données utilisée. A l'avenir, pour contourner cette limite deux possibilités s'offre à nous. La première est d'effectuer plus d'études sur les communautés eucaryotes associées à un hôte, en particulier pour les hôtes collectés en forêt ou dans les arbres car les fourmis de ces deux catégories étaient associées à la majorité des eucaryotes indéterminés dans notre étude. La seconde est d'utiliser l'approche du multi-marqueur métabarcoding, qui permettrait à la fois d'augmenter le nombre de séquences eucaryotes identifiées dans les échantillons, et de maximiser les possibilités d'identification de ces échantillons en utilisant plusieurs bases de données. En combinant ces deux possibilités, nous pourrions être capable d'identifier mieux les communautés eucaryotes associées à un hôte, et nous pourrions alors tester à nouveau le rôle des facteurs taxonomiques et écologiques de l'hôte dans la structuration de ces communautés, pour confirmer, renforcer ou infirmer les résultats présentés dans cette thèse. Malgré ces limites, nous avons tout de même identifié de nombreux eucaryotes connus pour être associés aux insectes, montrant l'intérêt de cette approche dans l'étude de la diversité et de la structuration des communautés eucaryotes associées à un hôte.

Les objectifs du chapitre 3 étaient d'étudier (1) la présence et (2) la diversité des clusters de gènes de biosynthèse des communautés bactériennes associées aux fourmis du genre *Cephalotes*. Les fourmis du genre *Cephalotes* sont un modèle intéressant pour cette étude pour plusieurs raisons. Tout d'abord, comme montré dans différentes études^{18,194,219} ainsi que dans le chapitre 1 de cette thèse, elles possèdent un noyau de communautés bactériennes majoritaires très conservé entre les espèces et composé de cinq ordres bactériens. Ensuite, ce noyau de communautés bactériennes joue un rôle majeur dans le recyclage d'azote en acides aminés essentiels pour l'hôte. Enfin, certains clusters de gènes de biosynthèse servent à la production de métabolites secondaires importants pour la maintenance des communautés microbiennes. Pour cette étude, nous avons utilisé les génomes et

métagénomiques de bactéries intestinales de différentes espèces de fourmis *Cephalotes* publiés récemment¹⁸. Ensuite, nous avons utilisé différents logiciels bio-informatiques pour étudier la diversité des clusters de gènes de biosynthèse présent dans ces génomes et métagénomiques.

Nos résultats révèlent que (1) les génomes et métagénomiques bactériens étudiés possèdent une grande diversité de types de clusters de gènes de biosynthèse, (2) la grande majorité de ces clusters provient de quatre des cinq ordres bactériens formant le noyau des communautés bactériennes des fourmis *Cephalotes*, (3) certains types de clusters identifiés dans ces ordres bactériens ont des fonctions connues et supposées pour la colonisation et la maintenance des symbiontes dans une communautés, et (4) le seul membre des cinq ordres bactériens formant le noyau des communautés bactériennes des fourmis *Cephalotes* qui est isolé spatialement des quatre autres ordres bactériens ne possède quasiment aucun cluster de gène de biosynthèse. Ces résultats suggèrent que certains clusters de gènes de biosynthèse trouvés dans les génomes et métagénomiques du noyau des communautés bactériennes associées aux *Cephalotes* leur permet de produire des métabolites spécialisés jouant un rôle dans le maintien de ces communautés bactériennes dans leur association avec les *Cephalotes*.

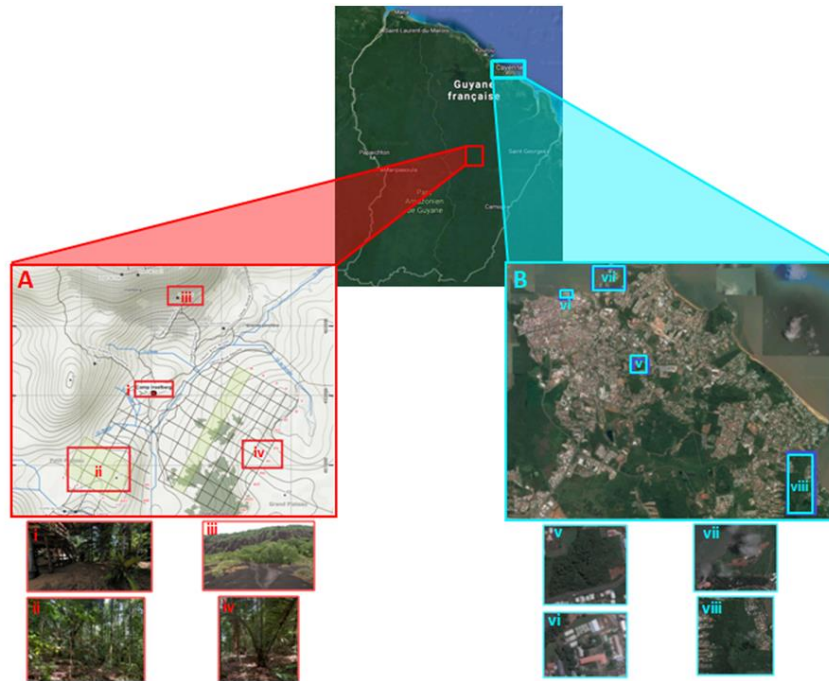
Ces résultats très encourageants montrent l'intérêt de la métagénomique dans l'analyse des clusters de gènes de biosynthèse des communautés bactériennes associées à un hôte. Idéalement, cette étude pourrait être reproduite et élargie aux différentes espèces des fourmis étudiés dans les chapitres 1 et 2, en particulier pour comparer la présence et la diversité de ces clusters entre les fourmis associées à des communautés bactériennes très conservées et les fourmis associées à des communautés bactériennes transitoires et très diversifiées. De plus, les champignons possédant également des clusters de gènes de biosynthèse, cette méthode pourrait être étendue à l'étude des clusters de gènes de biosynthèse des champignons associées à ces fourmis. Cependant, cette méthode étant très coûteuse en termes de temps et d'argent, cela la peut la rendre difficilement réalisable.

Une méthode alternative à la métagénomique est l'utilisation d'amorces de PCR ciblant spécifiquement des gènes essentiels dans certains types de clusters de gènes de biosynthèse. Cette méthode a été utilisée avec succès pour étudier la diversité des domaines KS et A dans les sols de

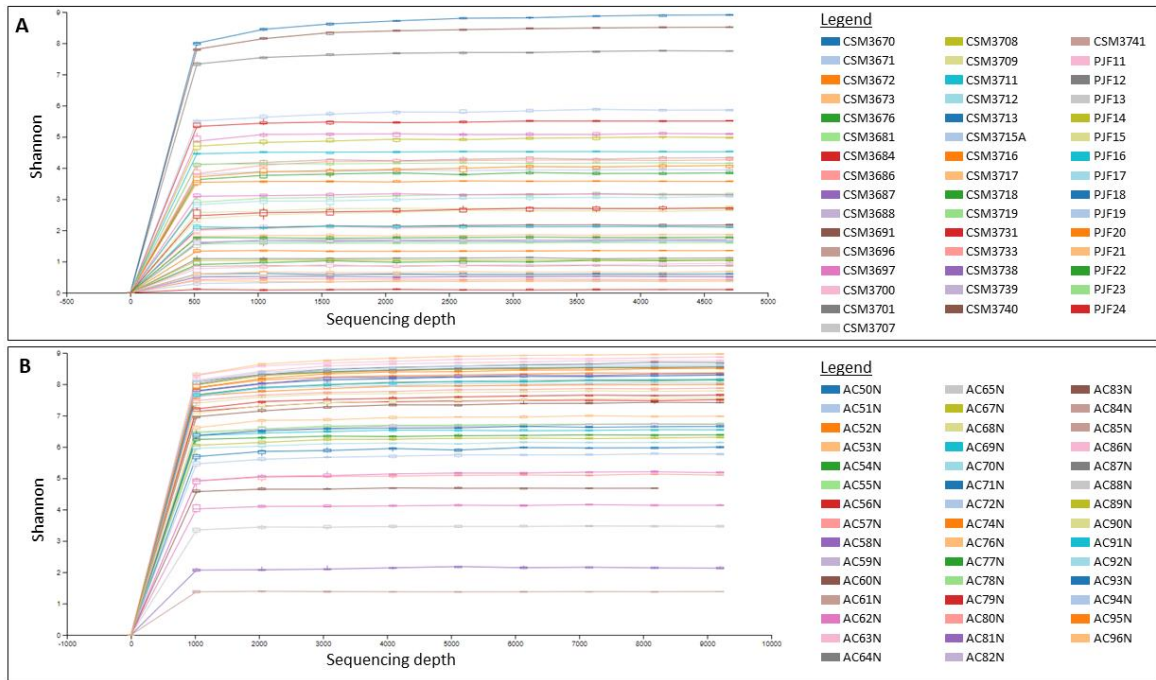
différents parcs New-Yorkais³⁷⁹. Le domaine KS est un gène essentiel dans les arylpolyènes et polycétides, tandis que le domaine A est un gène essentiel dans les peptides non-ribosomiques. En outre, ces trois types de cluster possèdent des fonctions connues et supposées pour la colonisation et la maintenance des symbiontes dans une communauté. Ainsi, cibler la présence et la diversité de ces deux gènes pourrait également être très intéressant dans un contexte d'étude de communautés microbiennes associées à un hôte. Nous avons au cours de cette thèse essayé d'appliquer cette méthode à nos échantillons de fourmis, mais malheureusement sans succès. Il apparaît donc nécessaire de passer plus de temps pour d'adapter cette méthode à dans notre système d'étude. Nous n'avons pas eu le temps de réaliser ces tests supplémentaires pendant cette thèse, mais cela reste une méthode intéressante à mettre en place dans de futures études dans un contexte d'étude de maintenance des communautés microbiennes associées à un hôte. D'autre la continuation de ce projet consistera à étudier le rôle fonctionnel de ces métabolites par des approches de co-cultures et de métabolomiques pour comprendre les interactions entre différentes communautés bactériennes symbiotiques des fourmis *Cephalotes*.

Annexes

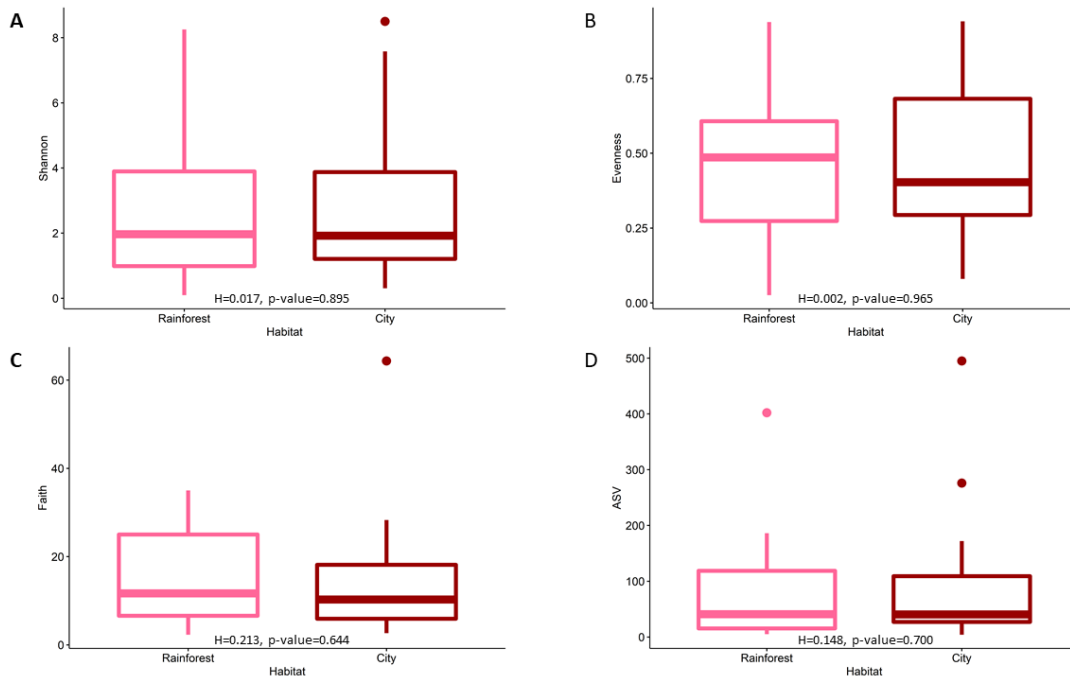
Annexe 1. Supporting Information de l'article Impact of Amazonian ants nesting mode, diet and taxonomy in structuring their bacterial communities



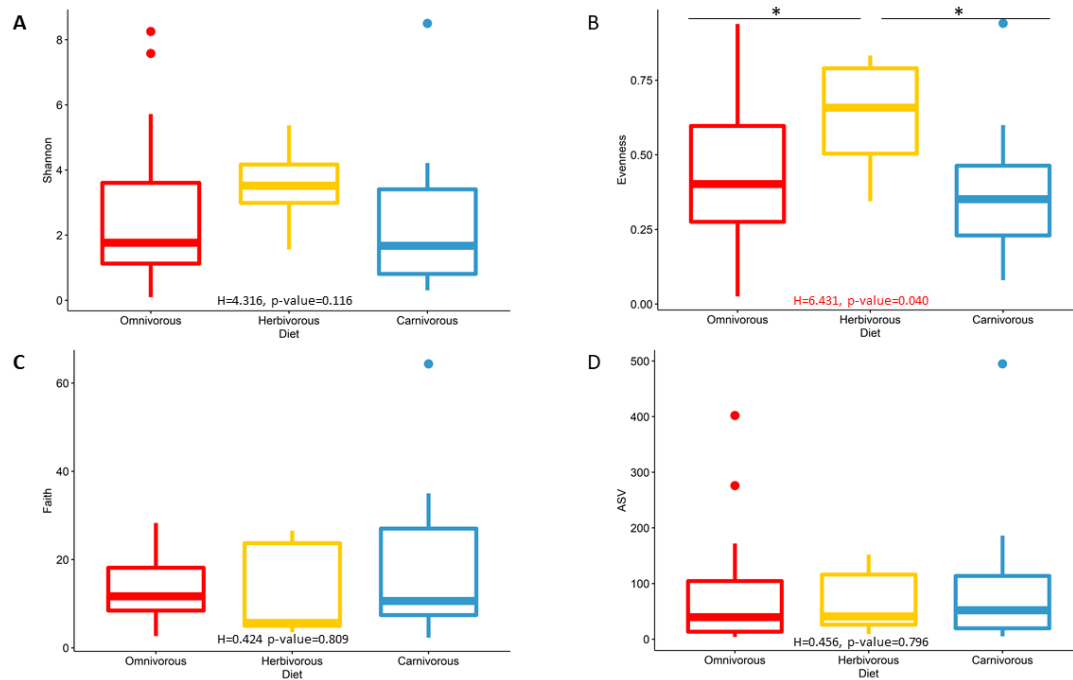
Supplementary Figure 1. Sampling Map. (A) Different sampling places in primary forest (Nouragues Field Station): (i) Inselberg camp, Inselberg little plateau (ii), Inselberg terrace (iii), Inselberg big plateau (iv). (B) Different sampling places in urban area (Cayenne): (v) Baduel hill, (vi) Pasteur Institute, (vii) Montabo trail, (viii) Loyola trail. The figure was made with the help of Peter Flynn. Satellites images are from Google Map.



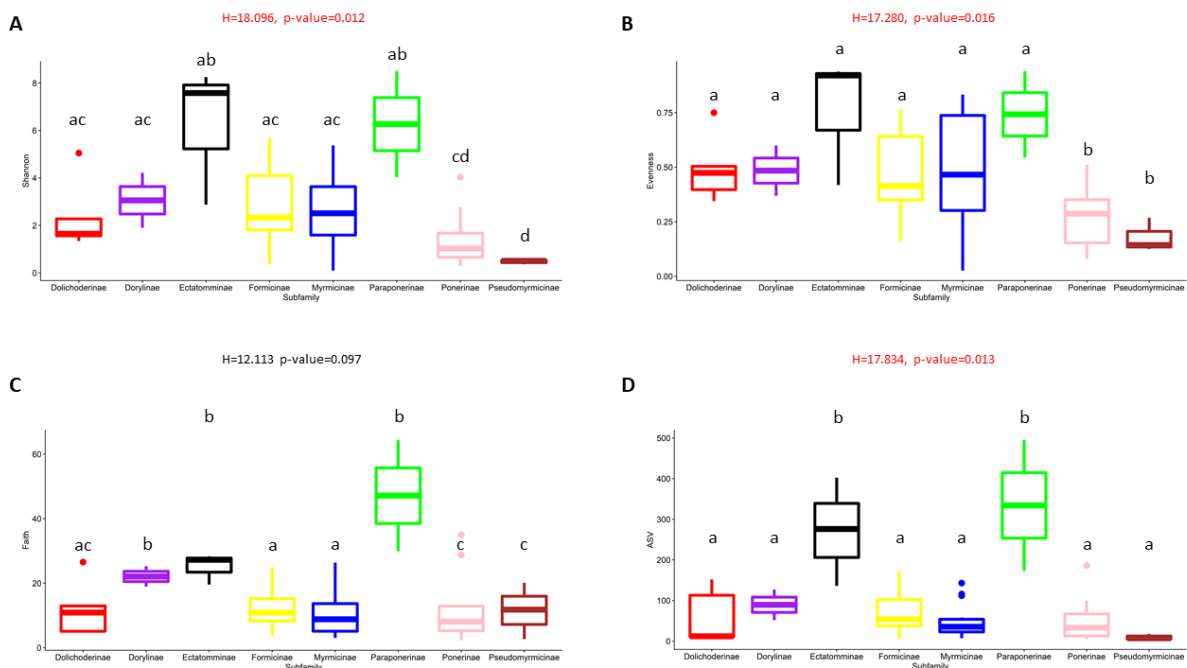
Supplementary Figure 2. Rarefaction curves used to estimate richness in the Shannon index for the ant dataset (A) and the nest dataset (B). The vertical axis shows the bacterial OTUs observed and the number of sequences per sample is shown on the horizontal axis. Our sequencing covers thousands of Illumina reads and samples have reached the plateau, indicating that the majority of the bacterial diversity has been recovered.



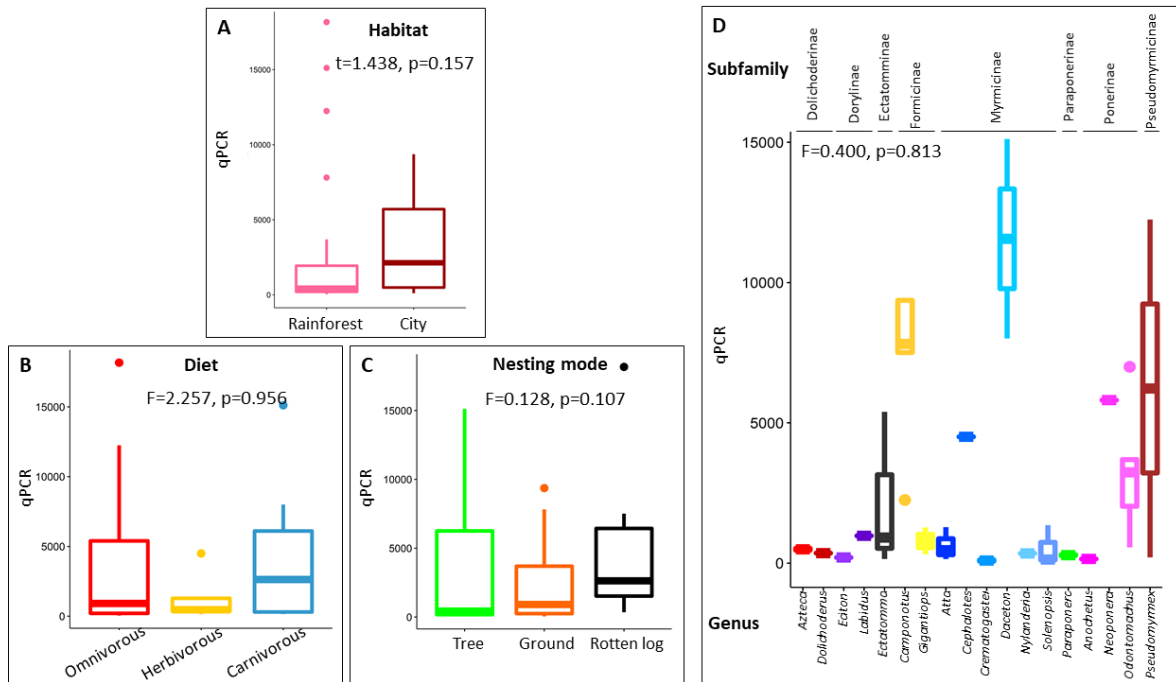
Supplementary Figure 3. Differences in bacterial alpha diversity by ant habitat. (A) Mean (\pm standard error) Shannon diversity. (B) Mean (\pm standard error) Pielou's evenness. (C) Mean (\pm standard error) Faith's phylogenetic diversity. (D) mean (\pm standard error) ASV count. The statistical p-values were obtained with pairwise Kruskal-Wallis. P-values for this figure are found in Supplementary File 2.



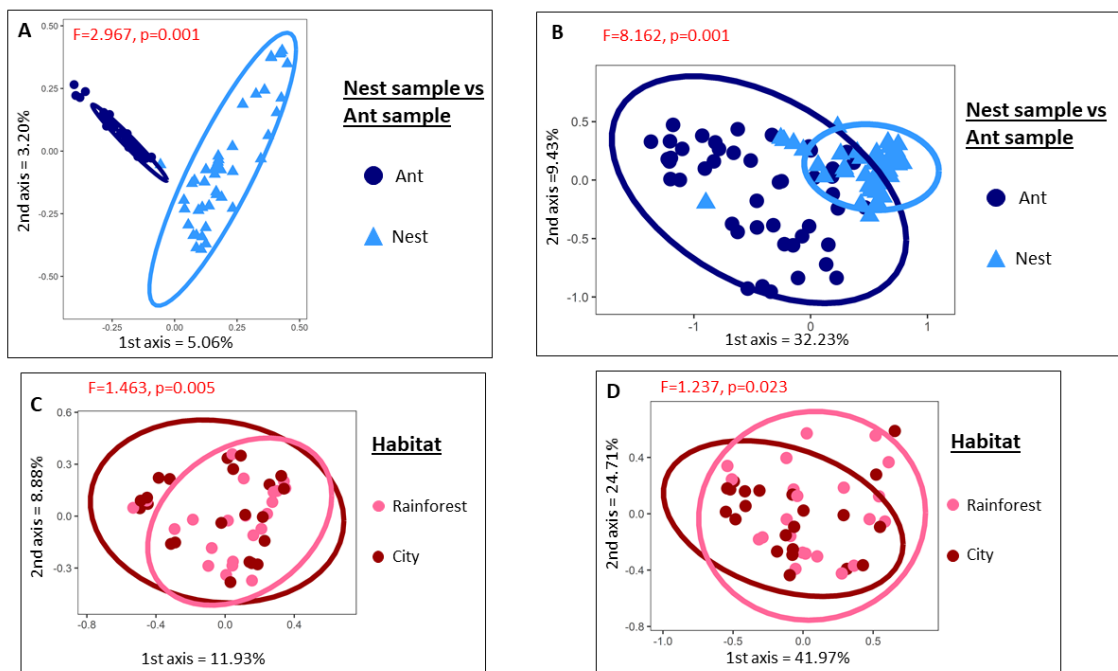
Supplementary Figure 4. Differences in bacterial alpha diversity by ant diet. (A) Mean (\pm standard error) Shannon diversity. (B) Mean (\pm standard error) Pielou's evenness. (C) Mean (\pm standard error) Faith's phylogenetic diversity. (D) mean (\pm standard error) ASV count. * <0.05 . The statistical p-values were obtained with pairwise Kruskal-Wallis. P-values for this figure are found in Supplementary File 2.



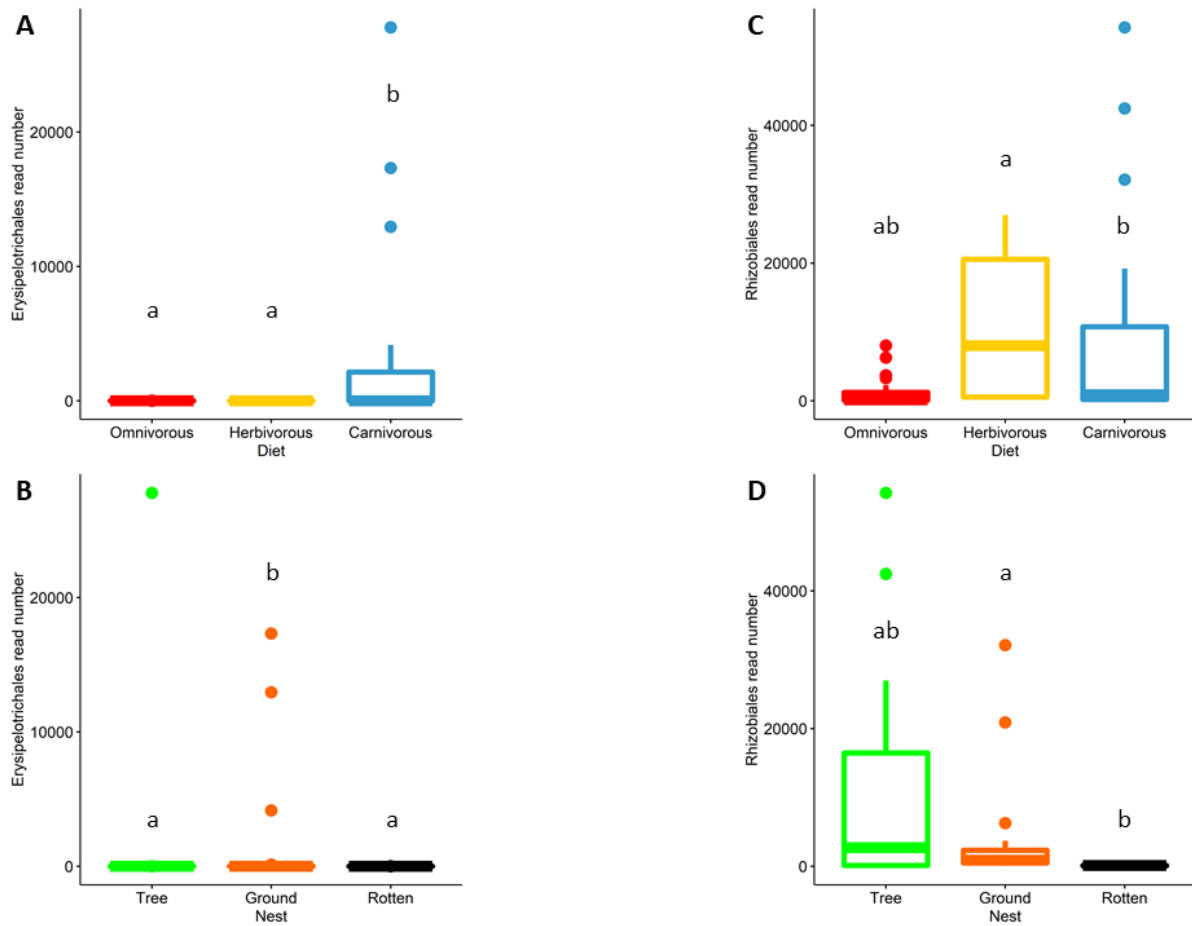
Supplementary Figure 5. Differences in bacterial alpha diversity by ant subfamily. (A) Mean (\pm standard error) Shannon diversity. (B) Mean (\pm standard error) Pielou's evenness. (C) Mean (\pm standard error) Faith's phylogenetic diversity. (D) mean (\pm standard error) ASV count. Different letters at the top of the figure illustrate subfamilies with significant differences ($P<0.05$) in this alpha diversity metric. The statistical p-values were obtained with pairwise Kruskal-Wallis. P-values for this figure are found in Supplementary File 2.



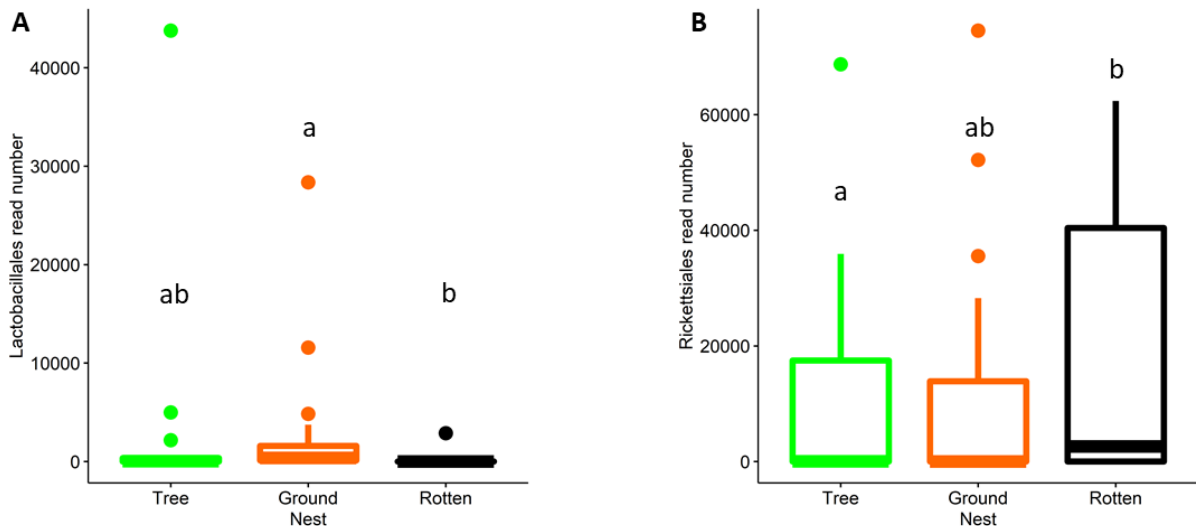
Supplementary Figure 6. Differences in mean qPCR values across subfamily and habitat (A), diet (B), nesting mode (C) and taxonomy (D). The statistical analyses were performed using t-test (A) and ANOVAs (B, C and D).



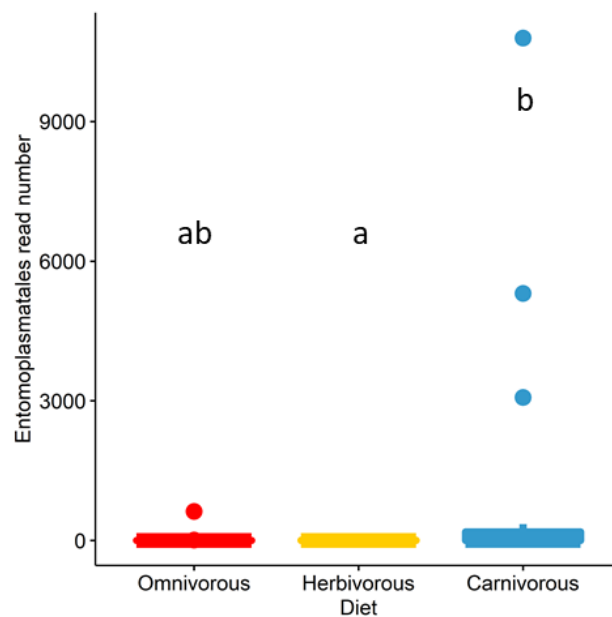
Supplementary Figure 7. Differences in bacterial communities. (A and B) Differences in bacterial communities between ants and their nest, using the Jaccard (A) or the unUniFrac distance (B). (C and D) Differences in bacterial communities associated with rainforest nests or city nests, using the Jaccard (C) or the unUniFrac distances (D). The statistical p-values were obtained with pairwise PERMANOVAs. The circles represent a confidence interval of 0.95. P-values for this figure are found in Supplementary File 4.



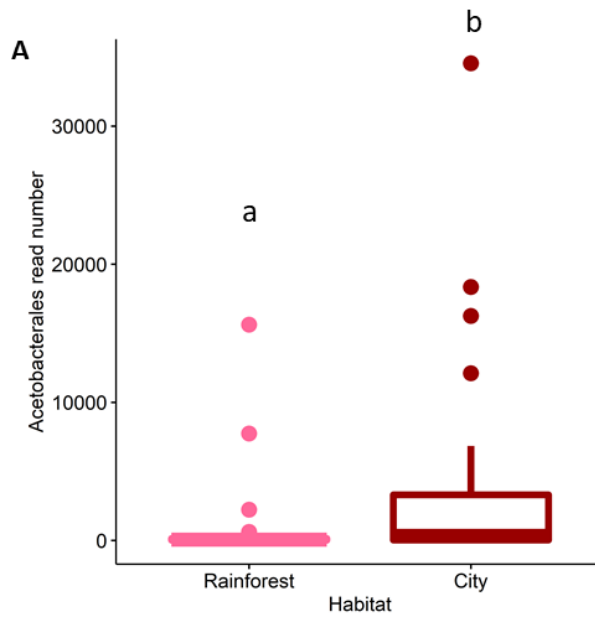
Supplementary Figure 8 SIMPER analysis using ASV abundance of different bacterial orders. (A and B) Mean (\pm standard error) Erysipelotrichales read number by ant diet (A) and nesting mode (B). (C and D) Mean (\pm standard error) Rhizobiales read number by ant diet (C) and nesting mode (D). Different letters at the top of the figure illustrate significant differences ($P < 0.05$) determined by Kruskal-Wallis. P-values for this figure are found in Supplementary File 7.



Supplementary Figure 9. SIMPER analysis using ASV abundance of different bacterial orders by ant nesting mode. (A) Mean (\pm standard error) Lactobacillales read number. (B) Mean (\pm standard error) Rickettsiales read number. Different letters at the top of the figure illustrate nesting mode with significant differences ($P < 0.05$) determined by Kruskal-Wallis. P-values for this figure are found in Supplementary File 7.



Supplementary Figure 10. Entomoplasmatales SIMPER analysis using ASV abundance by ant diet. Different letters at the top of the figure illustrate diet with significant differences ($P < 0.05$) determined by Kruskal-Wallis. P-values for this figure are found in Supplementary File 7.



Supplementary Figure 11. Acetobacterales SIMPER analysis using ASV abundance by ant habitat. Different letters at the top of the figure illustrate habitat with significant differences ($P < 0.05$) determined by Kruskal-Wallis. P-values for this figure are found in Supplementary File 7.

Supplementary File 1. List of collected samples.

Sample ID	Dataset	Corresponding nest ID	Corresponding AntID	Barcode sequence	Linker primer sequence	Barcode plate	Subfamily	Genus	Species	Sample type	Habitat mode	Habitat	Collection location	Diet	Nest	Nesting mode	qPCR	Qubit 1ng/1uL	Well
CSM3670	Ant	AC50N		AAGGCTG GGCAC	GTGTGYCA GCMGCCG CGGTAA	10	Ponerinae	Odontomachus	haematodus	Sample	City	City	Baduel	Carnivorous	Rotten log	Rotten	2027.91	1.111	B1
CSM3671	Ant	AC51N		CGATAGG CCTTA	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Cephalotes	minutus	Sample	City	City	Baduel	Herbivorous	Twig	Tree	4502.36	0.173	B2
CSM3672	Ant	AC52N		TAAGATG AGTC	GTGTGYCA GCMGCCG CGGTAA	10	Formicinae	Camponotus	sp1	Sample	City	City	Baduel	Omnivorous	NA	Ground	2247.18	1.411	B3
CSM3673	Ant	AC53N		TCCGCAAC CTGA	GTGTGYCA GCMGCCG CGGTAA	10	Paraponerinae	Paraponera	clavata	Sample	City	City	Baduel	Carnivorous	Ground	Ground	314.83	8.792	B4
CSM3676	Ant	AC54N		AGCCGAC TCTGT	GTGTGYCA GCMGCCG CGGTAA	10	Ectatomminae	Ectatomma	brunneum	Sample	City	City	Baduel	Omnivorous	Ground	Ground	909.98	1.320	B5
CSM3681	Ant	AC55N		GTGTTCC AGAA	GTGTGYCA GCMGCCG CGGTAA	10	Ponerinae	Odontomachus	hastatus	Sample	Rainforest	Rainforest	Camp	Carnivorous	Palm	Tree	109427.44	2.287	B6
CSM3684	Ant	AC56N		CTCGAGC GTACT	GTGTGYCA GCMGCCG CGGTAA	10	Formicinae	Camponotus	rapax	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Ground	Ground	7810.37	1.524	B7
CSM3686	Ant	AC57N		GCTGTGAT TCGA	GTGTGYCA GCMGCCG CGGTAA	10	Ectatomminae	Ectatomma	tuberculatum	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Ground	Ground	146.02	1.384	B8
CSM3687	Ant	AC58N		CTGCCCT GATG	GTGTGYCA GCMGCCG CGGTAA	10	Ponerinae	Odontomachus	scalptus	Sample	Rainforest	Rainforest	Petit plateau	Carnivorous	Ground	Ground	558.45	1.795	B9
CSM3688	Ant	AC59N		TTGGGTAC ACGT	GTGTGYCA GCMGCCG CGGTAA	10	Dolichoderinae	Dolichoderus	bispinosus	Sample	Rainforest	Rainforest	Petit plateau	Herbivorous	Palm	Tree	356.37	0.588	B10
CSM3691	Ant	AC60N		GATATCGA TGAA	GTGTGYCA GCMGCCG CGGTAA	10	Dolichoderinae	Dolichoderus	attelaboides	Sample	Rainforest	Rainforest	Petit plateau	Herbivorous	Palm	Tree	36888.05	1.533	B11

CSM3695 a	Ant	AC61N		AACGCGA AATTC	GTGTGYCA GCMGCCG CGGTAA	10	Formicinae	Camponotus	femoratus	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Ant garden	Tree	28058.12	3.700	B12
CSM3695 b	Ant	AC61N		GTGAGCA CGCAG	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Crematogaster	levior	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Ant garden	Tree	147.91	2.100	C1
CSM3696	Ant	AC62N		CGATCGA ACACT	GTGTGYCA GCMGCCG CGGTAA	10	Dolichoderinae	Azteca	sp	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Tree	Tree	559.36	0.100	C2
CSM3697	Ant	AC63N		TTCGCTAA CCTT	GTGTGYCA GCMGCCG CGGTAA	10	Formicinae	Camponotus	atriceps	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Rotten log	Rotten	18168.36	2.110	C3
CSM3700	Ant	AC64N		GCAGGGT CGAAC	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Solenopsis	virulens	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Rotten log	Rotten	1350.93	0.216	C4
CSM3701	Ant	AC65N		TAGAGGC GTAGG	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Crematogaster	limata	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	NA	Ground	38.01	0.100	C5
CSM3707	Ant			CAGGTAAT AATG	GTGTGYCA GCMGCCG CGGTAA	10	Dorylinae	Ecton	rapax	Sample	Rainforest	Rainforest	Grand plateau	Carnivorous	Bivouac	Ground	203.33	0.160	C6
CSM3708	Ant	AC67N		GCCGAAC GCCGA	GTGTGYCA GCMGCCG CGGTAA	10	Paraponerinae	Paraponera	clavata	Sample	Rainforest	Rainforest	Grand plateau	Carnivorous	Ground	Ground	252.79	1.407	C7
CSM3709	Ant	AC68N		GTCTAGTA TTTC	GTGTGYCA GCMGCCG CGGTAA	10	Ponerinae	Neoponera	commutata	Sample	Rainforest	Rainforest	Parare	Carnivorous	Ground	Ground	3996045.20	3.906	C8
CSM3711	Ant	AC69N		AAGCATTG AGAT	GTGTGYCA GCMGCCG CGGTAA	10	Pseudomyrmicinae	Pseudomyrmex	sp1	Sample	Rainforest	Rainforest	Terrace	Omnivorous	NA	Tree	202.28	0.022	C9
CSM3712	Ant	AC70N		CCGCTGAT GTCA	GTGTGYCA GCMGCCG CGGTAA	10	Ponerinae	Anochetus	emarginatus	Sample	Rainforest	Rainforest	Terrace	Carnivorous	Dead twig above ground	Tree	148.26	1.066	C10
CSM3713	Ant	AC71N		ACCCTCAG CCCA	GTGTGYCA GCMGCCG CGGTAA	10	Pseudomyrmicinae	Pseudomyrmex	gracilis	Sample	Rainforest	Rainforest camp	Camp	Omnivorous	Dead twig above ground	Tree	12245.70	2.355	C11
CSM3715 A	Ant	AC72N		CCGTCAA GATGT	GTGTGYCA GCMGCCG CGGTAA	10	Formicinae	Gigantiops	destructor	Sample	Rainforest	Rainforest	Camp	Omnivorous	Ground	Ground	NA	0.100	D1

CSM3716	Ant			GGTGACT AGTTC	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Cephalotes	atratus	Sample	Rainforest	Rainforest camp	Camp	Herbivorous	Tree	Tree	55993.14	1.306	C12
CSM3717	Ant	AC74N		GGAGTGC CCGAA	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Daceton	armigerum	Sample	Rainforest	Rainforest rocky savana	Mont	Omnivorous	Tree	Tree	15113.49	0.834	D3
CSM3718	Ant			CATGACG CCTCC	GTGTGYCA GCMGCCG CGGTAA	10	Dorylinae	Labidus	coecus	Sample	Rainforest	Rainforest rocky savana	Mont	Carnivorous	Bivouac	Ground	972.71	0.294	D4
CSM3719	Ant	AC76N		GTATTATG ACTT	GTGTGYCA GCMGCCG CGGTAA	10	Ponerinae	Odontomachus	haematodus	Sample	Rainforest	Rainforest rocky savana	Mont	Carnivorous	Ground	Ground	3691.90	1.599	D5
CSM3731	Ant	AC77N		CATAATCT TGCT	GTGTGYCA GCMGCCG CGGTAA	10	Dolichoderinae	Azteca	chartifex	Sample	Rainforest	Rainforest camp	Camp	Omnivorous	Carton	Tree	427.27	0.082	D6
CSM3733	Ant	AC78N		TGCGTGG GTGGA	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Cephalotes	pavonii	Sample	Rainforest	Rainforest	Terrace	Herbivorous	Twig	Tree	130878.47	0.336	D7
CSM3738	Ant	AC79N		ACTATATT AGCT	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Atta	cephalotes	Sample	Rainforest	Rainforest	Parare	Herbivorous	Ground	Ground	460.75	0.730	D8
CSM3739	Ant	AC80N		TTCTGGTC TTGT	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Daceton	armigerum	Sample	City	Forest city	Loyola	Omnivorous	Tree	Tree	8008.35	0.707	D9
CSM3740	Ant	AC81N		TCCAGGG CTATA	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Atta	cephalotes	Sample	City	Forest city	Montabo	Herbivorous	Ground	Ground	1282.31	1.816	D10
CSM3741	Ant	AC82N		GACTATAA TGGC	GTGTGYCA GCMGCCG CGGTAA	10	Ponerinae	Odontomachus	haematodus	Sample	City	Forest city	Montabo	Carnivorous	Rotten log	Rotten	3235.66	2.192	D11
PJF10	Ant	AC83N		CTATCATC CTCA	GTGTGYCA GCMGCCG CGGTAA	13	Myrmicinae	Solenopsis	geminata	Sample	City	City	Pasteur	Omnivorous	Ground	Ground	105.03	0.100	G9
PJF11	Ant	AC84N		AATTGAAT TTAC	GTGTGYCA GCMGCCG CGGTAA	10	Dolichoderinae	Azteca	chartifex	Sample	City	City	Pasteur	Omnivorous	Carton	Tree	32067.93	0.100	E1

PJF12	Ant	AC85N		AGGAAAAG CCAGA	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Atta	sexdens	Sample	City	City	Pasteur	Herbivorous	Ground	Ground	138.85	0.481	E2
PJF13	Ant	AC86N		ATGGCTGT CAGT	GTGTGYCA GCMGCCG CGGTAA	10	Formicinae	Camponotus	atriceps	Sample	City	City	Pasteur	Omnivorous	Rotten log	Rotten	7503.38	1.333	E3
PJF14	Ant	AC87N		GAGGGTT GTCGG	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Nylanderia	sp	Sample	City	City	Pasteur	Omnivorous	Rotten log	Rotten	345.53	0.100	E4
PJF15	Ant	AC88N		CCACGTAC GTAA	GTGTGYCA GCMGCCG CGGTAA	10	Ponerinae	Odontomachus	bori	Sample	City	City	Pasteur	Carnivorous	Ground	Ground	6998.29	0.929	E5
PJF16	Ant	AC89N		AGTAAGTC GATA	GTGTGYCA GCMGCCG CGGTAA	10	Formicinae	Camponotus	sp2	Sample	City	Forest city	Montabo	Omnivorous	Rotten log	Rotten	27158159.79	5.099	E6
PJF17	Ant	AC90N		TTGGGCC ACATA	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Cephalotes	atratus	Sample	City	Forest city	Loyola	Herbivorous	Tree	Tree	9772831.73	4.112	E7
PJF18	Ant			TCTATCTG GCTT	GTGTGYCA GCMGCCG CGGTAA	10	Formicinae	Gigantiops	destructor	Sample	City	Forest city	Loyola	Omnivorous	Ground	Ground	1268.32	1.161	E8
PJF19	Ant	AC91N		CCGTGGA CCAGG	GTGTGYCA GCMGCCG CGGTAA	10	Ponerinae	Neoponera	commutata	Sample	City	Forest city	Montabo	Carnivorous	Ground	Ground	5810.32	2.859	E12
PJF20	Ant	AC92N		GTCCCTAT TATC	GTGTGYCA GCMGCCG CGGTAA	10	Ectatomminae	Ectatomma	tuberculatum	Sample	City	Forest city	Loyola	Omnivorous	Ground	Ground	5391.95	0.124	F1
PJF21	Ant	AC93N		GCCTGCA GTACT	GTGTGYCA GCMGCCG CGGTAA	10	Pseudomyrmicinae	Pseudomyrmex	sp2	Sample	City	City	Pasteur	Omnivorous	Twig	Tree	NA	4.710	E10
PJF22	Ant	AC94N		ATACCTAA ATAT	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Solenopsis	sp	Sample	City	City	Pasteur	Omnivorous	Twig	Tree	127.12	0.887	E9
PJF23	Ant	AC95N		ATGGTTCA CCCCG	GTGTGYCA GCMGCCG CGGTAA	10	Formicinae	Camponotus	blandus	Sample	City	City	Pasteur	Omnivorous	Ground	Ground	9367.56	28.000	E11

PJF24	Ant	AC96N		GCGACGC GGCAT	GTGTGYCA GCMGCCG CGGTAA	10	Myrmicinae	Crematogaster	sp2	Sample	City	City	Pasteur	Omnivorous	Twig	Tree	NA	0.100	F2
negative1	Ant			CAGAGCT AATTG	GTGTGYCA GCMGCCG CGGTAA	18	Control	Control	Control	Control	Control	Control	Control	Control	Control	Control	931712.85	0.101	D1
negative2	Ant			ATATATAG TATC	GTGTGYCA GCMGCCG CGGTAA	15	Control	Control	Control	Control	Control	Control	Control	Control	Control	Control	303571.13	0.102	G5
AC50N	Nest		CSM3670	GTCCGCA AGTTA	GTGTGYCA GCMGCCG CGGTAA	15	Ponerinae	Odontomachus	haematodus	Sample	City	City	Baduel	Carnivorous	Rotten log	Rotten	42933.28	4.670	A1
AC51N	Nest		CSM3671	CAACACAT GCTG	GTGTGYCA GCMGCCG CGGTAA	15	Myrmicinae	Cephalotes	minutus	Sample	City	City	Baduel	Herbivorous	Twig	Tree	495.71	0.449	A2
AC52N	Nest		CSM3672	CATACCGT GAGT	GTGTGYCA GCMGCCG CGGTAA	15	Formicinae	Camponotus	sp1	Sample	City	City	Baduel	Omnivorous	NA	Ground	1392.04	0.337	A3
AC53N	Nest		CSM3673	GTCCATG GTTCCG	GTGTGYCA GCMGCCG CGGTAA	15	Paraponerinae	Paraponera	clavata	Sample	City	City	Baduel	Carnivorous	Ground	Ground	55219.30	11.800	A4
AC54N	Nest		CSM3676	ACCATTAC CATT	GTGTGYCA GCMGCCG CGGTAA	15	Ectatomminae	Ectatomma	brunneum	Sample	City	City	Baduel	Omnivorous	Ground	Ground	39084.65	8.650	A5
AC55N	Nest		CSM3681	TGGTAAG AGTCT	GTGTGYCA GCMGCCG CGGTAA	15	Ponerinae	Odontomachus	hastatus	Sample	Rainforest	Rainforest	Camp	Carnivorous	Palm	Tree	16722.33	4.880	A6
AC56N	Nest		CSM3684	CCAGCCTT CAGA	GTGTGYCA GCMGCCG CGGTAA	15	Formicinae	Camponotus	rapax	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Ground	Ground	43114.00	10.100	A7
AC57N	Nest		CSM3686	ATTCAGAT GGCA	GTGTGYCA GCMGCCG CGGTAA	15	Ectatomminae	Ectatomma	tuberculatum	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Ground	Ground	27546.76	3.270	A8
AC58N	Nest		CSM3687	GTATAGTC CGTG	GTGTGYCA GCMGCCG CGGTAA	12	Ponerinae	Odontomachus	scalptus	Sample	Rainforest	Rainforest	Petit plateau	Carnivorous	Ground	Ground	76439.16	14.300	F1

AC59N	Nest		CSM3688	CCTGTAG GTTGC	GTGTGYCA GCMGCCG CGGTAA	12	Dolichoderinae	Dolichoderus	bispinosus	Sample	Rainforest	Rainforest	Petit plateau	Herbivorous	Palm	Tree	777166.81	45.100	F2
AC60N	Nest		CSM3691	TACGTCGT TATT	GTGTGYCA GCMGCCG CGGTAA	12	Dolichoderinae	Dolichoderus	attelaboides	Sample	Rainforest	Rainforest	Petit plateau	Herbivorous	Palm	Tree	15488.15	1.560	F3
AC61N	Nest		CSM3695a	GAGTAGG GCACC	GTGTGYCA GCMGCCG CGGTAA	12	Formicinae	Camponotus	femoratus	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Ant garden	Tree	410.25	0.199	F4
AC61N	Nest		CSM3695b	GAGTAGG GCACC	GTGTGYCA GCMGCCG CGGTAA	12	Myrmicinae	Crematogaster	levior	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Ant garden	Tree	410.25	0.199	F4
AC62N	Nest		CSM3696	CTACTAGC GGTA	GTGTGYCA GCMGCCG CGGTAA	15	Dolichoderinae	Azteca	sp	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Tree	Tree	146182.38	28.100	B1
AC63N	Nest		CSM3697	CGATTAG GCCA	GTGTGYCA GCMGCCG CGGTAA	15	Formicinae	Camponotus	atriceps	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Rotten log	Rotten	991.21	0.758	B2
AC64N	Nest		CSM3700	GCTTGGTA GGTT	GTGTGYCA GCMGCCG CGGTAA	15	Myrmicinae	Solenopsis	virulens	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Rotten log	Rotten	335.11	0.531	B3
AC65N	Nest		CSM3701	AGGCGCT CTCCT	GTGTGYCA GCMGCCG CGGTAA	15	Myrmicinae	Crematogaster	limata	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	NA	Ground	4921.04	0.954	B4
AC67N	Nest		CSM3708	ACCTGATC CGCA	GTGTGYCA GCMGCCG CGGTAA	15	Paraponerinae	Paraponera	clavata	Sample	Rainforest	Rainforest	Grand plateau	Carnivorous	Ground	Ground	5980.62	2.550	B5
AC68N	Nest		CSM3709	GAGATTA AGCA	GTGTGYCA GCMGCCG CGGTAA	15	Ponerinae	Neoponera	commutata	Sample	Rainforest	Rainforest	Parare	Carnivorous	Ground	Ground	6150.59	2.550	B6
AC69N	Nest		CSM3711	TGGGTCC CACAT	GTGTGYCA GCMGCCG CGGTAA	15	Pseudomyrmicinae	Pseudomyrmex	sp1	Sample	Rainforest	Rainforest	Terrace	Omnivorous	NA	Tree	1474.96	0.525	B7
AC70N	Nest		CSM3712	ATTCTGCC GAAG	GTGTGYCA GCMGCCG CGGTAA	15	Ponerinae	Anochetus	emarginatus	Sample	Rainforest	Rainforest	Terrace	Carnivorous	Dead twig above ground	Tree	34987.56	7.200	B8

AC71N	Nest		CSM3713	CATTACTT ATGC	GTGTGYCA GCMGCCG CGGTAA	12	Pseudomyrmicinae	Pseudomyrmex	gracilis	Sample	Rainforest	Rainforest camp	Camp	Omnivorous	Dead twig above ground	Tree	32014.97	3.420	F5
AC72N	Nest		CSM3715A	CACGTCTG CGGAG	GTGTGYCA GCMGCCG CGGTAA	12	Formicinae	Gigantiops	destructor	Sample	Rainforest	Rainforest	Camp	Omnivorous	Ground	Ground	3011.89	3.440	F6
AC74N	Nest		CSM3717	CGTTACCG GACT	GTGTGYCA GCMGCCG CGGTAA	12	Myrmicinae	Daceton	armigerum	Sample	Rainforest	Rainforest rocky savana	Mont	Omnivorous	Tree	Tree	62417.44	3.260	F7
AC76N	Nest		CSM3719	GGACTCA ACTAA	GTGTGYCA GCMGCCG CGGTAA	12	Ponerinae	Odontomachus	haematodus	Sample	Rainforest	Rainforest rocky savana	Mont	Carnivorous	Ground	Ground	26203.48	3.860	F8
AC77N	Nest		CSM3731	ATGAAGC ACTGT	GTGTGYCA GCMGCCG CGGTAA	15	Dolichoderinae	Azteca	chartifex	Sample	Rainforest	Rainforest camp	Camp	Omnivorous	Carton	Tree	54321.97	6.210	D1
AC78N	Nest		CSM3733	TCGTCAG GACC	GTGTGYCA GCMGCCG CGGTAA	15	Myrmicinae	Cephalotes	pavonii	Sample	Rainforest	Rainforest	Terrace	Herbivorous	Twig	Tree	3362.22	0.538	D2
AC79N	Nest		CSM3738	CGTCGATT GCAC	GTGTGYCA GCMGCCG CGGTAA	15	Myrmicinae	Atta	cephalotes	Sample	Rainforest	Rainforest	Parare	Herbivorous	Ground	Ground	330662.52	21.300	D3
AC80N	Nest		CSM3739	CGAACAG AGCGC	GTGTGYCA GCMGCCG CGGTAA	15	Myrmicinae	Daceton	armigerum	Sample	City	Forest city	Loyola	Omnivorous	Tree	Tree	168836.52	9.190	D4
AC81N	Nest		CSM3740	GCCGACG TGCCG	GTGTGYCA GCMGCCG CGGTAA	15	Myrmicinae	Atta	cephalotes	Sample	City	Forest city	Montabo	Herbivorous	Ground	Ground	28851.52	4.610	D5
AC82N	Nest		CSM3741	GTGCTGC GCTTA	GTGTGYCA GCMGCCG CGGTAA	15	Ponerinae	Odontomachus	haematodus	Sample	City	Forest city	Montabo	Carnivorous	Rotten log	Rotten	542018.66	29.200	D6
AC83N	Nest		PJF10	TACGCAG CACTA	GTGTGYCA GCMGCCG CGGTAA	15	Myrmicinae	Solenopsis	geminata	Sample	City	City	Pasteur	Omnivorous	Ground	Ground	303625.83	12.500	D7
AC84N	Nest		PJF11	TGACAACC GAAT	GTGTGYCA GCMGCCG CGGTAA	15	Dolichoderinae	Azteca	chartifex	Sample	City	City	Pasteur	Omnivorous	Carton	Tree	164.38	0.309	D8

AC85N	Nest		PJF12	GCCGGTA CTCTA	GTGTGYCA GCMGCCG CGGTAA	15	Myrmicinae	Atta	sexdens	Sample	City	City	Pasteur	Herbivorous	Ground	Ground	18689.52	1.380	D9
AC86N	Nest		PJF13	CGTTAAGT CAGC	GTGTGYCA GCMGCCG CGGTAA	12	Formicinae	Camponotus	atriceps	Sample	City	City	Pasteur	Omnivorous	Rotten log	Rotten	200753.90	7.210	F9
AC87N	Nest		PJF14	TTATAGAT CTCC	GTGTGYCA GCMGCCG CGGTAA	12	Myrmicinae	Nylanderia	sp	Sample	City	City	Pasteur	Omnivorous	Rotten log	Rotten	141260.50	4.760	F10
AC88N	Nest		PJF15	GGTAGCG ACGCC	GTGTGYCA GCMGCCG CGGTAA	12	Ponerinae	Odontomachus	bori	Sample	City	City	Pasteur	Carnivorous	Ground	Ground	160027.60	7.130	F11
AC89N	Nest		PJF16	CGGACTC GTTAC	GTGTGYCA GCMGCCG CGGTAA	15	Formicinae	Camponotus	sp2	Sample	City	Forest city	Montabo	Omnivorous	Rotten log	Rotten	1548645.00	18.000	E1
AC90N	Nest		PJF17	CAGACGA GGAAC	GTGTGYCA GCMGCCG CGGTAA	15	Myrmicinae	Cephalotes	atratus	Sample	City	Forest city	Loyola	Herbivorous	Tree	Tree	201516.64	5.610	E2
AC91N	Nest		PJF19	CTCGATGT AAGC	GTGTGYCA GCMGCCG CGGTAA	15	Ponerinae	Neoponera	commutata	Sample	City	Forest city	Montabo	Carnivorous	Ground	Ground	14781.64	0.797	E3
AC92N	Nest		PJF20	TCTGAATG GTAG	GTGTGYCA GCMGCCG CGGTAA	15	Ectatomminae	Ectatomma	tuberculatum	Sample	City	Forest city	Loyola	Omnivorous	Ground	Ground	91209.69	7.020	E4
AC93N	Nest		PJF21	CAGGGAG GATCC	GTGTGYCA GCMGCCG CGGTAA	15	Pseudomyrmicinae	Pseudomyrmex	sp2	Sample	City	City	Pasteur	Omnivorous	Twig	Tree	84747.32	4.380	E5
AC94N	Nest		PJF22	GGACTATC GTTG	GTGTGYCA GCMGCCG CGGTAA	15	Myrmicinae	Solenopsis	sp	Sample	City	City	Pasteur	Omnivorous	Twig	Tree	113402.27	4.140	E6
AC95N	Nest		PJF23	TTCGCAGA TACG	GTGTGYCA GCMGCCG CGGTAA	15	Formicinae	Camponotus	blandus	Sample	City	City	Pasteur	Omnivorous	Ground	Ground	280176.57	5.920	E7
AC96N	Nest		PJF24	AACACATG GGTT	GTGTGYCA GCMGCCG CGGTAA	15	Myrmicinae	Crematogaster	sp2	Sample	City	City	Pasteur	Omnivorous	Twig	Tree	29890.24	2.980	E8

T-Kit1	Nest			CACGAGC TACTC	GTGTGYCA GCMGCCG CGGTAA	12	Control	Control	Control	Control	Control	Control	Control	Control	Control	Control	27.31	0.211	F12
T-Kit2	Nest			ATGAGAT GATAC	GTGTGYCA GCMGCCG CGGTAA	12	Control	Control	Control	Control	Control	Control	Control	Control	Control	Control	27.04	0.134	G1

Supplementary File 2. Statistic calculations of alpha diversity metrics for each factor. The four tested alpha diversity metrics are Shannon diversity, Pielou's evenness, Faith's PD and ASV richness. Each metric is calculated for the whole dataset and the forest/city dataset. The cells colored in red in each table represent p-values < 0.05. The statistical p-values were obtained with pairwise Kruskal-Wallis.

Shannon diversity whole dataset

Factor	Group 1	Group 2	H	p-value	q-value
Subfamily	Dolichoderinae (n=5)	Dorylinae (n=2)	0.600	0.439	0.559
		Ectatomminae (n=3)	3.756	0.053	0.167
		Formicinae (n=8)	0.536	0.464	0.565
		Myrmicinae (n=14)	0.214	0.643	0.630
		Paraponerinae (n=2)	2.400	0.121	0.184
		Ponerinae (n=9)	2.351	0.125	0.184
		Pseudomyrmicinae (n=3)	5.000	0.025	0.167
	Dorylinae (n=2)	Ectatomminae (n=3)	1.333	0.248	0.347
		Formicinae (n=8)	0.068	0.794	0.794
		Myrmicinae (n=14)	0.403	0.525	0.613
		Paraponerinae (n=2)	0.600	0.439	0.559
		Ponerinae (n=9)	2.722	0.099	0.178
		Pseudomyrmicinae (n=3)	3.000	0.083	0.167
	Ectatomminae (n=3)	Formicinae (n=8)	3.375	0.066	0.167
		Myrmicinae (n=14)	3.111	0.078	0.078
		Paraponerinae (n=2)	0.333	0.564	0.564
		Ponerinae (n=9)	5.342	0.021	0.021
		Pseudomyrmicinae (n=3)	3.857	0.049	0.049
	Formicinae (n=8)	Myrmicinae (n=14)	0.116	0.733	0.733
		Paraponerinae (n=2)	2.455	0.117	0.117
		Ponerinae (n=9)	2.676	0.102	0.102
		Pseudomyrmicinae (n=3)	4.167	0.042	0.167
	Myrmicinae (n=14)	Paraponerinae (n=2)	3.050	0.081	0.167
		Ponerinae (n=9)	3.571	0.059	0.167
		Pseudomyrmicinae (n=3)	5.143	0.023	0.167
	Paraponerinae (n=2)	Ponerinae (n=9)	4.500	0.034	0.167
		Pseudomyrmicinae (n=3)	3.000	0.083	0.167
Ponerinae (n=9)	Pseudomyrmicinae (n=3)	3.769	0.052	0.167	
Genus	<i>Anochetus</i> (n=1)	<i>Atta</i> (n=3)	1.800	0.180	0.388
		<i>Azteca</i> (n=3)	0.200	0.655	0.795
		<i>Camponotus</i> (n=6)	0.250	0.617	0.780
		<i>Cephalotes</i> (n=4)	2.000	0.157	0.388
		<i>Crematogaster</i> (n=2)	1.500	0.221	0.388
		<i>Daceton</i> (n=2)	0.000	1.000	1.000
		<i>Dolichoderus</i> (n=2)	0.000	1.000	1.000
		<i>Eciton</i> (n=1)	1.000	0.317	0.462

		<i>Ectatomma (n=3)</i>	1.800	0.180	0.388
		<i>Gigantiops (n=2)</i>	1.500	0.221	0.388
		<i>Labidus (n=1)</i>	1.000	0.317	0.462
		<i>Neoponera (n=2)</i>	0.000	1.000	1.000
		<i>Nylanderia (n=1)</i>	1.000	0.317	0.462
		<i>Odontomachus (n=6)</i>	1.000	0.317	0.462
		<i>Paraponera (n=2)</i>	1.500	0.221	0.388
		<i>Pseudomyrmex (n=3)</i>	1.800	0.180	0.388
		<i>Solenopsis (n=2)</i>	1.500	0.221	0.388
	<i>Atta (n=3)</i>	<i>Azteca (n=3)</i>	1.190	0.275	0.443
		<i>Camponotus (n=6)</i>	4.267	0.039	0.335
		<i>Cephalotes (n=4)</i>	0.125	0.724	0.863
		<i>Crematogaster (n=2)</i>	3.000	0.083	0.335
		<i>Daceton (n=2)</i>	3.000	0.083	0.335
		<i>Dolichoderus (n=2)</i>	3.000	0.083	0.335
		<i>Eciton (n=1)</i>	0.200	0.655	0.795
		<i>Ectatomma (n=3)</i>	0.429	0.513	0.682
		<i>Gigantiops (n=2)</i>	1.333	0.248	0.408
		<i>Labidus (n=1)</i>	1.800	0.180	0.388
		<i>Neoponera (n=2)</i>	3.000	0.083	0.335
		<i>Nylanderia (n=1)</i>	1.800	0.180	0.388
		<i>Odontomachus (n=6)</i>	3.267	0.071	0.335
		<i>Paraponera (n=2)</i>	1.333	0.248	0.408
		<i>Pseudomyrmex (n=3)</i>	3.857	0.049	0.335
		<i>Solenopsis (n=2)</i>	3.000	0.083	0.335
		<i>Azteca (n=3)</i>	<i>Camponotus (n=6)</i>	0.000	1.000
	<i>Cephalotes (n=4)</i>		0.500	0.480	0.644
	<i>Crematogaster (n=2)</i>		0.333	0.564	0.719
	<i>Daceton (n=2)</i>		0.333	0.564	0.719
	<i>Dolichoderus (n=2)</i>		0.000	1.000	1.000
	<i>Eciton (n=1)</i>		0.200	0.655	0.795
	<i>Ectatomma (n=3)</i>		2.333	0.127	0.387
	<i>Gigantiops (n=2)</i>		1.333	0.248	0.408
	<i>Labidus (n=1)</i>		0.200	0.655	0.795
	<i>Neoponera (n=2)</i>		0.333	0.564	0.719
	<i>Nylanderia (n=1)</i>		1.800	0.180	0.388
	<i>Odontomachus (n=6)</i>		3.267	0.071	0.335
	<i>Paraponera (n=2)</i>		1.333	0.248	0.408
	<i>Pseudomyrmex (n=3)</i>		3.857	0.049	0.335
	<i>Solenopsis (n=2)</i>	1.333	0.248	0.408	
	<i>Camponotus (n=6)</i>	<i>Cephalotes (n=4)</i>	4.545	0.033	0.335
		<i>Crematogaster (n=2)</i>	0.111	0.739	0.863
		<i>Daceton (n=2)</i>	1.000	0.317	0.462
		<i>Dolichoderus (n=2)</i>	0.111	0.739	0.863
		<i>Eciton (n=1)</i>	2.250	0.134	0.388
		<i>Ectatomma (n=3)</i>	4.267	0.039	0.335
		<i>Gigantiops (n=2)</i>	4.000	0.046	0.335

	<i>Labidus (n=1)</i>	0.000	1.000	1.000
	<i>Neoponera (n=2)</i>	0.000	1.000	1.000
	<i>Nylanderia (n=1)</i>	1.000	0.317	0.463
	<i>Odontomachus (n=6)</i>	1.256	0.262	0.463
	<i>Paraponera (n=2)</i>	4.000	0.046	0.335
	<i>Pseudomyrmex (n=3)</i>	3.267	0.071	0.335
	<i>Solenopsis (n=2)</i>	1.778	0.182	0.388
<i>Cephalotes (n=4)</i>	<i>Crematogaster (n=2)</i>	3.429	0.064	0.335
	<i>Daceton (n=2)</i>	3.429	0.064	0.335
	<i>Dolichoderus (n=2)</i>	3.429	0.064	0.335
	<i>Eciton (n=1)</i>	0.500	0.480	0.644
	<i>Ectatomma (n=3)</i>	0.500	0.480	0.644
	<i>Gigantiops (n=2)</i>	3.429	0.064	0.335
	<i>Labidus (n=1)</i>	2.000	0.157	0.388
	<i>Neoponera (n=2)</i>	3.429	0.064	0.335
	<i>Nylanderia (n=1)</i>	2.000	0.157	0.388
	<i>Odontomachus (n=6)</i>	4.545	0.033	0.335
	<i>Paraponera (n=2)</i>	0.857	0.355	0.512
	<i>Pseudomyrmex (n=3)</i>	4.500	0.034	0.335
	<i>Solenopsis (n=2)</i>	3.429	0.064	0.335
<i>Crematogaster (n=2)</i>	<i>Daceton (n=2)</i>	0.600	0.439	0.605
	<i>Dolichoderus (n=2)</i>	0.000	1.000	1.000
	<i>Eciton (n=1)</i>	1.500	0.221	0.388
	<i>Ectatomma (n=3)</i>	3.000	0.083	0.335
	<i>Gigantiops (n=2)</i>	2.400	0.121	0.379
	<i>Labidus (n=1)</i>	0.000	1.000	1.000
	<i>Neoponera (n=2)</i>	0.000	1.000	1.000
	<i>Nylanderia (n=1)</i>	1.500	0.221	0.388
	<i>Odontomachus (n=6)</i>	1.778	0.182	0.388
	<i>Paraponera (n=2)</i>	2.400	0.121	0.379
	<i>Pseudomyrmex (n=3)</i>	3.000	0.006	0.335
<i>Solenopsis (n=2)</i>	2.400	0.121	0.379	
<i>Daceton (n=2)</i>	<i>Dolichoderus (n=2)</i>	0.600	0.439	0.605
	<i>Eciton (n=1)</i>	1.500	0.221	0.388
	<i>Ectatomma (n=3)</i>	3.000	0.083	0.335
	<i>Gigantiops (n=2)</i>	2.400	0.121	0.379
	<i>Labidus (n=1)</i>	1.500	0.221	0.388
	<i>Neoponera (n=2)</i>	0.000	1.000	1.000
	<i>Nylanderia (n=1)</i>	0.000	1.000	1.000
	<i>Odontomachus (n=6)</i>	0.111	0.739	0.863
	<i>Paraponera (n=2)</i>	2.400	0.121	0.379
	<i>Pseudomyrmex (n=3)</i>	3.000	0.083	0.335
	<i>Solenopsis (n=2)</i>	0.600	0.439	0.605
<i>Dolichoderus (n=2)</i>	<i>Eciton (n=1)</i>	1.500	0.221	0.388
	<i>Ectatomma (n=3)</i>	3.000	0.083	0.335
	<i>Gigantiops (n=2)</i>	2.400	0.121	0.379
	<i>Labidus (n=1)</i>	0.000	1.000	1.000

	<i>Neoponera (n=2)</i>	0.000	1.000	1.000
	<i>Nylanderia (n=1)</i>	1.500	0.221	0.388
	<i>Odontomachus (n=6)</i>	1.778	0.182	0.388
	<i>Paraponera (n=2)</i>	2.400	0.121	0.379
	<i>Pseudomyrmex (n=3)</i>	3.000	0.083	0.335
	<i>Solenopsis (n=2)</i>	0.600	0.439	0.605
<i>Eciton (n=1)</i>	<i>Ectatomma (n=3)</i>	0.200	0.655	0.795
	<i>Gigantiops (n=2)</i>	1.500	0.221	0.388
	<i>Labidus (n=1)</i>	1.000	0.317	0.462
	<i>Neoponera (n=2)</i>	1.500	0.221	0.388
	<i>Nylanderia (n=1)</i>	1.000	0.317	0.462
	<i>Odontomachus (n=6)</i>	2.250	0.134	0.388
	<i>Paraponera (n=2)</i>	0.000	1.000	1.000
	<i>Pseudomyrmex (n=3)</i>	1.800	0.180	0.388
	<i>Solenopsis (n=2)</i>	1.500	0.221	0.388
<i>Ectatomma (n=3)</i>	<i>Gigantiops (n=2)</i>	0.333	0.564	0.719
	<i>Labidus (n=1)</i>	1.800	0.180	0.388
	<i>Neoponera (n=2)</i>	3.000	0.083	0.335
	<i>Nylanderia (n=1)</i>	1.800	0.180	0.388
	<i>Odontomachus (n=6)</i>	4.267	0.039	0.335
	<i>Paraponera (n=2)</i>	0.333	0.564	0.719
	<i>Pseudomyrmex (n=3)</i>	3.857	0.049	0.335
	<i>Solenopsis (n=2)</i>	3.000	0.083	0.335
<i>Gigantiops (n=2)</i>	<i>Labidus (n=1)</i>	1.500	0.221	0.388
	<i>Neoponera (n=2)</i>	2.400	0.121	0.379
	<i>Nylanderia (n=1)</i>	1.500	0.221	0.388
	<i>Odontomachus (n=6)</i>	4.000	0.046	0.335
	<i>Paraponera (n=2)</i>	0.000	1.000	1.000
	<i>Pseudomyrmex (n=3)</i>	3.000	0.083	0.335
	<i>Solenopsis (n=2)</i>	2.400	0.121	0.379
<i>Labidus (n=1)</i>	<i>Neoponera (n=2)</i>	0.000	1.000	1.000
	<i>Nylanderia (n=1)</i>	1.000	0.317	0.462
	<i>Odontomachus (n=6)</i>	1.000	0.317	0.462
	<i>Paraponera (n=2)</i>	1.500	0.221	0.388
	<i>Pseudomyrmex (n=3)</i>	1.800	0.180	0.388
	<i>Solenopsis (n=2)</i>	1.500	0.221	0.388
<i>Neoponera (n=2)</i>	<i>Nylanderia (n=1)</i>	0.000	1.000	1.000
	<i>Odontomachus (n=6)</i>	0.000	1.000	1.000
	<i>Paraponera (n=2)</i>	2.400	0.121	0.379
	<i>Pseudomyrmex (n=3)</i>	1.333	0.248	0.408
	<i>Solenopsis (n=2)</i>	0.600	0.439	0.605
<i>Nylanderia (n=1)</i>	<i>Odontomachus (n=6)</i>	0.000	1.000	1.000
	<i>Paraponera (n=2)</i>	1.500	0.221	0.388
	<i>Pseudomyrmex (n=3)</i>	1.800	0.180	0.388
	<i>Solenopsis (n=2)</i>	0.000	1.000	1.000
<i>Odontomachus (n=6)</i>	<i>Paraponera (n=2)</i>	40.000	0.046	0.335
	<i>Pseudomyrmex (n=3)</i>	3.267	0.071	0.335

		<i>Solenopsis</i> (n=2)	0.111	0.739	0.863
	<i>Paraponera</i> (n=2)	<i>Pseudomyrmex</i> (n=3)	3.000	0.083	0.335
		<i>Solenopsis</i> (n=2)	2.400	0.121	0.379
	<i>Pseudomyrmex</i> (n=3)	<i>Solenopsis</i> (n=2)	0.000	1.000	1.000
Diet	Carnivorous (n=13)	Herbivorous (n=9)	4.232	0.040	0.119
		Omnivorous (n=24)	0.138	0.710	0.710
	Herbivorous (n=9)	Omnivorous (n=24)	2.880	0.090	0.134
Nest	Ground (n=21)	Rotten (n=7)	11.711	0.001	0.002
		Tree (n=18)	7.013	0.008	0.012
	Rotten (n=7)	Tree (n=18)	4.487	0.034	0.034
Habitat	Rainforest (n=24)	City (n=22)	0.017	0.895	0.895

Shannon diversity forest/city dataset

Factor	Group 1	Group 2	H	p-value	q-value
Subfamily	<i>Dolichoderinae</i> (n=2)	<i>Ectatomminae</i> (n=2)	2.400	0.121	0.202
		<i>Formicinae</i> (n=4)	0.214	0.643	0.742
		<i>Myrmicinae</i> (n=6)	0.111	0.739	0.792
		<i>Paraponerinae</i> (n=2)	0.600	0.439	0.598
		<i>Ponerinae</i> (n=5)	2.400	0.121	0.202
	<i>Ectatomminae</i> (n=2)	<i>Formicinae</i> (n=4)	3.429	0.064	0.137
		<i>Myrmicinae</i> (n=6)	4.000	0.046	0.132
		<i>Paraponerinae</i> (n=2)	0.000	1.000	1.000
		<i>Ponerinae</i> (n=5)	3.750	0.052	0.132
	<i>Formicinae</i> (n=4)	<i>Myrmicinae</i> (n=6)	0.409	0.522	0.653
		<i>Paraponerinae</i> (n=2)	0.857	0.355	0.532
		<i>Ponerinae</i> (n=5)	3.840	0.050	0.132
	<i>Myrmicinae</i> (n=6)	<i>Paraponerinae</i> (n=2)	4.000	0.046	0.132
		<i>Ponerinae</i> (n=5)	4.800	0.028	0.132
<i>Paraponerinae</i> (n=2)	<i>Ponerinae</i> (n=5)	3.750	0.053	0.132	
Genus	<i>A. cephalotes</i> (n=2)	<i>A. chartifex</i> (n=2)	0.000	1.000	1.000
		<i>C. atriceps</i> (n=2)	2.400	1.000	0.182
		<i>C. atratus</i> (n=2)	0.600	0.121	0.533
		<i>D. armigerum</i> (n=2)	2.400	0.439	0.182
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	2.400	0.121	0.182
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>A. chartifex</i> (n=2)	<i>C. atriceps</i> (n=2)	0.000	1.000	1.000
		<i>C. atratus</i> (n=2)	0.000	1.000	1.000
		<i>D. armigerum</i> (n=2)	0.600	0.439	0.533
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	0.600	0.439	0.533
<i>N. commutata</i> (n=2)		0.600	0.439	0.533	
<i>O. haematodus</i> (n=2)		3.000	0.083	0.182	

		<i>P. clavata</i> (n=2)	0.600	0.439	0.533
	<i>C. atriceps</i> (n=2)	<i>C. atratus</i> (n=2)	2.400	0.121	0.182
		<i>D. armigerum</i> (n=2)	0.600	0.439	0.533
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.000	1.000	1.000
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>C. atratus</i> (n=2)	<i>D. armigerum</i> (n=2)	2.400	0.121	0.182
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	2.400	0.121	0.182
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>D. armigerum</i> (n=2)	<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.000	1.000	1.000
		<i>O. haematodus</i> (n=2)	1.333	0.248	0.360
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>E. tuberculatum</i> (n=2)	<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	2.400	0.121	0.182
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	0.000	1.000	1.000
	<i>G. destructor</i> (n=2)	<i>N. commutata</i> (n=2)	2.400	0.121	0.182
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	0.000	1.000	1.000
	<i>N. commutata</i> (n=2)	<i>O. haematodus</i> (n=2)	0.333	0.564	0.668
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>O. haematodus</i> (n=2)	<i>P. clavata</i> (n=2)	3.000	0.083	0.182
Diet	Carnivorous (n=7)	Herbivorous (n=4)	2.381	0.123	0.184
		Omnivorous (n=10)	3.704	0.054	0.163
	Herbivorous (n=4)	Omnivorous (n=10)	0.462	0.497	0.497
Nest	Ground (n=11)	Rotten (n=4)	4.926	0.026	0.080
		Tree (n=6)	1.455	0.228	0.228
	Rotten (n=4)	Tree (n=6)	2.227	0.136	0.203
Habitat	Rainforest (n=10)	City (n=11)	0.045	0.833	0.833

Pielou's evenness whole dataset

Factor	Group 1	Group 2	H	p-value	q-value
Subfamily	Dolichoderinae (n=5)	Dorylinae (n=2)	0.000	1.000	1.000
		Ectatomminae (n=3)	1.800	0.180	0.296
		Formicinae (n=8)	0.086	0.770	0.897
		Myrmicinae (n=14)	0.000	1.000	1.000
		Paraponerinae (n=2)	2.400	0.121	0.226
		Ponerinae (n=9)	4.271	0.039	0.153

		Pseudomyrmicinae (n=3)	5.000	0.025	0.153
	Dorylinae (n=2)	Ectatomminae (n=3)	1.333	0.248	0.347
		Formicinae (n=8)	0.000	1.000	1.000
		Myrmicinae (n=14)	0.000	1.000	1.000
		Paraponerinae (n=2)	0.600	0.439	0.585
		Ponerinae (n=9)	2.722	0.099	0.205
		Pseudomyrmicinae (n=3)	3.000	0.083	0.205
	Ectatomminae (n=3)	Formicinae (n=8)	2.667	0.102	0.205
		Myrmicinae (n=14)	2.683	0.101	0.205
		Paraponerinae (n=2)	0.333	0.564	0.717
		Ponerinae (n=9)	5.342	0.021	0.153
		Pseudomyrmicinae (n=3)	3.857	0.049	0.154
	Formicinae (n=8)	Myrmicinae (n=14)	0.117	0.733	0.892
		Paraponerinae (n=2)	1.705	0.192	0.298
		Ponerinae (n=9)	4.481	0.034	0.153
		Pseudomyrmicinae (n=3)	5.042	0.025	0.153
	Myrmicinae (n=14)	Paraponerinae (n=2)	1.613	0.204	0.300
		Ponerinae (n=9)	4.063	0.044	0.153
		Pseudomyrmicinae (n=3)	4.063	0.044	0.153
	Paraponerinae (n=2)	Ponerinae (n=9)	4.500	0.034	0.153
		Pseudomyrmicinae (n=3)	3.000	0.083	0.205
	Ponerinae (n=9)	Pseudomyrmicinae (n=3)	1.923	0.166	0.290
Genus	Anochetus (n=1)	<i>Atta</i> (n=3)	1.800	0.180	0.402
		<i>Azteca</i> (n=3)	1.800	0.180	0.402
		<i>Camponotus</i> (n=6)	1.000	0.317	0.450
		<i>Cephalotes</i> (n=4)	2.000	0.157	0.402
		<i>Crematogaster</i> (n=2)	1.500	0.221	0.402
		<i>Daceton</i> (n=2)	0.000	1.000	1.000
		<i>Dolichoderus</i> (n=2)	1.500	0.221	0.402
		<i>Eciton</i> (n=1)	1.000	0.317	0.450
		<i>Ectatomma</i> (n=3)	1.800	0.180	0.402
		<i>Gigantiops</i> (n=2)	1.500	0.221	0.402
		<i>Labidus</i> (n=1)	1.000	0.317	0.450
		<i>Neoponera</i> (n=2)	0.000	1.000	1.000
		<i>Nylanderia</i> (n=1)	1.000	0.317	0.450
		<i>Odontomachus</i> (n=6)	0.000	1.000	1.000
		<i>Paraponera</i> (n=2)	1.500	0.221	0.402
		<i>Pseudomyrmex</i> (n=3)	1.800	0.180	0.402
		<i>Solenopsis</i> (n=2)	0.000	1.000	1.000
		<i>Atta</i> (n=3)	<i>Azteca</i> (n=3)	0.429	0.513
	<i>Camponotus</i> (n=6)		3.267	0.071	0.398
	<i>Cephalotes</i> (n=4)		2.000	0.157	0.402
	<i>Crematogaster</i> (n=2)		1.333	0.248	0.404
	<i>Daceton</i> (n=2)		3.000	0.083	0.398
	<i>Dolichoderus</i> (n=2)		1.333	0.248	0.404
	<i>Eciton</i> (n=1)		0.200	0.655	0.742
	<i>Ectatomma</i> (n=3)		0.429	0.513	0.633

		<i>Gigantiops</i> (n=2)	0.333	0.564	0.663
		<i>Labidus</i> (n=1)	1.800	0.180	0.402
		<i>Neoponera</i> (n=2)	3.000	0.083	0.398
		<i>Nylanderia</i> (n=1)	1.800	0.180	0.402
		<i>Odontomachus</i> (n=6)	4.267	0.039	0.398
		<i>Paraponera</i> (n=2)	0.333	0.564	0.663
		<i>Pseudomyrmex</i> (n=3)	3.857	0.049	0.368
		<i>Solenopsis</i> (n=2)	3.000	0.083	0.398
	Azteca (n=3)	<i>Camponotus</i> (n=6)	1.667	0.197	0.402
		<i>Cephalotes</i> (n=4)	3.125	0.077	0.398
		<i>Crematogaster</i> (n=2)	0.333	0.564	0.663
		<i>Daceton</i> (n=2)	1.333	0.248	0.404
		<i>Dolichoderus</i> (n=2)	0.333	0.564	0.663
		<i>Eciton</i> (n=1)	0.200	0.655	0.742
		<i>Ectatomma</i> (n=3)	1.190	0.275	0.443
		<i>Gigantiops</i> (n=2)	1.333	0.248	0.404
		<i>Labidus</i> (n=1)	1.800	0.180	0.402
		<i>Neoponera</i> (n=2)	1.333	0.248	0.404
		<i>Nylanderia</i> (n=1)	1.800	0.180	0.402
		<i>Odontomachus</i> (n=6)	3.267	0.071	0.398
		<i>Paraponera</i> (n=2)	1.333	0.248	0.404
		<i>Pseudomyrmex</i> (n=3)	3.857	0.049	0.398
		<i>Solenopsis</i> (n=2)	3.000	0.083	0.398
		Camponotus (n=6)	<i>Cephalotes</i> (n=4)	6.545	0.011
	<i>Crematogaster</i> (n=2)		0.111	0.739	0.819
	<i>Daceton</i> (n=2)		0.444	0.505	0.633
	<i>Dolichoderus</i> (n=2)		0.111	0.739	0.819
	<i>Eciton</i> (n=1)		1.000	0.317	0.450
	<i>Ectatomma</i> (n=3)		3.267	0.071	0.398
	<i>Gigantiops</i> (n=2)		4.000	0.046	0.398
	<i>Labidus</i> (n=1)		0.000	1.000	1.000
	<i>Neoponera</i> (n=2)		0.111	0.739	0.819
	<i>Nylanderia</i> (n=1)		1.000	0.317	0.450
	<i>Odontomachus</i> (n=6)		2.564	0.109	0.402
	<i>Paraponera</i> (n=2)		2.778	0.096	0.402
	<i>Pseudomyrmex</i> (n=3)		4.267	0.039	0.398
	<i>Solenopsis</i> (n=2)		1.778	0.182	0.402
	Cephalotes (n=4)	<i>Crematogaster</i> (n=2)	3.429	0.064	0.398
		<i>Daceton</i> (n=2)	3.429	0.064	0.398
		<i>Dolichoderus</i> (n=2)	3.429	0.064	0.398
		<i>Eciton</i> (n=1)	2.000	0.157	0.402
		<i>Ectatomma</i> (n=3)	0.500	0.480	0.622
		<i>Gigantiops</i> (n=2)	0.214	0.643	0.742
		<i>Labidus</i> (n=1)	2.000	0.157	0.402
		<i>Neoponera</i> (n=2)	3.429	0.064	0.398
		<i>Nylanderia</i> (n=1)	2.000	0.157	0.402
		<i>Odontomachus</i> (n=6)	6.545	0.011	0.398

	<i>Paraponera</i> (n=2)	0.000	1.000	1.000	
	<i>Pseudomyrmex</i> (n=3)	4.500	0.034	0.398	
	<i>Solenopsis</i> (n=2)	3.429	0.064	0.398	
<i>Crematogaster</i> (n=2)	<i>Daceton</i> (n=2)	0.600	0.439	0.574	
	<i>Dolichoderus</i> (n=2)	0.600	0.439	0.574	
	<i>Eciton</i> (n=1)	1.500	0.221	0.402	
	<i>Ectatomma</i> (n=3)	1.333	0.248	0.404	
	<i>Gigantiops</i> (n=2)	2.400	0.121	0.402	
	<i>Labidus</i> (n=1)	0.000	1.000	1.000	
	<i>Neoponera</i> (n=2)	0.600	0.439	0.573	
	<i>Nylanderia</i> (n=1)	1.500	0.221	0.402	
	<i>Odontomachus</i> (n=6)	1.000	0.317	0.450	
	<i>Paraponera</i> (n=2)	2.400	0.121	0.402	
	<i>Pseudomyrmex</i> (n=3)	3.000	0.083	0.398	
	<i>Solenopsis</i> (n=2)	0.600	0.439	0.574	
	<i>Daceton</i> (n=2)	<i>Dolichoderus</i> (n=2)	0.600	0.439	0.574
		<i>Eciton</i> (n=1)	1.500	0.221	0.402
		<i>Ectatomma</i> (n=3)	3.000	0.083	0.398
<i>Gigantiops</i> (n=2)		2.400	0.121	0.402	
<i>Labidus</i> (n=1)		0.000	1.000	1.000	
<i>Neoponera</i> (n=2)		0.600	0.439	0.574	
<i>Nylanderia</i> (n=1)		1.500	0.221	0.402	
<i>Odontomachus</i> (n=6)		0.444	0.505	0.633	
<i>Paraponera</i> (n=2)		2.400	0.121	0.402	
<i>Pseudomyrmex</i> (n=3)		1.333	0.248	0.404	
<i>Solenopsis</i> (n=2)		0.600	0.439	0.574	
<i>Dolichoderus</i> (n=2)	<i>Eciton</i> (n=1)	1.500	0.221	0.402	
	<i>Ectatomma</i> (n=3)	1.333	0.248	0.404	
	<i>Gigantiops</i> (n=2)	2.400	0.121	0.402	
	<i>Labidus</i> (n=1)	0.000	1.000	1.000	
	<i>Neoponera</i> (n=2)	0.600	0.439	0.574	
	<i>Nylanderia</i> (n=1)	1.500	0.221	0.402	
	<i>Odontomachus</i> (n=6)	1.000	0.317	0.450	
	<i>Paraponera</i> (n=2)	2.400	0.121	0.402	
	<i>Pseudomyrmex</i> (n=3)	3.000	0.083	0.398	
	<i>Solenopsis</i> (n=2)	2.400	0.121	0.402	
<i>Eciton</i> (n=1)	<i>Ectatomma</i> (n=3)	0.200	0.654	0.742	
	<i>Gigantiops</i> (n=2)	1.500	0.221	0.402	
	<i>Labidus</i> (n=1)	1.000	0.317	0.450	
	<i>Neoponera</i> (n=2)	1.500	0.221	0.402	
	<i>Nylanderia</i> (n=1)	1.000	0.137	0.450	
	<i>Odontomachus</i> (n=6)	2.250	0.133	0.402	
	<i>Paraponera</i> (n=2)	0.000	1.000	1.000	
	<i>Pseudomyrmex</i> (n=3)	1.800	0.180	0.402	
<i>Ectatomma</i> (n=3)	<i>Solenopsis</i> (n=2)	1.500	0.221	0.402	
	<i>Gigantiops</i> (n=2)	0.333	0.564	0.663	
	<i>Labidus</i> (n=1)	1.800	0.180	0.402	

		<i>Neoponera</i> (n=2)	3.000	0.083	0.398
		<i>Nylanderia</i> (n=1)	1.800	0.180	0.402
		<i>Odontomachus</i> (n=6)	4.267	0.039	0.398
		<i>Paraponera</i> (n=2)	0.333	0.564	0.663
		<i>Pseudomyrmex</i> (n=3)	3.857	0.049	0.398
		<i>Solenopsis</i> (n=2)	3.000	0.083	0.398
	<i>Gigantiops</i> (n=2)	<i>Labidus</i> (n=1)	1.500	0.221	0.402
		<i>Neoponera</i> (n=2)	2.400	0.121	0.402
		<i>Nylanderia</i> (n=1)	1.500	0.221	0.402
		<i>Odontomachus</i> (n=6)	4.000	0.046	0.398
		<i>Paraponera</i> (n=2)	0.000	1.000	1.000
		<i>Pseudomyrmex</i> (n=3)	3.000	0.083	0.398
		<i>Solenopsis</i> (n=2)	2.400	0.121	0.402
	<i>Labidus</i> (n=1)	<i>Neoponera</i> (n=2)	0.000	1.000	1.000
		<i>Nylanderia</i> (n=1)	1.000	0.317	0.450
		<i>Odontomachus</i> (n=6)	1.000	0.317	0.450
		<i>Paraponera</i> (n=2)	1.500	0.221	0.402
		<i>Pseudomyrmex</i> (n=3)	1.800	0.180	0.402
		<i>Solenopsis</i> (n=2)	1.500	0.221	0.402
	<i>Neoponera</i> (n=2)	<i>Nylanderia</i> (n=1)	1.500	0.221	0.402
		<i>Odontomachus</i> (n=6)	0.444	0.505	0.633
		<i>Paraponera</i> (n=2)	2.400	0.121	0.402
		<i>Pseudomyrmex</i> (n=3)	1.333	0.248	0.404
		<i>Solenopsis</i> (n=2)	0.600	0.439	0.574
	<i>Nylanderia</i> (n=1)	<i>Odontomachus</i> (n=6)	0.000	1.000	1.000
		<i>Paraponera</i> (n=2)	1.500	0.221	0.402
		<i>Pseudomyrmex</i> (n=3)	0.200	0.655	0.742
		<i>Solenopsis</i> (n=2)	0.000	1.000	1.000
	<i>Odontomachus</i> (n=6)	<i>Paraponera</i> (n=2)	4.000	0.046	0.398
		<i>Pseudomyrmex</i> (n=3)	1.067	0.302	0.450
		<i>Solenopsis</i> (n=2)	0.444	0.405	0.633
	<i>Paraponera</i> (n=2)	<i>Pseudomyrmex</i> (n=3)	3.000	0.083	0.398
		<i>Solenopsis</i> (n=2)	2.400	0.121	0.402
	<i>Pseudomyrmex</i> (n=3)	<i>Solenopsis</i> (n=2)	0.000	1.000	1.000
Diet	Carnivorous (n=13)	Herbivorous (n=9)	6.124	0.013	0.040
		Omnivorous (n=24)	0.551	0.458	0.458
	Herbivorous (n=9)	Omnivorous (n=24)	4.008	0.045	0.068
Nest	Ground (n=21)	Rotten (n=7)	12.077	0.001	0.002
		Tree (n=18)	3.051	0.081	0.081
	Rotten (n=7)	Tree (n=18)	6.773	0.009	0.014
Habitat	Rainforest (n=24)	City (n=22)	0.002	0.965	0.965

Pielou's evenness forest/city dataset

Factor	Group 1	Group 2	H	p-value	q-value
Subfamily	Dolichoderinae (n=2)	Ectatomminae (n=2)	2.400	0.121	0.228
		Formicinae (n=4)	0.214	0.643	0.804
		Myrmicinae (n=6)	0.000	1.000	1.000
		Paraponerinae (n=2)	0.600	0.439	0.598
		Ponerinae (n=5)	2.400	0.121	0.228
	Ectatomminae (n=2)	Formicinae (n=4)	3.429	0.064	0.192
		Myrmicinae (n=6)	4.000	0.046	0.192
		Paraponerinae (n=2)	0.000	1.000	1.000
		Ponerinae (n=5)	3.750	0.053	0.192
	Formicinae (n=4)	Myrmicinae (n=6)	0.045	0.831	0.959
		Paraponerinae (n=2)	0.857	0.355	0.532
		Ponerinae (n=5)	2.940	0.086	0.216
	Myrmicinae (n=6)	Paraponerinae (n=2)	1.778	0.182	0.304
		Ponerinae (n=5)	4.800	0.028	0.192
Paraponerinae (n=2)	Ponerinae (n=5)	3.750	0.053	0.192	
Genus	<i>A. cephalotes</i> (n=2)	<i>A. chartifex</i> (n=2)	0.000	1.000	1.000
		<i>C. atriceps</i> (n=2)	2.400	0.121	0.182
		<i>C. atratus</i> (n=2)	2.400	0.121	0.182
		<i>D. armigerum</i> (n=2)	0.600	0.439	0.506
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.600	0.439	0.506
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>A. chartifex</i> (n=2)	<i>C. atriceps</i> (n=2)	2.400	0.121	0.182
		<i>C. atratus</i> (n=2)	0.600	0.439	0.506
		<i>D. armigerum</i> (n=2)	0.600	0.439	0.506
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	0.600	0.439	0.506
		<i>N. commutata</i> (n=2)	0.600	0.439	0.506
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	0.600	0.439	0.506
	<i>C. atriceps</i> (n=2)	<i>C. atratus</i> (n=2)	2.400	0.121	0.182
		<i>D. armigerum</i> (n=2)	0.000	1.000	1.000
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.600	0.439	0.506
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>C. atratus</i> (n=2)	<i>D. armigerum</i> (n=2)	2.400	0.121	0.182
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	0.600	0.439	0.506
<i>N. commutata</i> (n=2)		2.400	0.121	0.182	
<i>O. haematodus</i> (n=2)		3.000	0.083	0.182	
<i>P. clavata</i> (n=2)		0.000	1.000	1.000	

	<i>D. armigerum</i> (n=2)	<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.000	1.000	1.000
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>E. tuberculatum</i> (n=2)	<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	2.400	0.121	0.182
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	0.000	1.000	1.000
	<i>G. destructor</i> (n=2)	<i>N. commutata</i> (n=2)	2.400	0.121	0.182
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	0.000	1.000	1.000
	<i>N. commutata</i> (n=2)	<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
<i>O. haematodus</i> (n=2)	<i>P. clavata</i> (n=2)	3.000	0.083	0.182	
Diet	Carnivorous (n=7)	Herbivorous (n=4)	2.381	0.123	0.184
		Omnivorous (n=10)	3.000	0.083	0.184
	Herbivorous (n=4)	Omnivorous (n=10)	0.000	1.000	1.000
Nest	Ground (n=11)	Rotten (n=4)	5.523	0.019	0.049
		Tree (n=6)	0.162	0.688	0.688
	Rotten (n=4)	Tree (n=6)	4.545	0.033	0.049
Habitat	Rainforest (n=10)	City (n=11)	0.005	0.944	0.944

Faith's PD whole dataset

Factor	Group 1	Group 2	H	p-value	q-value
Subfamily	Dolichoderinae (n=5)	Dorylinae (n=2)	1.350	0.245	0.386
		Ectatomminae (n=3)	3.756	0.053	0.237
		Formicinae (n=8)	0.000	1.000	1.000
		Myrmicinae (n=14)	1.137	0.711	0.948
		Paraponerinae (n=2)	3.750	0.053	0.237
		Ponerinae (n=9)	0.040	0.841	0.961
		Pseudomyrmicinae (n=3)	0.022	0.881	0.961
	Dorylinae (n=2)	Ectatomminae (n=3)	1.333	0.248	0.386
		Formicinae (n=8)	2.455	0.117	0.273
		Myrmicinae (n=14)	2.521	0.112	0.273
		Paraponerinae (n=2)	2.400	0.121	0.273
		Ponerinae (n=9)	1.389	0.239	0.386
		Pseudomyrmicinae (n=3)	1.333	0.248	0.386
	Ectatomminae (n=3)	Formicinae (n=8)	4.167	0.041	0.237
		Myrmicinae (n=14)	5.143	0.023	0.237
		Paraponerinae (n=2)	3.000	0.083	0.259
		Ponerinae (n=9)	1.923	0.166	0.331
		Pseudomyrmicinae (n=3)	2.333	0.127	0.273
	Formicinae (n=8)	Myrmicinae (n=14)	0.377	0.539	0.789
		Paraponerinae (n=2)	4.364	0.037	0.237

		Ponerinae (n=9)	0.333	0.564	0.789	
		Pseudomyrmicinae (n=3)	0.042	0.838	0.961	
	Myrmicinae (n=14)	Paraponerinae (n=2)	4.941	0.026	0.237	
		Ponerinae (n=9)	0.016	0.900	0.961	
		Pseudomyrmicinae (n=3)	0.016	0.900	0.961	
	Paraponerinae (n=2)	Ponerinae (n=9)	3.556	0.059	0.237	
		Pseudomyrmicinae (n=3)	3.000	0.083	0.259	
	Ponerinae (n=9)	Pseudomyrmicinae (n=3)	0.009	0.926	0.961	
Genus	Anochetus (n=1)	<i>Atta</i> (n=3)	1.800	0.180	0.393	
		<i>Azteca</i> (n=3)	1.800	0.180	0.393	
		<i>Camponotus</i> (n=6)	2.250	0.134	0.393	
		<i>Cephalotes</i> (n=4)	2.000	0.157	0.393	
		<i>Crematogaster</i> (n=2)	1.500	0.221	0.393	
		<i>Daceton</i> (n=2)	1.500	0.221	0.393	
		<i>Dolichoderus</i> (n=2)	1.500	0.221	0.393	
		<i>Eciton</i> (n=1)	1.000	0.317	0.501	
		<i>Ectatomma</i> (n=3)	1.800	0.180	0.393	
		<i>Gigantiops</i> (n=2)	1.500	0.221	0.393	
		<i>Labidus</i> (n=1)	1.000	0.317	0.501	
		<i>Neoponera</i> (n=2)	1.500	0.221	0.393	
		<i>Nylanderia</i> (n=1)	1.000	0.317	0.501	
		<i>Odontomachus</i> (n=6)	1.000	0.317	0.501	
		<i>Paraponera</i> (n=2)	1.500	0.221	0.393	
		<i>Pseudomyrmex</i> (n=3)	1.800	0.180	0.180	
		<i>Solenopsis</i> (n=2)	1.500	0.221	0.221	
		<i>Atta</i> (n=3)	<i>Azteca</i> (n=3)	3.857	0.049	0.049
			<i>Camponotus</i> (n=6)	5.400	0.020	0.020
			<i>Cephalotes</i> (n=4)	4.500	0.034	0.034
			<i>Crematogaster</i> (n=2)	3.000	0.083	0.083
			<i>Daceton</i> (n=2)	3.000	0.083	0.083
			<i>Dolichoderus</i> (n=2)	0.000	1.000	1.000
			<i>Eciton</i> (n=1)	1.800	0.180	0.393
			<i>Ectatomma</i> (n=3)	0.429	0.513	0.740
			<i>Gigantiops</i> (n=2)	0.000	1.000	1.000
			<i>Labidus</i> (n=1)	0.200	0.655	0.801
			<i>Neoponera</i> (n=2)	3.000	0.083	0.393
			<i>Nylanderia</i> (n=1)	1.800	0.180	0.393
			<i>Odontomachus</i> (n=6)	2.400	0.121	0.393
			<i>Paraponera</i> (n=2)	3.000	0.083	0.393
			<i>Pseudomyrmex</i> (n=3)	3.857	0.049	0.393
		<i>Solenopsis</i> (n=2)	3.000	0.083	0.393	
		<i>Azteca</i> (n=3)	<i>Camponotus</i> (n=6)	0.267	0.606	0.785
			<i>Cephalotes</i> (n=4)	2.000	0.157	0.393
			<i>Crematogaster</i> (n=2)	0.333	0.564	0.750
			<i>Daceton</i> (n=2)	0.333	0.564	0.750
			<i>Dolichoderus</i> (n=2)	0.333	0.564	0.750
			<i>Eciton</i> (n=1)	1.800	0.180	0.393

	<i>Ectatomma (n=3)</i>	3.857	0.049	0.393
	<i>Gigantiops (n=2)</i>	3.000	0.083	0.393
	<i>Labidus (n=1)</i>	1.800	0.180	0.393
	<i>Neoponera (n=2)</i>	0.333	0.564	0.750
	<i>Nylanderia (n=1)</i>	0.200	0.655	0.801
	<i>Odontomachus (n=6)</i>	0.267	0.606	0.785
	<i>Paraponera (n=2)</i>	3.000	0.083	0.392
	<i>Pseudomyrmex (n=3)</i>	0.048	0.827	0.967
	<i>Solenopsis (n=2)</i>	0.000	1.000	1.000
<i>Camponotus (n=6)</i>	<i>Cephalotes (n=4)</i>	3.682	0.056	0.393
	<i>Crematogaster (n=2)</i>	0.444	0.505	0.736
	<i>Daceton (n=2)</i>	0.111	0.739	0.876
	<i>Dolichoderus (n=2)</i>	0.111	0.739	0.876
	<i>Eciton (n=1)</i>	2.250	0.134	0.393
	<i>Ectatomma (n=3)</i>	5.400	0.020	0.393
	<i>Gigantiops (n=2)</i>	4.000	0.046	0.393
	<i>Labidus (n=1)</i>	2.250	0.133	0.393
	<i>Neoponera (n=2)</i>	0.000	1.000	1.000
	<i>Nylanderia (n=1)</i>	0.250	0.617	0.787
	<i>Odontomachus (n=6)</i>	0.410	0.522	0.746
	<i>Paraponera (n=2)</i>	4.000	0.046	0.393
	<i>Pseudomyrmex (n=3)</i>	0.267	0.606	0.785
	<i>Solenopsis (n=2)</i>	0.000	1.000	1.000
<i>Cephalotes (n=4)</i>	<i>Crematogaster (n=2)</i>	3.429	0.064	0.393
	<i>Daceton (n=2)</i>	3.439	0.064	0.393
	<i>Dolichoderus (n=2)</i>	1.929	0.165	0.393
	<i>Eciton (n=1)</i>	2.000	0.157	0.393
	<i>Ectatomma (n=3)</i>	4.500	0.034	0.393
	<i>Gigantiops (n=2)</i>	3.429	0.064	0.393
	<i>Labidus (n=1)</i>	2.000	0.157	0.393
	<i>Neoponera (n=2)</i>	0.000	1.000	1.000
	<i>Nylanderia (n=1)</i>	2.000	0.157	0.393
	<i>Odontomachus (n=6)</i>	2.227	0.136	0.393
	<i>Paraponera (n=2)</i>	3.429	0.064	0.393
	<i>Pseudomyrmex (n=3)</i>	0.500	0.480	0.712
<i>Crematogaster (n=2)</i>	<i>Solenopsis (n=2)</i>	0.000	1.000	1.000
	<i>Daceton (n=2)</i>	0.000	1.000	1.000
	<i>Dolichoderus (n=2)</i>	0.000	1.000	1.000
	<i>Eciton (n=1)</i>	1.500	0.221	0.393
	<i>Ectatomma (n=3)</i>	3.000	0.083	0.393
	<i>Gigantiops (n=2)</i>	2.400	0.121	0.393
	<i>Labidus (n=1)</i>	1.500	0.221	0.393
	<i>Neoponera (n=2)</i>	0.600	0.439	0.659
	<i>Nylanderia (n=1)</i>	0.000	1.000	1.000
	<i>Odontomachus (n=6)</i>	1.000	0.317	0.501
	<i>Paraponera (n=2)</i>	2.400	0.121	0.393
<i>Pseudomyrmex (n=3)</i>	0.000	1.000	1.000	

	<i>Solenopsis (n=2)</i>	0.600	0.439	0.658
<i>Daceton (n=2)</i>	<i>Dolichoderus (n=2)</i>	0.000	1.000	1.000
	<i>Eciton (n=1)</i>	1.500	0.221	0.393
	<i>Ectatomma (n=3)</i>	3.000	0.083	0.393
	<i>Gigantiops (n=2)</i>	2.400	0.121	0.393
	<i>Labidus (n=1)</i>	1.500	0.221	0.393
	<i>Neoponera (n=2)</i>	0.000	1.000	1.000
	<i>Nylanderia (n=1)</i>	1.500	0.221	0.393
	<i>Odontomachus (n=6)</i>	0.444	0.505	0.736
	<i>Paraponera (n=2)</i>	2.400	0.121	0.393
	<i>Pseudomyrmex (n=3)</i>	0.333	0.564	0.750
		<i>Solenopsis (n=2)</i>	0.000	1.000
<i>Dolichoderus (n=2)</i>	<i>Eciton (n=1)</i>	0.000	1.000	1.000
	<i>Ectatomma (n=3)</i>	1.333	0.248	0.437
	<i>Gigantiops (n=2)</i>	0.000	1.000	1.000
	<i>Labidus (n=1)</i>	0.000	1.000	1.000
	<i>Neoponera (n=2)</i>	0.600	0.439	0.658
	<i>Nylanderia (n=1)</i>	0.000	1.000	1.000
	<i>Odontomachus (n=6)</i>	0.000	1.000	1.000
	<i>Paraponera (n=2)</i>	2.400	0.121	0.393
	<i>Pseudomyrmex (n=3)</i>	0.333	0.564	0.750
		<i>Solenopsis (n=2)</i>	0.600	0.439
<i>Eciton (n=1)</i>	<i>Ectatomma (n=3)</i>	1.800	0.180	0.393
	<i>Gigantiops (n=2)</i>	1.500	0.221	0.393
	<i>Labidus (n=1)</i>	1.000	0.317	0.501
	<i>Neoponera (n=2)</i>	1.500	0.221	0.393
	<i>Nylanderia (n=1)</i>	1.000	0.317	0.501
	<i>Odontomachus (n=6)</i>	1.000	0.317	0.501
	<i>Paraponera (n=2)</i>	1.500	0.221	0.393
	<i>Pseudomyrmex (n=3)</i>	0.200	0.655	0.801
		<i>Solenopsis (n=2)</i>	1.500	0.221
<i>Ectatomma (n=3)</i>	<i>Gigantiops (n=2)</i>	0.333	0.564	0.750
	<i>Labidus (n=1)</i>	0.200	0.655	0.801
	<i>Neoponera (n=2)</i>	3.000	0.083	0.393
	<i>Nylanderia (n=1)</i>	1.800	0.180	0.393
	<i>Odontomachus (n=6)</i>	2.400	0.121	0.393
	<i>Paraponera (n=2)</i>	3.000	0.083	0.393
	<i>Pseudomyrmex (n=3)</i>	2.333	0.127	0.393
		<i>Solenopsis (n=2)</i>	3.000	0.083
<i>Gigantiops (n=2)</i>	<i>Labidus (n=1)</i>	1.500	0.221	0.393
	<i>Neoponera (n=2)</i>	2.400	0.121	0.393
	<i>Nylanderia (n=1)</i>	1.500	0.221	0.393
	<i>Odontomachus (n=6)</i>	1.778	0.182	0.393
	<i>Paraponera (n=2)</i>	2.400	0.121	0.393
	<i>Pseudomyrmex (n=3)</i>	3.000	0.083	0.393
		<i>Solenopsis (n=2)</i>	2.400	0.221
<i>Labidus (n=1)</i>	<i>Neoponera (n=2)</i>	1.500	0.221	0.393

		<i>Nylanderia</i> (n=1)	1.000	0.317	0.501
		<i>Odontomachus</i> (n=6)	1.000	0.317	0.501
		<i>Paraponera</i> (n=2)	1.500	0.221	0.393
		<i>Pseudomyrmex</i> (n=3)	1.800	0.180	0.393
		<i>Solenopsis</i> (n=2)	1.500	0.221	0.393
	<i>Neoponera</i> (n=2)	<i>Nylanderia</i> (n=1)	0.000	1.000	1.000
		<i>Odontomachus</i> (n=6)	0.111	0.739	0.876
		<i>Paraponera</i> (n=2)	2.400	0.121	0.393
		<i>Pseudomyrmex</i> (n=3)	0.330	0.564	0.750
		<i>Solenopsis</i> (n=2)	0.600	0.439	0.658
	<i>Nylanderia</i> (n=1)	<i>Odontomachus</i> (n=6)	0.250	0.617	0.787
		<i>Paraponera</i> (n=2)	1.500	0.221	0.393
		<i>Pseudomyrmex</i> (n=3)	0.200	0.655	0.801
		<i>Solenopsis</i> (n=2)	0.000	1.000	1.000
	<i>Odontomachus</i> (n=6)	<i>Paraponera</i> (n=2)	2.778	0.096	0.393
		<i>Pseudomyrmex</i> (n=3)	0.067	0.796	0.937
		<i>Solenopsis</i> (n=2)	0.111	0.739	0.876
	<i>Paraponera</i> (n=2)	<i>Pseudomyrmex</i> (n=3)	3.000	0.083	0.393
		<i>Solenopsis</i> (n=2)	2.400	0.121	0.393
	<i>Pseudomyrmex</i> (n=3)	<i>Solenopsis</i> (n=2)	0.000	1.000	1.000
Diet	Carnivorous (n=13)	Herbivorous (n=9)	0.320	0.571	0.857
		Omnivorous (n=24)	0.001	0.975	0.975
	Herbivorous (n=9)	Omnivorous (n=24)	0.371	0.542	0.857
Nest	Ground (n=21)	Rotten (n=7)	9.307	0.002	0.006
		Tree (n=18)	1.013	0.008	0.012
	Rotten (n=7)	Tree (n=18)	0.824	0.364	0.386
Habitat	Rainforest (n=24)	City (n=22)	0.213	0.644	0.644

Faith's PD forest/city dataset

Factor	Group 1	Group 2	H	p-value	q-value
Subfamily	Dolichoderinae (n=2)	Ectatomminae (n=2)	2.400	0.121	0.228
		Formicinae (n=4)	0.929	0.165	0.275
		Myrmicinae (n=6)	0.111	0.739	0.739
		Paraponerinae (n=2)	2.400	0.121	0.228
		Ponerinae (n=5)	0.150	0.699	0.739
	Ectatomminae (n=2)	Formicinae (n=4)	0.857	0.355	0.443
		Myrmicinae (n=6)	4.000	0.046	0.192
		Paraponerinae (n=2)	2.400	0.121	0.228
		Ponerinae (n=5)	3.750	0.053	0.192
	Formicinae (n=4)	Myrmicinae (n=6)	1.636	0.201	0.301
		Paraponerinae (n=2)	3.429	0.064	0.192
		Ponerinae (n=5)	1.500	0.221	0.301
	Myrmicinae (n=6)	Paraponerinae (n=2)	4.000	0.046	0.192
		Ponerinae (n=5)	0.133	0.715	0.739
	Paraponerinae (n=2)	Ponerinae (n=5)	3.750	0.053	0.192

Genus	<i>A. cephalotes</i> (n=2)	<i>A. chartifex</i> (n=2)	2.400	0.121	0.182
		<i>C. atriceps</i> (n=2)	2.400	0.121	0.182
		<i>C. atratus</i> (n=2)	2.400	0.121	0.182
		<i>D. armigerum</i> (n=2)	2.400	0.121	0.182
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	2.400	0.121	0.182
		<i>O. haematodus</i> (n=2)	1.330	0.248	0.349
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>A. chartifex</i> (n=2)	<i>C. atriceps</i> (n=2)	0.600	0.439	0.580
		<i>C. atratus</i> (n=2)	0.600	0.439	0.580
		<i>D. armigerum</i> (n=2)	0.000	1.000	1.000
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.000	1.000	1.000
		<i>O. haematodus</i> (n=2)	0.333	0.564	0.705
	<i>C. atriceps</i> (n=2)	<i>C. atratus</i> (n=2)	2.400	0.121	0.182
		<i>D. armigerum</i> (n=2)	0.000	1.000	1.000
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.000	1.000	1.000
		<i>O. haematodus</i> (n=2)	0.000	1.000	1.000
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>C. atratus</i> (n=2)	<i>D. armigerum</i> (n=2)	2.400	0.121	0.182
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.000	1.000	1.000
		<i>O. haematodus</i> (n=2)	1.333	0.248	0.349
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>D. armigerum</i> (n=2)	<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.000	1.000	1.000
		<i>O. haematodus</i> (n=2)	0.000	1.000	1.000
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>E. tuberculatum</i> (n=2)	<i>G. destructor</i> (n=2)	0.000	1.000	1.000
		<i>N. commutata</i> (n=2)	2.400	0.121	0.182
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>G. destructor</i> (n=2)	<i>N. commutata</i> (n=2)	2.400	0.121	0.182
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>N. commutata</i> (n=2)	<i>O. haematodus</i> (n=2)	0.333	0.564	0.705
<i>P. clavata</i> (n=2)		2.400	0.121	0.182	
<i>O. haematodus</i> (n=2)	<i>P. clavata</i> (n=2)	3.000	0.083	0.182	
	Carnivorous (n=7)	Herbivorous (n=4)	0.095	0.758	0.758
Omnivorous (n=10)		0.231	0.630	0.758	
Diet					

	Herbivorous (n=4)	Omnivorous (n=10)	1.413	0.234	0.703
Nest	Ground (n=11)	Rotten (n=4)	5.523	0.019	0.028
		Tree (n=6)	7.364	0.007	0.020
	Rotten (n=4)	Tree (n=6)	0.409	0.522	0.522
Habitat	Rainforest (n=10)	City (n=11)	0.079	0.778	0.778

ASV richness whole dataset

Factor	Group 1	Group 2	H	p-value	q-value	
Subfamily	Dolichoderinae (n=5)	Dorylinae (n=2)	0.600	0.439	0.572	
		Ectatomminae (n=3)	3.756	0.053	0.148	
		Formicinae (n=8)	0.537	0.464	0.572	
		Myrmicinae (n=14)	0.309	0.579	0.647	
		Paraponerinae (n=2)	3.750	0.053	0.148	
		Ponerinae (n=9)	0.004	0.947	0.947	
		Pseudomyrmicinae (n=3)	1.800	0.180	0.280	
	Dorylinae (n=2)	Ectatomminae (n=3)	3.000	0.083	0.155	
		Formicinae (n=8)	0.274	0.600	0.647	
		Myrmicinae (n=14)	2.042	0.153	0.252	
		Paraponerinae (n=2)	2.400	0.121	0.212	
		Ponerinae (n=9)	0.889	0.346	0.484	
		Pseudomyrmicinae (n=3)	3.000	0.083	0.155	
	Ectatomminae (n=3)	Formicinae (n=8)	4.186	0.041	0.148	
		Myrmicinae (n=14)	6.349	0.012	0.148	
		Paraponerinae (n=2)	0.333	0.564	0.647	
		Ponerinae (n=9)	5.342	0.021	0.148	
		Pseudomyrmicinae (n=3)	3.857	0.049	0.148	
	Formicinae (n=8)	Myrmicinae (n=14)	1.683	0.195	0.287	
		Paraponerinae (n=2)	4.390	0.036	0.148	
		Ponerinae (n=9)	0.522	0.470	0.572	
		Pseudomyrmicinae (n=3)	4.186	0.041	0.148	
	Myrmicinae (n=14)	Paraponerinae (n=2)	4.941	0.026	0.148	
		Ponerinae (n=9)	0.036	0.850	0.882	
		Pseudomyrmicinae (n=3)	5.439	0.020	0.148	
	Paraponerinae (n=2)	Ponerinae (n=9)	3.555	0.059	0.151	
		Pseudomyrmicinae (n=3)	3.000	0.083	0.155	
	Ponerinae (n=9)	Pseudomyrmicinae (n=3)	3.085	0.079	0.155	
	Genus	<i>Anochetus</i> (n=1)	<i>Atta</i> (n=3)	1.800	0.180	0.411
			<i>Azteca</i> (n=3)	0.200	0.655	0.814
<i>Camponotus</i> (n=6)			0.250	0.617	0.800	
<i>Cephalotes</i> (n=4)			2.000	0.157	0.410	
<i>Crematogaster</i> (n=2)			1.500	0.221	0.412	
<i>Daceton</i> (n=2)			1.500	0.221	0.412	
<i>Dolichoderus</i> (n=2)			0.000	1.000	1.000	
<i>Eciton</i> (n=1)			1.000	0.317	0.511	
<i>Ectatomma</i> (n=3)			1.800	0.180	0.410	

		<i>Gigantiops (n=2)</i>	2.000	0.157	0.410
		<i>Labidus (n=1)</i>	1.000	0.317	0.511
		<i>Neoponera (n=2)</i>	0.000	1.000	1.000
		<i>Nylanderia (n=1)</i>	1.000	0.170	0.511
		<i>Odontomachus (n=6)</i>	0.250	0.617	0.800
		<i>Paraponera (n=2)</i>	1.500	0.221	0.412
		<i>Pseudomyrmex (n=3)</i>	1.800	0.178	0.410
		<i>Solenopsis (n=2)</i>	1.500	0.221	0.412
	<i>Atta (n=3)</i>	<i>Azteca (n=3)</i>	2.330	0.127	0.410
		<i>Camponotus (n=6)</i>	5.400	0.020	0.410
		<i>Cephalotes (n=4)</i>	4.500	0.034	0.410
		<i>Crematogaster (n=2)</i>	3.000	0.083	0.410
		<i>Daceton (n=2)</i>	3.000	0.083	0.410
		<i>Dolichoderus (n=2)</i>	0.000	1.000	1.000
		<i>Eciton (n=1)</i>	0.200	0.655	0.814
		<i>Ectatomma (n=3)</i>	2.330	0.127	0.410
		<i>Gigantiops (n=2)</i>	3.158	0.076	0.410
		<i>Labidus (n=1)</i>	1.800	0.180	0.410
		<i>Neoponera (n=2)</i>	3.000	0.083	0.410
		<i>Nylanderia (n=1)</i>	1.800	0.180	0.410
		<i>Odontomachus (n=6)</i>	2.400	0.121	0.410
		<i>Paraponera (n=2)</i>	3.000	0.083	0.410
		<i>Pseudomyrmex (n=3)</i>	3.857	0.049	0.410
		<i>Solenopsis (n=2)</i>	3.000	0.083	0.410
	<i>Azteca (n=3)</i>	<i>Camponotus (n=6)</i>	0.067	0.796	0.952
		<i>Cephalotes (n=4)</i>	0.500	0.480	0.719
		<i>Crematogaster (n=2)</i>	0.330	0.564	0.777
		<i>Daceton (n=2)</i>	0.330	0.564	0.777
		<i>Dolichoderus (n=2)</i>	0.000	1.000	1.000
		<i>Eciton (n=1)</i>	1.800	0.180	0.410
		<i>Ectatomma (n=3)</i>	3.857	0.049	0.410
		<i>Gigantiops (n=2)</i>	3.158	0.076	0.410
		<i>Labidus (n=1)</i>	0.200	0.655	0.814
		<i>Neoponera (n=2)</i>	0.330	0.564	0.777
		<i>Nylanderia (n=1)</i>	0.200	0.655	0.814
		<i>Odontomachus (n=6)</i>	0.267	0.606	0.800
		<i>Paraponera (n=2)</i>	3.000	0.083	0.410
	<i>Pseudomyrmex (n=3)</i>	1.190	0.275	0.490	
	<i>Solenopsis (n=2)</i>	0.333	0.564	0.777	
	<i>Camponotus (n=6)</i>	<i>Cephalotes (n=4)</i>	1.136	0.286	0.504
		<i>Crematogaster (n=2)</i>	0.110	0.739	0.897
		<i>Daceton (n=2)</i>	1.778	0.182	0.410
		<i>Dolichoderus (n=2)</i>	0.111	0.739	0.897
		<i>Eciton (n=1)</i>	2.250	0.134	0.410
		<i>Ectatomma (n=3)</i>	5.400	0.020	0.410
		<i>Gigantiops (n=2)</i>	4.048	0.044	0.410
		<i>Labidus (n=1)</i>	0.250	0.617	0.800

	<i>Neoponera (n=2)</i>	0.028	0.867	1.000	
	<i>Nylanderia (n=1)</i>	0.250	0.617	0.800	
	<i>Odontomachus (n=6)</i>	0.103	0.749	0.902	
	<i>Paraponera (n=2)</i>	4.000	0.046	0.410	
	<i>Pseudomyrmex (n=3)</i>	3.267	0.070	0.410	
	<i>Solenopsis (n=2)</i>	1.000	0.317	0.511	
<i>Cephalotes (n=4)</i>	<i>Crematogaster (n=2)</i>	0.000	1.000	1.000	
	<i>Daceton (n=2)</i>	3.429	0.064	0.410	
	<i>Dolichoderus (n=2)</i>	0.000	1.000	1.000	
	<i>Eciton (n=1)</i>	2.000	0.157	0.410	
	<i>Ectatomma (n=3)</i>	4.500	0.034	0.410	
	<i>Gigantiops (n=2)</i>	3.529	0.060	0.410	
	<i>Labidus (n=1)</i>	2.000	0.157	0.410	
	<i>Neoponera (n=2)</i>	0.000	1.000	1.000	
	<i>Nylanderia (n=1)</i>	0.500	0.480	0.719	
	<i>Odontomachus (n=6)</i>	0.046	0.830	0.985	
	<i>Paraponera (n=2)</i>	3.429	0.064	0.410	
	<i>Pseudomyrmex (n=3)</i>	4.500	0.034	0.410	
		<i>Solenopsis (n=2)</i>	0.214	0.643	0.814
	<i>Crematogaster (n=2)</i>	<i>Daceton (n=2)</i>	0.600	0.439	0.671
<i>Dolichoderus (n=2)</i>		0.000	1.000	1.000	
<i>Eciton (n=1)</i>		1.500	0.221	0.412	
<i>Ectatomma (n=3)</i>		3.000	0.083	0.410	
<i>Gigantiops (n=2)</i>		2.667	0.102	0.410	
<i>Labidus (n=1)</i>		0.000	1.000	1.000	
<i>Neoponera (n=2)</i>		0.000	1.000	1.000	
<i>Nylanderia (n=1)</i>		0.000	1.000	1.000	
<i>Odontomachus (n=6)</i>		0.000	1.000	1.000	
<i>Paraponera (n=2)</i>		2.400	0.121	0.410	
<i>Pseudomyrmex (n=3)</i>		3.000	0.083	0.410	
		<i>Solenopsis (n=2)</i>	0.600	0.439	0.671
<i>Daceton (n=2)</i>	<i>Dolichoderus (n=2)</i>	0.000	1.000	1.000	
	<i>Eciton (n=1)</i>	1.500	0.221	0.412	
	<i>Ectatomma (n=3)</i>	3.000	0.083	0.410	
	<i>Gigantiops (n=2)</i>	2.667	0.102	0.410	
	<i>Labidus (n=1)</i>	1.500	0.221	0.412	
	<i>Neoponera (n=2)</i>	0.000	1.000	1.000	
	<i>Nylanderia (n=1)</i>	1.500	0.221	0.412	
	<i>Odontomachus (n=6)</i>	0.444	0.505	0.743	
	<i>Paraponera (n=2)</i>	2.400	0.121	0.410	
	<i>Pseudomyrmex (n=3)</i>	1.333	0.248	0.447	
		<i>Solenopsis (n=2)</i>	0.000	1.000	1.000
<i>Dolichoderus (n=2)</i>	<i>Eciton (n=1)</i>	0.000	1.000	1.000	
	<i>Ectatomma (n=3)</i>	1.333	0.248	0.447	
	<i>Gigantiops (n=2)</i>	2.667	0.102	0.410	
	<i>Labidus (n=1)</i>	0.000	1.000	1.000	
	<i>Neoponera (n=2)</i>	0.600	0.439	0.671	

	<i>Nylanderia (n=1)</i>	0.000	1.000	1.000
	<i>Odontomachus (n=6)</i>	0.000	1.000	1.000
	<i>Paraponera (n=2)</i>	2.400	0.121	0.410
	<i>Pseudomyrmex (n=3)</i>	1.333	0.248	0.447
	<i>Solenopsis (n=2)</i>	0.600	0.439	0.671
<i>Eciton (n=1)</i>	<i>Ectatomma (n=3)</i>	1.800	0.180	0.410
	<i>Gigantiops (n=2)</i>	2.000	0.157	0.410
	<i>Labidus (n=1)</i>	1.000	0.317	0.511
	<i>Neoponera (n=2)</i>	1.500	0.221	0.412
	<i>Nylanderia (n=1)</i>	1.000	0.317	0.511
	<i>Odontomachus (n=6)</i>	1.000	0.317	0.511
	<i>Paraponera (n=2)</i>	1.500	0.221	0.412
	<i>Pseudomyrmex (n=3)</i>	1.800	0.180	0.410
	<i>Solenopsis (n=2)</i>	1.500	0.221	0.412
<i>Ectatomma (n=3)</i>	<i>Gigantiops (n=2)</i>	0.351	0.554	0.777
	<i>Labidus (n=1)</i>	1.800	0.180	0.410
	<i>Neoponera (n=2)</i>	3.000	0.083	0.410
	<i>Nylanderia (n=1)</i>	1.800	0.180	0.410
	<i>Odontomachus (n=6)</i>	4.267	0.039	0.410
	<i>Paraponera (n=2)</i>	0.333	0.564	0.777
	<i>Pseudomyrmex (n=3)</i>	3.857	0.049	0.410
	<i>Solenopsis (n=2)</i>	3.000	0.083	0.410
<i>Gigantiops (n=2)</i>	<i>Labidus (n=1)</i>	2.000	0.157	0.410
	<i>Neoponera (n=2)</i>	2.667	0.102	0.410
	<i>Nylanderia (n=1)</i>	2.000	0.157	0.410
	<i>Odontomachus (n=6)</i>	1.799	0.180	0.410
	<i>Paraponera (n=2)</i>	2.667	0.102	0.410
	<i>Pseudomyrmex (n=3)</i>	3.158	0.076	0.410
	<i>Solenopsis (n=2)</i>	2.667	0.102	0.410
<i>Labidus (n=1)</i>	<i>Neoponera (n=2)</i>	0.000	1.000	1.000
	<i>Nylanderia (n=1)</i>	1.000	0.317	0.511
	<i>Odontomachus (n=6)</i>	0.250	0.617	0.800
	<i>Paraponera (n=2)</i>	1.500	0.221	0.412
	<i>Pseudomyrmex (n=3)</i>	1.800	0.180	0.410
	<i>Solenopsis (n=2)</i>	1.500	0.221	0.412
<i>Neoponera (n=2)</i>	<i>Nylanderia (n=1)</i>	0.000	1.000	1.000
	<i>Odontomachus (n=6)</i>	0.111	0.739	0.897
	<i>Paraponera (n=2)</i>	2.400	0.121	0.410
	<i>Pseudomyrmex (n=3)</i>	0.333	0.564	0.777
	<i>Solenopsis (n=2)</i>	0.000	1.000	1.000
<i>Nylanderia (n=1)</i>	<i>Odontomachus (n=6)</i>	0.250	0.617	0.800
	<i>Paraponera (n=2)</i>	1.500	0.221	0.412
	<i>Pseudomyrmex (n=3)</i>	1.800	0.180	0.410
	<i>Solenopsis (n=2)</i>	0.000	1.000	1.000
<i>Odontomachus (n=6)</i>	<i>Paraponera (n=2)</i>	2.778	0.096	0.410
	<i>Pseudomyrmex (n=3)</i>	3.267	0.071	0.410
	<i>Solenopsis (n=2)</i>	0.444	0.505	0.743

	<i>Paraponera</i> (n=2)	<i>Pseudomyrmex</i> (n=3)	3.000	0.083	0.410
		<i>Solenopsis</i> (n=2)	2.400	0.121	0.410
	<i>Pseudomyrmex</i> (n=3)	<i>Solenopsis</i> (n=2)	0.789	0.374	0.596
Diet	Carnivorous (n=13)	Herbivorous (n=9)	0.356	0.551	0.826
		Omnivorous (n=24)	0.024	0.877	0.877
	Herbivorous (n=9)	Omnivorous (n=24)	0.372	0.542	0.826
Nest	Ground (n=21)	Rotten (n=7)	8.832	0.003	0.004
		Tree (n=18)	12.802	0.006	0.001
	Rotten (n=7)	Tree (n=18)	0.023	0.880	0.880
Habitat	Rainforest (n=24)	City (n=22)	0.148	0.700	0.700

ASV richness forest/city dataset

Factor	Group 1	Group 2	H	p-value	q-value
Subfamily	Dolichoderinae (n=2)	Ectatomminae (n=2)	2.400	0.121	0.213
		Formicinae (n=4)	0.857	0.355	0.483
		Myrmicinae (n=6)	0.000	1.000	1.000
		Paraponerinae (n=2)	2.400	0.121	0.213
		Ponerinae (n=5)	0.150	0.699	0.806
	Ectatomminae (n=2)	Formicinae (n=4)	0.429	0.064	0.160
		Myrmicinae (n=6)	4.000	0.046	0.160
		Paraponerinae (n=2)	0.000	1.000	1.000
		Ponerinae (n=5)	3.750	0.053	0.160
	Formicinae (n=4)	Myrmicinae (n=6)	2.227	0.136	0.212
		Paraponerinae (n=2)	3.429	0.064	0.160
		Ponerinae (n=5)	2.160	0.142	0.212
	Myrmicinae (n=6)	Paraponerinae (n=2)	4.000	0.046	0.160
		Ponerinae (n=5)	0.410	0.522	0.652
	Paraponerinae (n=2)	Ponerinae (n=5)	3.750	0.053	0.160
	Genus	<i>A. cephalotes</i> (n=2)	<i>A. chartifex</i> (n=2)	0.000	1.000
<i>C. atriceps</i> (n=2)			2.400	0.121	0.182
<i>C. atratus</i> (n=2)			2.400	0.121	0.182
<i>D. armigerum</i> (n=2)			2.400	0.121	0.182
<i>E. tuberculatum</i> (n=2)			2.400	0.121	0.182
<i>G. destructor</i> (n=2)			2.400	0.121	0.182
<i>N. commutata</i> (n=2)			2.400	0.121	0.182
<i>O. haematodus</i> (n=2)			3.000	0.083	0.182
<i>P. clavata</i> (n=2)			2.400	0.121	0.182
<i>A. chartifex</i> (n=2)		<i>C. atriceps</i> (n=2)	0.000	1.000	1.000
		<i>C. atratus</i> (n=2)	0.000	1.000	1.000
		<i>D. armigerum</i> (n=2)	0.000	1.000	1.000
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.600	0.439	0.599
		<i>O. haematodus</i> (n=2)	0.000	1.000	1.000
<i>P. clavata</i> (n=2)	2.400	0.121	0.182		

	<i>C. atriceps</i> (n=2)	<i>C. atratus</i> (n=2)	0.600	0.439	0.598
		<i>D. armigerum</i> (n=2)	2.400	0.121	0.182
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.000	1.000	1.000
		<i>O. haematodus</i> (n=2)	0.333	0.563	0.746
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>C. atratus</i> (n=2)	<i>D. armigerum</i> (n=2)	2.400	0.121	0.182
		<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.000	1.000	1.000
		<i>O. haematodus</i> (n=2)	0.088	0.767	0.986
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>D. armigerum</i> (n=2)	<i>E. tuberculatum</i> (n=2)	2.400	0.121	0.182
		<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	0.000	1.000	1.000
		<i>O. haematodus</i> (n=2)	1.333	0.248	0.360
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>E. tuberculatum</i> (n=2)	<i>G. destructor</i> (n=2)	2.400	0.121	0.182
		<i>N. commutata</i> (n=2)	2.400	0.121	0.182
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	0.000	1.000	1.000
	<i>G. destructor</i> (n=2)	<i>N. commutata</i> (n=2)	2.400	0.121	0.182
		<i>O. haematodus</i> (n=2)	3.000	0.083	0.182
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
	<i>N. commutata</i> (n=2)	<i>O. haematodus</i> (n=2)	0.000	1.000	1.000
		<i>P. clavata</i> (n=2)	2.400	0.121	0.182
<i>O. haematodus</i> (n=2)	<i>P. clavata</i> (n=2)	3.000	0.083	0.182	
Diet	Carnivorous (n=7)	Herbivorous (n=4)	0.722	0.395	0.395
		Omnivorous (n=10)	1.333	0.248	0.372
	Herbivorous (n=4)	Omnivorous (n=10)	1.413	0.234	0.372
Nest	Ground (n=11)	Rotten (n=4)	5.523	0.019	0.031
		Tree (n=6)	5.343	0.021	0.031
	Rotten (n=4)	Tree (n=6)	0.286	0.593	0.593
Habitat	Rainforest (n=10)	City (n=11)	0.011	0.916	0.916

Supplementary File 3. Bacteria relative abundance of the 15 most abundant bacterial order for each factor. The cells colored in red in each table bacterial abundance higher than 10%.

Factors		Relative abundance of the 15 most abundant bacterial orders (%)														
		Rickettsiales	Rhizobiales	Acetobacterales	Enterobacteriales	Lactobacillales	Burkholderiales	Xanthomonadales	Erysipelotrichales	Flavobacteriales	Pseudomonadales	Corynebacteriales	Entomoplasmatales	Opitutales	Sphingomonadales	Micrococcales
Subfamily	Dolichoderinae (n=6)	39.33	15.36	8.96	0.14	14.41	0.66	0.55	0.00	0.10	0.73	6.98	0.00	0.02	0.24	4.55
	Dorylinae (n=1)	0.05	1.83	0.26	1.87	3.68	5.11	1.39	48.64	10.69	4.13	0.20	10.08	0.00	0.52	0.71
	Ectatomminae (n=3)	0.08	9.62	1.04	0.85	25.04	5.94	3.62	0.03	0.53	2.73	2.28	0.00	0.19	3.81	0.25
	Formicinae (n=6)	13.21	2.88	19.50	33.98	3.27	1.85	1.18	0.01	0.18	0.41	3.81	0.00	0.03	2.83	0.20
	Myrmicinae (n=14)	23.04	20.49	8.59	0.46	3.74	7.72	6.78	0.01	1.22	2.52	2.05	1.76	2.34	0.72	0.32
	Paraponerinae (n=2)	0.05	5.03	0.70	0.55	1.17	4.10	6.03	0.05	0.10	20.60	0.84	0.00	0.06	0.37	0.43
	Ponerinae (n=9)	48.58	18.00	0.22	0.11	1.54	0.66	0.79	5.71	7.86	0.29	0.16	0.00	0.10	0.12	0.11
	Pseudomyrmicinae (n=3)	0.00	0.07	3.40	0.00	0.10	0.15	0.00	0.00	0.03	0.03	0.00	0.55	0.00	0.00	0.03
Genus	Cephalotes (n=4)	2.32	19.19	0.09	0.08	1.44	22.71	24.65	0.00	3.58	0.02	0.00	0.00	8.72	0.00	0.63
	Crematogaster (n=2)	78.58	10.85	0.07	0.00	4.74	0.71	0.04	0.00	0.49	0.31	0.27	0.00	0.06	0.17	0.17
	Daceton (n=2)	0.10	72.49	11.31	0.05	0.05	0.05	0.86	0.00	0.01	0.27	0.00	13.20	0.00	0.00	0.00
	Dolichoderus (n=2)	49.05	38.06	0.31	0.29	0.56	0.41	0.19	0.00	0.13	0.05	0.79	0.00	0.00	0.45	0.38
	Ectiton (n=1)	0.00	3.29	0.00	3.73	1.61	8.83	2.67	34.47	21.37	4.68	0.29	0.93	0.00	0.87	0.00
	Ectatomma (n=3)	0.08	9.62	1.04	0.85	25.04	5.94	3.62	0.03	0.53	2.73	2.28	0.00	0.22	3.81	0.71
	Gigantiops (n=2)	0.05	10.46	22.59	0.32	0.38	6.61	3.33	0.00	0.22	1.87	11.28	0.00	0.00	11.04	0.71
	Labidus (n=1)	0.00	0.37	0.09	0.00	5.75	1.40	0.00	62.80	0.00	3.75	0.17	19.24	0.00	0.18	0.00
	Neoponera (n=2)	0.08	45.07	0.08	0.00	1.77	2.05	3.10	5.70	35.31	0.39	0.25	0.02	0.10	0.31	0.27
	Nylanderia (n=1)	0.00	0.77	89.50	0.27	0.10	0.35	0.10	0.10	0.84	0.92	0.56	0.00	0.00	0.17	0.20
	Odontomachus (n=6)	72.87	11.60	0.10	0.17	1.66	0.21	0.12	6.66	0.03	0.29	0.06	0.00	0.02	0.02	0.06
	Paraponera (n=2)	0.05	5.03	0.70	0.55	1.17	4.10	6.03	0.05	0.20	20.60	0.84	0.00	0.06	0.36	0.43
	Pseudomyrmex (n=3)	62.61	0.07	3.40	0.00	0.10	0.15	0.00	0.00	0.03	0.03	0.00	0.55	0.00	0.00	0.02
	Solenopsis (n=2)	59.59	1.38	0.04	1.62	3.94	5.21	0.38	0.03	0.03	3.76	0.08	0.00	0.00	0.00	0.00
	Diet	Carnivorous (n=13)	29.14	21.38	1.73	0.38	1.57	1.65	1.57	9.92	6.16	3.49	0.23	3.10	0.03	0.18
Herbivorous (n=9)		11.90	23.63	1.90	0.51	4.40	11.00	10.98	0.00	1.86	2.72	3.45	0.00	3.89	1.25	0.58
Omnivorous (n=24)		33.08	3.57	13.76	12.71	8.61	2.40	1.08	0.02	0.22	2.23	3.27	0.08	0.04	1.59	1.21
Nest	Ground (n=21)	19.90	10.25	4.36	3.90	7.65	3.51	1.99	4.98	4.44	5.42	3.13	0.92	0.07	2.13	0.34
	Rotten (n=7)	41.62	0.28	24.53	28.69	0.82	0.16	0.10	0.01	0.16	0.21	0.24	0.00	0.01	0.07	0.07
	Tree (n=18)	30.24	21.36	4.40	0.06	4.40	5.45	5.77	2.20	0.88	0.31	1.99	1.56	1.95	0.07	1.43
Habitat	City (n=22)	23.60	8.98	10.93	8.39	8.52	4.45	3.34	0.00	3.81	1.78	3.33	0.44	1.03	1.85	0.28
	Rainforest (n=24)	30.39	16.87	3.95	3.94	2.34	3.09	2.97	6.19	1.08	3.55	2.13	1.59	0.50	0.25	1.13

Supplementary File 4. Statistic calculations of beta diversity metrics for each factor. The two tested beta diversity metrics are Bray-Curtis and Weighted unifracs. Each metric is calculated for the whole dataset and the forest/city dataset. The cells colored in red in each table represent p-values < 0.05. The statistical p-values were obtained with PERMANOVAs.

Bray-Curtis distance whole dataset

Factor	Group 1	Group 2	pseudo-F	p-value	q-value	
Subfamily	Dolichoderinae (n=5)	Dorylinae (n=2)	1.694	0.045	0.165	
		Ectatomminae (n=3)	1.149	0.106	0.165	
		Formicinae (n=8)	1.471	0.037	0.165	
		Myrmicinae (n=14)	1.117	0.230	0.293	
		Paraponerinae (n=2)	1.164	0.097	0.165	
		Ponerinae (n=9)	1.021	0.287	0.349	
		Pseudomyrmicinae (n=3)	0.833	0.726	0.726	
	Dorylinae (n=2)	Ectatomminae (n=3)	1.692	0.088	0.165	
		Formicinae (n=8)	1.832	0.071	0.165	
		Myrmicinae (n=14)	1.675	0.030	0.165	
		Paraponerinae (n=2)	1.948	0.368	0.429	
		Ponerinae (n=9)	1.981	0.023	0.165	
		Pseudomyrmicinae (n=3)	1.650	0.123	0.172	
	Ectatomminae (n=3)	Formicinae (n=8)	1.295	0.145	0.193	
		Myrmicinae (n=14)	1.183	0.105	0.165	
		Paraponerinae (n=2)	0.978	0.633	0.656	
		Ponerinae (n=9)	1.424	0.076	0.165	
		Pseudomyrmicinae (n=3)	1.079	0.093	0.165	
	Formicinae (n=8)	Myrmicinae (n=14)	1.712	0.001	0.028	
		Paraponerinae (n=2)	1.315	0.084	0.165	
		Ponerinae (n=9)	2.068	0.006	0.084	
		Pseudomyrmicinae (n=3)	1.282	0.117	0.172	
	Myrmicinae (n=14)	Paraponerinae (n=2)	1.196	0.047	0.165	
		Ponerinae (n=9)	1.421	0.079	0.165	
		Pseudomyrmicinae (n=3)	0.966	0.526	0.589	
	Paraponerinae (n=2)	Ponerinae (n=9)	1.396	0.068	0.165	
		Pseudomyrmicinae (n=3)	1.089	0.103	0.165	
	Ponerinae (n=9)	Pseudomyrmicinae (n=3)	0.756	0.577	0.621	
	Genus	<i>Atta</i> (n=3)	<i>Azteca</i> (n=3)	1.255	0.104	0.284
			<i>Camponotus</i> (n=6)	1.761	0.092	0.284
<i>Cephalotes</i> (n=4)			2.111	0.022	0.284	
<i>Crematogaster</i> (n=2)			1.245	0.221	0.357	
<i>Daceton</i> (n=2)			2.681	0.118	0.295	
<i>Dolichoderus</i> (n=2)			1.249	0.199	0.349	
<i>Ectatomma</i> (n=3)			1.222	0.212	0.349	
<i>Gigantiops</i> (n=2)			1.240	0.187	0.344	

	<i>NA (n=3)</i>	1.428	0.084	0.284
	<i>Neoponera (n=2)</i>	1.229	0.207	0.349
	<i>Odontomachus (n=6)</i>	2.284	0.035	0.284
	<i>Paraponera (n=2)</i>	1.317	0.201	0.349
	<i>Pseudomyrmex (n=3)</i>	1.245	0.107	0.284
	<i>Solenopsis (n=2)</i>	1.239	0.185	0.344
<i>Azteca (n=3)</i>	<i>Camponotus (n=6)</i>	1.511	0.109	0.284
	<i>Cephalotes (n=4)</i>	1.789	0.028	0.284
	<i>Crematogaster (n=2)</i>	0.866	0.486	0.544
	<i>Daceton (n=2)</i>	2.206	0.093	0.284
	<i>Dolichoderus (n=2)</i>	0.890	0.487	0.544
	<i>Ectatomma (n=3)</i>	1.087	0.108	0.284
	<i>Gigantiops (n=2)</i>	1.053	0.213	0.349
	<i>NA (n=3)</i>	1.225	0.167	0.337
	<i>Neoponera (n=2)</i>	1.030	0.321	0.406
	<i>Odontomachus (n=6)</i>	1.263	0.210	0.349
	<i>Paraponera (n=2)</i>	1.116	0.106	0.284
	<i>Pseudomyrmex (n=3)</i>	0.809	0.611	0.642
	<i>Solenopsis (n=2)</i>	0.757	0.598	0.634
	<i>Camponotus (n=6)</i>	<i>Cephalotes (n=4)</i>	2.490	0.007
<i>Crematogaster (n=2)</i>		1.438	0.136	0.304
<i>Daceton (n=2)</i>		2.648	0.069	0.284
<i>Dolichoderus (n=2)</i>		1.428	0.146	0.307
<i>Ectatomma (n=3)</i>		1.519	0.100	0.284
<i>Gigantiops (n=2)</i>		1.442	0.142	0.304
<i>NA (n=3)</i>		1.752	0.053	0.284
<i>Neoponera (n=2)</i>		1.418	0.177	0.344
<i>Odontomachus (n=6)</i>		2.880	0.001	0.105
<i>Paraponera (n=2)</i>		1.525	0.096	0.284
<i>Pseudomyrmex (n=3)</i>		1.473	0.136	0.304
<i>Solenopsis (n=2)</i>		1.420	0.165	0.337
<i>Cephalotes (n=4)</i>	<i>Crematogaster (n=2)</i>	1.746	0.064	0.284
	<i>Daceton (n=2)</i>	3.509	0.087	0.284
	<i>Dolichoderus (n=2)</i>	1.744	0.074	0.284
	<i>Ectatomma (n=3)</i>	1.844	0.035	0.284
	<i>Gigantiops (n=2)</i>	1.779	0.081	0.284
	<i>NA (n=3)</i>	2.056	0.022	0.284
	<i>Neoponera (n=2)</i>	1.735	0.074	0.284
	<i>Odontomachus (n=6)</i>	3.176	0.011	0.284
	<i>Paraponera (n=2)</i>	1.880	0.066	0.284
	<i>Pseudomyrmex (n=3)</i>	1.750	0.023	0.284
	<i>Solenopsis (n=2)</i>	1.732	0.067	0.284
<i>Crematogaster (n=2)</i>	<i>Daceton (n=2)</i>	2.578	0.334	0.406
	<i>Dolichoderus (n=2)</i>	0.834	0.695	0.708
	<i>Ectatomma (n=3)</i>	1.075	0.125	0.304
	<i>Gigantiops (n=2)</i>	1.030	0.322	0.406
	<i>NA (n=3)</i>	1.180	0.263	0.406

	<i>Neoponera (n=2)</i>	1.008	0.320	0.406
	<i>Odontomachus (n=6)</i>	0.999	0.349	0.406
	<i>Paraponera (n=2)</i>	1.105	0.338	0.406
	<i>Pseudomyrmex (n=3)</i>	0.840	0.503	0.550
	<i>Solenopsis (n=2)</i>	0.795	0.649	0.674
<i>Daceton (n=2)</i>	<i>Dolichoderus (n=2)</i>	2.579	0.327	0.406
	<i>Ectatomma (n=3)</i>	2.317	0.101	0.284
	<i>Gigantiops (n=2)</i>	2.657	0.334	0.406
	<i>NA (n=3)</i>	2.373	0.071	0.284
	<i>Neoponera (n=2)</i>	2.569	0.348	0.406
	<i>Odontomachus (n=6)</i>	3.392	0.044	0.284
	<i>Paraponera (n=2)</i>	2.879	0.332	0.406
	<i>Pseudomyrmex (n=3)</i>	2.137	0.092	0.284
	<i>Solenopsis (n=2)</i>	2.552	0.323	0.406
	<i>Ectatomma (n=3)</i>	1.062	0.111	0.284
	<i>Gigantiops (n=2)</i>	1.035	0.341	0.406
	<i>NA (n=3)</i>	1.191	0.132	0.304
	<i>Neoponera (n=2)</i>	1.003	0.337	0.406
	<i>Odontomachus (n=6)</i>	1.373	0.140	0.304
	<i>Paraponera (n=2)</i>	1.089	0.337	0.406
	<i>Pseudomyrmex (n=3)</i>	0.861	0.503	0.550
	<i>Solenopsis (n=2)</i>	0.826	0.655	0.674
<i>Ectatomma (n=3)</i>	<i>Gigantiops (n=2)</i>	1.029	0.419	0.478
	<i>NA (n=3)</i>	1.223	0.267	0.406
	<i>Neoponera (n=2)</i>	1.042	0.198	0.349
	<i>Odontomachus (n=6)</i>	1.955	0.053	0.284
	<i>Paraponera (n=2)</i>	0.978	0.589	0.632
	<i>Pseudomyrmex (n=3)</i>	1.079	0.099	0.284
	<i>Solenopsis (n=2)</i>	1.071	0.096	0.284
<i>Gigantiops (n=2)</i>	<i>NA (n=3)</i>	1.178	0.330	0.406
	<i>Neoponera (n=2)</i>	1.028	0.346	0.406
	<i>Odontomachus (n=6)</i>	1.856	0.037	0.284
	<i>Paraponera (n=2)</i>	1.093	0.352	0.406
	<i>Pseudomyrmex (n=3)</i>	1.029	0.108	0.284
	<i>Solenopsis (n=2)</i>	1.019	0.299	0.406
<i>NA (n=3)</i>	<i>Neoponera (n=2)</i>	1.167	0.312	0.406
	<i>Odontomachus (n=6)</i>	2.293	0.009	0.284
	<i>Paraponera (n=2)</i>	1.255	0.142	0.304
	<i>Pseudomyrmex (n=3)</i>	1.195	0.171	0.339
	<i>Solenopsis (n=2)</i>	1.169	0.312	0.406
<i>Neoponera (n=2)</i>	<i>Odontomachus (n=6)</i>	1.811	0.186	0.344
	<i>Paraponera (n=2)</i>	1.058	0.350	0.406
	<i>Pseudomyrmex (n=3)</i>	1.000	0.295	0.406
	<i>Solenopsis (n=2)</i>	0.996	1.000	1.000
<i>Odontomachus (n=6)</i>	<i>Paraponera (n=2)</i>	1.936	0.042	0.284
	<i>Pseudomyrmex (n=3)</i>	0.934	0.290	0.406
	<i>Solenopsis (n=2)</i>	0.676	0.590	0.632

	<i>Paraponera</i> (n=2)	<i>Pseudomyrmex</i> (n=3)	1.089	0.088	0.284
		<i>Solenopsis</i> (n=2)	1.093	0.313	0.406
	<i>Pseudomyrmex</i> (n=3)	<i>Solenopsis</i> (n=2)	0.600	1.000	1.000
Diet	Carnivorous (n=13)	Herbivorous (n=9)	1.618	0.012	0.018
		Omnivorous (n=24)	1.096	0.255	0.255
	Herbivorous (n=9)	Omnivorous (n=24)	1.482	0.007	0.018
Nest	Ground (n=21)	Rotten (n=7)	1.659	0.016	0.042
		Tree (n=18)	1.268	0.049	0.049
	Rotten (n=7)	Tree (n=18)	1.601	0.028	0.042
Habitat	Rainforest (n=24)	City (n=22)	0.920	0.619	0.619

Bray-Curtis distance forest/city dataset

Factor	Group 1	Group 2	pseudo-F	p-value	q-value
Subfamily	<i>Dolichoderinae</i> (n=2)	<i>Ectatomminae</i> (n=2)	1.111	0.366	0.392
		<i>Formicinae</i> (n=4)	1.379	0.112	0.202
		<i>Myrmicinae</i> (n=6)	1.249	0.148	0.202
		<i>Paraponerinae</i> (n=2)	1.125	0.325	0.375
		<i>Ponerinae</i> (n=5)	1.828	0.049	0.140
	<i>Ectatomminae</i> (n=2)	<i>Formicinae</i> (n=4)	1.380	0.142	0.202
		<i>Myrmicinae</i> (n=6)	1.250	0.193	0.241
		<i>Paraponerinae</i> (n=2)	0.971	0.664	0.664
		<i>Ponerinae</i> (n=5)	1.868	0.056	0.140
	<i>Formicinae</i> (n=4)	<i>Myrmicinae</i> (n=6)	1.704	0.019	0.140
		<i>Paraponerinae</i> (n=2)	1.398	0.148	0.202
		<i>Ponerinae</i> (n=5)	2.402	0.035	0.140
	<i>Myrmicinae</i> (n=6)	<i>Paraponerinae</i> (n=2)	1.282	0.111	0.202
		<i>Ponerinae</i> (n=5)	2.301	0.009	0.135
<i>Paraponerinae</i> (n=2)	<i>Ponerinae</i> (n=5)	1.855	0.051	0.140	
Genus	<i>A. cephalotes</i> (n=2)	<i>A. chartifex</i> (n=2)	1.097	0.358	0.374
		<i>C. atriceps</i> (n=2)	3.169	0.332	0.374
		<i>C. atratus</i> (n=2)	1.811	0.345	0.374
		<i>D. armigerum</i> (n=2)	2.735	0.344	0.374
		<i>E. tuberculatum</i> (n=2)	1.067	0.308	0.374
		<i>G. destructor</i> (n=2)	1.075	0.350	0.374
		<i>N. commutata</i> (n=2)	1.039	0.365	0.374
		<i>O. haematodus</i> (n=2)	5.695	0.098	0.374
		<i>P. clavata</i> (n=2)	1.105	0.343	0.374
	<i>A. chartifex</i> (n=2)	<i>C. atriceps</i> (n=2)	3.086	0.312	0.374
		<i>C. atratus</i> (n=2)	1.768	0.340	0.374
		<i>D. armigerum</i> (n=2)	2.651	0.353	0.374
		<i>E. tuberculatum</i> (n=2)	1.111	0.352	0.374
		<i>G. destructor</i> (n=2)	1.066	0.331	0.374
<i>N. commutata</i> (n=2)	1.041	0.340	0.374		
<i>O. haematodus</i> (n=2)	5.568	0.110	0.374		

		<i>P. clavata</i> (n=2)	1.125	0.318	0.374
	<i>C. atriceps</i> (n=2)	<i>C. atratus</i> (n=2)	6.976	0.321	0.374
		<i>D. armigerum</i> (n=2)	25.789	0.356	0.374
		<i>E. tuberculatum</i> (n=2)	3.276	0.316	0.374
		<i>G. destructor</i> (n=2)	3.031	0.336	0.374
		<i>N. commutata</i> (n=2)	1.920	0.337	0.374
		<i>O. haematodus</i> (n=2)	187.893	0.115	0.374
		<i>P. clavata</i> (n=2)	3.258	0.341	0.374
	<i>C. atratus</i> (n=2)	<i>D. armigerum</i> (n=2)	5.518	0.327	0.374
		<i>E. tuberculatum</i> (n=2)	1.832	0.364	0.374
		<i>G. destructor</i> (n=2)	1.749	0.311	0.374
		<i>N. commutata</i> (n=2)	1.697	0.354	0.374
		<i>O. haematodus</i> (n=2)	12.626	0.093	0.374
		<i>P. clavata</i> (n=2)	1.852	0.342	0.374
	<i>D. armigerum</i> (n=2)	<i>E. tuberculatum</i> (n=2)	2.818	0.327	0.374
		<i>G. destructor</i> (n=2)	2.620	0.343	0.374
		<i>N. commutata</i> (n=2)	2.542	0.330	0.374
		<i>O. haematodus</i> (n=2)	47.111	0.111	0.374
		<i>P. clavata</i> (n=2)	2.808	0.335	0.374
	<i>E. tuberculatum</i> (n=2)	<i>G. destructor</i> (n=2)	1.046	0.346	0.374
		<i>N. commutata</i> (n=2)	1.057	0.317	0.374
		<i>O. haematodus</i> (n=2)	5.911	0.101	0.374
		<i>P. clavata</i> (n=2)	0.971	0.680	0.680
	<i>G. destructor</i> (n=2)	<i>N. commutata</i> (n=2)	1.020	0.366	0.374
		<i>O. haematodus</i> (n=2)	5.468	0.110	0.374
		<i>P. clavata</i> (n=2)	1.070	0.332	0.374
	<i>N. commutata</i> (n=2)	<i>O. haematodus</i> (n=2)	5.214	0.087	0.374
		<i>P. clavata</i> (n=2)	1.042	0.349	0.374
	<i>O. haematodus</i> (n=2)	<i>P. clavata</i> (n=2)	5.871	0.099	0.374
Diet	Carnivorous (n=7)	Herbivorous (n=4)	1.636	0.041	0.062
		Omnivorous (n=10)	1.694	0.082	0.062
	Herbivorous (n=4)	Omnivorous (n=10)	1.357	0.039	0.093
Nest	Ground (n=11)	Rotten (n=4)	1.669	0.027	0.027
		Tree (n=6)	1.458	0.001	0.003
	Rotten (n=4)	Tree (n=6)	2.192	0.006	0.009
Habitat	Rainforest (n=10)	City (n=11)	0.645	0.987	0.987

Weighted unifracs distance whole dataset

Factor	Group 1	Group 2	pseudo-F	p-value	q-value
Subfamily	Dolichoderinae (n=5)	Dorylinae (n=2)	3.169	0.043	0.100
		Ectatomminae (n=3)	1.332	0.152	0.193
		Formicinae (n=8)	1.865	0.063	0.126
		Myrmicinae (n=14)	0.933	0.465	0.482
		Paraponerinae (n=2)	2.987	0.051	0.110

		Ponerinae (n=9)	0.331	0.964	0.964
		Pseudomyrmicinae (n=3)	1.083	0.285	0.333
	Dorylinae (n=2)	Ectatomminae (n=3)	3.643	0.106	0.156
		Formicinae (n=8)	3.814	0.025	0.095
		Myrmicinae (n=14)	5.631	0.013	0.073
		Paraponerinae (n=2)	3.102	0.335	0.375
		Ponerinae (n=9)	3.043	0.027	0.095
		Pseudomyrmicinae (n=3)	3.447	0.099	0.156
		Ectatomminae (n=3)	Formicinae (n=8)	2.068	0.112
	Myrmicinae (n=14)		2.021	0.042	0.100
	Paraponerinae (n=2)		1.925	0.106	0.156
	Ponerinae (n=9)		1.659	0.117	0.156
	Pseudomyrmicinae (n=3)		3.033	0.105	0.156
	Formicinae (n=8)	Myrmicinae (n=14)	3.119	0.001	0.028
		Paraponerinae (n=2)	2.837	0.040	0.100
		Ponerinae (n=9)	2.954	0.002	0.028
		Pseudomyrmicinae (n=3)	2.833	0.013	0.073
	Myrmicinae (n=14)	Paraponerinae (n=2)	5.045	0.004	0.037
		Ponerinae (n=9)	1.387	0.174	0.212
		Pseudomyrmicinae (n=3)	2.679	0.026	0.095
	Paraponerinae (n=2)	Ponerinae (n=9)	3.184	0.031	0.096
		Pseudomyrmicinae (n=3)	2.743	0.105	0.156
	Ponerinae (n=9)	Pseudomyrmicinae (n=3)	1.057	0.459	0.482
Genus	<i>Atta</i> (n=3)	<i>Azteca</i> (n=3)	1.180	0.310	0.469
		<i>Camponotus</i> (n=6)	3.105	0.025	0.184
		<i>Cephalotes</i> (n=4)	3.374	0.020	0.184
		<i>Crematogaster</i> (n=2)	2.697	0.216	0.391
		<i>Daceton</i> (n=2)	2.745	0.214	0.391
		<i>Dolichoderus</i> (n=2)	1.347	0.309	0.469
		<i>Ectatomma</i> (n=3)	0.964	0.516	0.616
		<i>Gigantiops</i> (n=2)	1.277	0.196	0.389
		<i>NA</i> (n=3)	0.993	0.526	0.621
		<i>Neoponera</i> (n=2)	1.436	0.197	0.389
		<i>Odontomachus</i> (n=6)	4.582	0.011	0.184
		<i>Paraponera</i> (n=2)	2.110	0.100	0.250
		<i>Pseudomyrmex</i> (n=3)	2.497	0.108	0.250
		<i>Solenopsis</i> (n=2)	3.121	0.190	0.339
	<i>Azteca</i> (n=3)	<i>Camponotus</i> (n=6)	2.731	0.015	0.184
		<i>Cephalotes</i> (n=4)	3.047	0.023	0.184
		<i>Crematogaster</i> (n=2)	0.683	0.821	0.871
		<i>Daceton</i> (n=2)	1.183	0.305	0.469
		<i>Dolichoderus</i> (n=2)	0.798	0.599	0.670
		<i>Ectatomma</i> (n=3)	1.150	0.389	0.511
		<i>Gigantiops</i> (n=2)	1.453	0.210	0.391
		<i>NA</i> (n=3)	0.855	0.563	0.657
		<i>Neoponera</i> (n=2)	1.079	0.410	0.525
<i>Odontomachus</i> (n=6)		1.641	0.110	0.250	

	<i>Paraponera (n=2)</i>	2.601	0.124	0.271	
	<i>Pseudomyrmex (n=3)</i>	1.392	0.192	0.389	
	<i>Solenopsis (n=2)</i>	0.881	0.486	0.587	
<i>Camponotus (n=6)</i>	<i>Cephalotes (n=4)</i>	6.473	0.005	0.184	
	<i>Crematogaster (n=2)</i>	3.240	0.042	0.184	
	<i>Daceton (n=2)</i>	4.775	0.036	0.184	
	<i>Dolichoderus (n=2)</i>	2.664	0.073	0.247	
	<i>Ectatomma (n=3)</i>	4.580	0.012	0.184	
	<i>Gigantiops (n=2)</i>	3.897	0.039	0.184	
	<i>NA (n=3)</i>	3.098	0.009	0.184	
	<i>Neoponera (n=2)</i>	3.727	0.042	0.184	
	<i>Odontomachus (n=6)</i>	6.688	0.006	0.184	
	<i>Paraponera (n=2)</i>	5.508	0.035	0.084	
	<i>Pseudomyrmex (n=3)</i>	4.158	0.018	0.184	
	<i>Solenopsis (n=2)</i>	3.538	0.042	0.184	
	<i>Cephalotes (n=4)</i>	<i>Crematogaster (n=2)</i>	7.716	0.087	0.250
		<i>Daceton (n=2)</i>	8.368	0.060	0.233
		<i>Dolichoderus (n=2)</i>	3.867	0.071	0.247
<i>Ectatomma (n=3)</i>		3.635	0.027	0.184	
<i>Gigantiops (n=2)</i>		3.909	0.079	0.250	
<i>NA (n=3)</i>		2.477	0.035	0.184	
<i>Neoponera (n=2)</i>		2.702	0.060	0.233	
<i>Odontomachus (n=6)</i>		9.008	0.003	0.184	
<i>Paraponera (n=2)</i>		5.933	0.072	0.247	
<i>Pseudomyrmex (n=3)</i>		4.824	0.028	0.184	
<i>Solenopsis (n=2)</i>		8.435	0.052	0.218	
<i>Crematogaster (n=2)</i>	<i>Daceton (n=2)</i>	16.566	0.325	0.469	
	<i>Dolichoderus (n=2)</i>	0.731	1.000	1.000	
	<i>Ectatomma (n=3)</i>	2.928	0.103	0.250	
	<i>Gigantiops (n=2)</i>	1.974	0.294	0.469	
	<i>NA (n=3)</i>	1.049	0.597	0.669	
	<i>Neoponera (n=2)</i>	1.458	0.344	0.469	
	<i>Odontomachus (n=6)</i>	0.185	1.000	1.000	
	<i>Paraponera (n=2)</i>	4.154	0.329	0.469	
	<i>Pseudomyrmex (n=3)</i>	0.252	0.892	0.926	
	<i>Solenopsis (n=2)</i>	0.510	1.000	1.000	
<i>Daceton (n=2)</i>	<i>Dolichoderus (n=2)</i>	1.007	0.668	0.723	
	<i>Ectatomma (n=3)</i>	3.092	0.109	0.250	
	<i>Gigantiops (n=2)</i>	2.035	0.332	0.469	
	<i>NA (n=3)</i>	1.206	0.349	0.469	
	<i>Neoponera (n=2)</i>	1.306	0.324	0.469	
	<i>Odontomachus (n=6)</i>	3.769	0.082	0.250	
	<i>Paraponera (n=2)</i>	4.500	0.335	0.469	
	<i>Pseudomyrmex (n=3)</i>	1.649	0.416	0.526	
	<i>Solenopsis (n=2)</i>	16.626	0.333	0.469	
<i>Dolichoderus (n=2)</i>	<i>Ectatomma (n=3)</i>	1.631	0.105	0.250	
	<i>Gigantiops (n=2)</i>	1.241	0.330	0.469	

		NA (n=3)	0.690	0.585	0.668
		<i>Neoponera</i> (n=2)	0.862	0.686	0.735
		<i>Odontomachus</i> (n=6)	0.947	0.449	0.553
		<i>Paraponera</i> (n=2)	1.991	0.338	0.469
		<i>Pseudomyrmex</i> (n=3)	0.331	0.900	0.926
		<i>Solenopsis</i> (n=2)	0.946	0.665	0.723
	<i>Ectatomma</i> (n=3)	<i>Gigantiops</i> (n=2)	1.526	0.093	0.250
		NA (n=3)	1.158	0.408	0.525
		<i>Neoponera</i> (n=2)	1.378	0.204	0.389
		<i>Odontomachus</i> (n=6)	4.765	0.015	0.184
		<i>Paraponera</i> (n=2)	1.925	0.106	0.250
		<i>Pseudomyrmex</i> (n=3)	3.033	0.100	0.250
		<i>Solenopsis</i> (n=2)	3.315	0.097	0.250
	<i>Gigantiops</i> (n=2)	NA (n=3)	1.290	0.355	0.472
		<i>Neoponera</i> (n=2)	1.366	0.332	0.469
		<i>Odontomachus</i> (n=6)	4.719	0.035	0.184
		<i>Paraponera</i> (n=2)	1.372	0.325	0.469
		<i>Pseudomyrmex</i> (n=3)	2.156	0.112	0.250
		<i>Solenopsis</i> (n=2)	2.108	0.335	0.469
	NA (n=3)	<i>Neoponera</i> (n=2)	0.967	0.609	0.673
		<i>Odontomachus</i> (n=6)	2.469	0.039	0.184
		<i>Paraponera</i> (n=2)	1.392	0.203	0.389
		<i>Pseudomyrmex</i> (n=3)	1.703	0.085	0.250
		<i>Solenopsis</i> (n=2)	1.179	0.453	0.553
	<i>Neoponera</i> (n=2)	<i>Odontomachus</i> (n=6)	3.184	0.073	0.247
		<i>Paraponera</i> (n=2)	2.046	0.341	0.469
		<i>Pseudomyrmex</i> (n=3)	1.734	0.191	0.389
		<i>Solenopsis</i> (n=2)	13636.000	0.340	0.469
	<i>Odontomachus</i> (n=6)	<i>Paraponera</i> (n=2)	7.780	0.032	0.184
		<i>Pseudomyrmex</i> (n=3)	0.857	0.445	0.553
		<i>Solenopsis</i> (n=2)	0.488	0.573	0.661
	<i>Paraponera</i> (n=2)	<i>Pseudomyrmex</i> (n=3)	2.743	0.100	0.250
		<i>Solenopsis</i> (n=2)	4.314	0.343	0.469
	<i>Pseudomyrmex</i> (n=3)	<i>Solenopsis</i> (n=2)	0.317	0.891	0.926
Diet	Carnivorous (n=13)	Herbivorous (n=9)	1.418	0.046	0.157
		Omnivorous (n=24)	1.751	0.062	0.093
	Herbivorous (n=9)	Omnivorous (n=24)	1.797	0.059	0.093
Nest	Ground (n=21)	Rotten (n=7)	2.533	0.006	0.018
		Tree (n=18)	2.122	0.026	0.036
	Rotten (n=7)	Tree (n=18)	1.996	0.057	0.057
Habitat	Rainforest (n=24)	City (n=22)	0.945	0.502	0.502

Weighted unifracs distance forest/city dataset

Factor	Group 1	Group 2	pseudo-F	p-value	q-value
Subfamily	Dolichoderinae (n=2)	Ectatomminae (n=2)	1.418	0.332	0.385
		Formicinae (n=4)	1.291	0.382	0.409
		Myrmicinae (n=6)	1.728	0.105	0.225
		Paraponerinae (n=2)	2.300	0.334	0.385
		Ponerinae (n=5)	1.734	0.140	0.250
	Ectatomminae (n=2)	Formicinae (n=4)	1.174	0.529	0.529
		Myrmicinae (n=6)	1.374	0.244	0.366
		Paraponerinae (n=2)	1.507	0.323	0.385
		Ponerinae (n=5)	2.098	0.150	0.250
	Formicinae (n=4)	Myrmicinae (n=6)	2.034	0.035	0.153
		Paraponerinae (n=2)	2.164	0.051	0.153
		Ponerinae (n=5)	2.507	0.009	0.350
	Myrmicinae (n=6)	Paraponerinae (n=2)	4.061	0.040	0.153
		Ponerinae (n=5)	1.769	0.094	0.225
	Paraponerinae (n=2)	Ponerinae (n=5)	4.152	0.042	0.153
Genus	<i>A. cephalotes</i> (n=2)	<i>A. chartifex</i> (n=2)	0.778	0.681	0.684
		<i>C. atriceps</i> (n=2)	3.318	0.340	0.441
		<i>C. atratus</i> (n=2)	1.535	0.500	0.441
		<i>D. armigerum</i> (n=2)	2.340	0.332	0.441
		<i>E. tuberculatum</i> (n=2)	0.892	0.653	0.684
		<i>G. destructor</i> (n=2)	0.902	0.653	0.684
		<i>N. commutata</i> (n=2)	0.914	0.641	0.684
		<i>O. haematodus</i> (n=2)	5.029	0.100	0.441
		<i>P. clavata</i> (n=2)	1.798	0.337	0.441
	<i>A. chartifex</i> (n=2)	<i>C. atriceps</i> (n=2)	2.718	0.341	0.441
		<i>C. atratus</i> (n=2)	1.887	0.320	0.441
		<i>D. armigerum</i> (n=2)	1.997	0.339	0.441
		<i>E. tuberculatum</i> (n=2)	1.418	0.323	0.441
		<i>G. destructor</i> (n=2)	1.155	0.669	0.684
		<i>N. commutata</i> (n=2)	1.093	0.683	0.684
		<i>O. haematodus</i> (n=2)	3.524	0.101	0.441
		<i>P. clavata</i> (n=2)	2.300	0.320	0.441
	<i>C. atriceps</i> (n=2)	<i>C. atratus</i> (n=2)	8.345	0.355	0.441
		<i>D. armigerum</i> (n=2)	58.267	0.325	0.441
		<i>E. tuberculatum</i> (n=2)	7.600	0.317	0.441
		<i>G. destructor</i> (n=2)	2.037	0.334	0.441
		<i>N. commutata</i> (n=2)	2.995	0.345	0.441
		<i>O. haematodus</i> (n=2)	115.463	0.097	0.441
		<i>P. clavata</i> (n=2)	4.749	0.356	0.441
	<i>C. atratus</i> (n=2)	<i>D. armigerum</i> (n=2)	6.552	0.346	0.441
		<i>E. tuberculatum</i> (n=2)	1.979	0.338	0.441
		<i>G. destructor</i> (n=2)	1.497	0.324	0.441
<i>N. commutata</i> (n=2)		1.124	0.684	0.684	
<i>O. haematodus</i> (n=2)		12.411	0.103	0.441	
<i>P. clavata</i> (n=2)		2.706	0.340	0.441	

	<i>D. armigerum</i> (n=2)	<i>E. tuberculatum</i> (n=2)	5.169	0.331	0.441
		<i>G. destructor</i> (n=2)	1.920	0.363	0.441
		<i>N. commutata</i> (n=2)	1.397	0.322	0.441
		<i>O. haematodus</i> (n=2)	50.139	0.109	0.441
		<i>P. clavata</i> (n=2)	4.399	0.337	0.441
	<i>E. tuberculatum</i> (n=2)	<i>G. destructor</i> (n=2)	0.924	0.642	0.684
		<i>N. commutata</i> (n=2)	1.262	0.316	0.441
		<i>O. haematodus</i> (n=2)	10.109	0.099	0.441
		<i>P. clavata</i> (n=2)	1.507	0.357	0.441
	<i>G. destructor</i> (n=2)	<i>N. commutata</i> (n=2)	1.314	0.332	0.441
		<i>O. haematodus</i> (n=2)	3.469	0.095	0.441
		<i>P. clavata</i> (n=2)	1.325	0.313	0.441
	<i>N. commutata</i> (n=2)	<i>O. haematodus</i> (n=2)	3.024	0.101	0.441
		<i>P. clavata</i> (n=2)	2.109	0.326	0.441
	<i>O. haematodus</i> (n=2)	<i>P. clavata</i> (n=2)	7.357	0.097	0.441
Diet	Carnivorous (n=7)	Herbivorous (n=4)	1.049	0.360	0.519
		Omnivorous (n=10)	1.691	0.050	0.150
	Herbivorous (n=4)	Omnivorous (n=10)	0.935	0.519	0.519
Nest	Ground (n=11)	Rotten (n=4)	1.906	0.052	0.092
		Tree (n=6)	1.347	0.155	0.155
	Rotten (n=4)	Tree (n=6)	1.875	0.061	0.092
Habitat	Rainforest (n=10)	City (n=11)	0.515	0.965	0.965

Supplementary File 5. SIMPER statistics for each factor in the whole dataset.

SIMPER statistics according to the ant subfamily

15 most influential bacterial orders	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)							
				Dolichoderinae	Dorylinae	Ectatomminae	Formicinae	Myrmicinae	Paraponerinae	Ponerinae	Pseudomyrmicinae
Rickettsiales	25.01	28.98	28.98	1.97E+04	6.5	25.3	7.65E+03	8.66E+03	3	2.46E+04	1.91E+04
Rhizobiales	12.19	14.13	43.12	4.22E+03	669	3.33E+03	992	1.02E+04	1.53E+03	8.73E+03	15.7
Acetobacterales	7.158	8.295	51.41	6.96E+03	92	363	7.14E+03	1.79E+03	216	4.56E+01	1.32E+03
Enterobacteriales	6.956	8.062	59.47	40	701	323	1.47E+04	42.3	197	56.3	0
Lactobacillales	5.900	6.838	66.31	9.87E+03	1.10E+03	9.57+03	1.11E+03	1.21E+03	3.97E+02	629	17
Erysipelotrichales	3.907	4.528	70.84	0	1.51E+04	3.33	1	0.6	4.5	3.56E+03	0
Burkholderiales	3.535	4.097	74.94	199	1.85E+03	2.15+03	620	3.68E+03	1.31E+03	283	24
Xanthomonadales	3.225	3.738	78.67	171	508	1.32+03	358	3.58E+03	2.10E+03	336	0
Flavobacteriales	2.888	3.348	82.02	31.8	4.01E+03	186	49	619	65	3.60E+03	1.67
Pseudomonadales	2.602	3.016	85.04	244	1.37E+03	1.02E+03	812	324	7.41E+03	113	1
Corynebacteriales	2.088	2.420	87.46	2.30E+03	78	838	1.34E+03	358	263	50.7	2.33
Entomoplasmatales	1.088	1.261	88.72	3	2.83E+03	0	0	926	1	3	208
Sphingomonadales	0.990	1.148	89.87	73	187	1.41E+03	872	98.5	142	43.4	3.67
Opitutales	0.863	1.000	90.87	9.2	4.5	71	3.63	1.28E+03	24	13.1	0
Micrococcales	0.781	0.905	91.77	1.47E+03	5	255	92.9	145	142	45.3	9

SIMPER statistics according to the ant genus

15 most influential bacterial orders	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)																		
				<i>Anochetus</i>	<i>Atta</i>	<i>Azteca</i>	<i>Camponotus</i>	<i>Cephalotes</i>	<i>Crematogaster</i>	<i>Daceton</i>	<i>Dolichoderus</i>	<i>Eciton</i>	<i>Ectatomma</i>	<i>Gigantiops</i>	<i>Labidus</i>	<i>Neoponera</i>	<i>Nylanderia</i>	<i>Odontomachus</i>	<i>Paraponera</i>	<i>Pseudomyrmex</i>	<i>Solenopsis</i>	
Rickettsiales	25.04	28.93	28.93	0	14.7	9.84E+03	1.02E+04	1.22E+03	2.83E+04	24.5	3.44E+04	13	25.3	5	0	14	0	3.69E+04	3	1.91E+04	2.28E+04	
Rhizobiales	13.11	15.14	44.07	255	7.30E+03	94	157	1.21E+04	4.15E+03	3.67E+04	1.04E+04	1.24E+03	3.33E+03	3.50E+03	101	1.65E+04	140	7.56E+03	1.53E+03	15.7	118	
Acetobacteriales	7.169	8.282	52.36	115	1.16E+03	1.15E+04	6.71E+03	31.3	30.5	3.44E+03	83.5	159	363	8.46E+03	25	30.5	1.63E+04	39	216	1.32E+03	6.33	
Enterobacteriales	6.400	7.393	59.75	0	133	14.7	1.95E+04	35	0	0	78	1.40E+03	323	104	0	0	49	84.5	197	0	15	
Lactobacillales	5.587	6.454	66.20	44	4.04E+03	1.64E+04	1.43E+03	677	1.57E+03	5.5	283	604	9.57E+03	124	1.59E+03	793	5	671	397	17	34	
Burkholderiales	3.654	4.221	75.00	22	280	254	126	0	16.5	282	46.5	1.00E+03	1.32E+03	1.05E+03	12	1.42E+03	12	28.8	2.10E+03	0	23	
Xanthomonadales	3.538	4.087	79.08	0	0	0	0	2.03E+03	0	0	0	1.29E+04	3.33	4	1.73E+04	2.08E+03	7	4.65E+03	4.5	0	0.667	
Erysipelotrichales	3.959	4.573	70.78	69	528	264	96.5	1.32E+04	271	63.5	103	3.32E+03	2.15E+03	2.19E+03	385	940	63	99.8	1.31E+03	24	103	
Flavobacteriales	2.956	3.414	82.50	0	203	35	39.8	3	177	8.5	27	8.02E+03	186	76.5	0	1.62E+04	153	11.2	65	1.67	11.3	
Pseudomonadales	2.447	2.827	87.73	16	1.35E+03	400	888	0	126	114	9.5	1.7E+03	1.02E+03	585	986	163	167	113	7.41E+03	1	47.7	
Corynebacteriales	2.081	2.404	89.09	21	1.67E+03	3.60E+03	642	3.5	107	5	355	109	838	3.45E+03	47	116	101	33.8	263	2.33	6	
Entomoplasmatales	1.177	1.359	90.29	0	0	3.67	0	4.78E+03	0	6.49E+03	2	349	0	0	5.31E+03	0	0	4.5	1	208	0	
Opitutales	0.976	1.127	91.42	6	433	39.7	42.5	410	60.5	10.5	123	325	1.41E+03	3.36E+03	49	142	30	16.7	142	3.67	0	
Sphingomonadales	1.039	1.201	91.42	0	3	11	4.17	1.5	22.5	10.5	6.5	9	71	2	0	15.5	3	14.5	24	0	5	
Micrococcales	0.740	0.855	92.27	11	1.19E+02	2.39E+03	39.2	0	60	9.5	102	10	2.55E+02	254	0	125	36	24.5	1.42E+02	9	0	

SIMPER statistics according to the ant diet

15 most influential bacterial orders	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)		
				Carnivorous	Omnivorous	Herbivorous
Rickettsiales	24.27	27.91	27.91	1.48E+04	1.19E+04	8.19E+03
Rhizobiales	15.02	17.28	45.19	1.04E+04	1.18E+03	1.01E+04
Acetobacteriales	6.201	7.132	52.32	527	4.92E+03	420
Enterobacteriales	5.333	6.135	58.46	124	5.15E+03	77.3
Lactobacillales	4.978	5.726	64.18	577	3.90E+03	1.71E+03
Burkholderiales	4.611	5.304	69.49	600	574	6.05E+03
Xanthomonadales	4.384	5.042	74.53	587	335	5.91E+03
Erysipelotrichales	4.009	4.611	79.14	4.32E+06	1.17	0
Flavobacteriales	3.259	3.749	82.89	2.71E+03	69.6	975
Pseudomonadales	2.519	2.897	85.78	1.25E+03	492	455
Corynebacteriales	1.916	2.204	87.99	76.5	1.06E+03	637
Opitutales	1.344	1.546	89.53	13.1	14.7	2.19E+03
Entomoplasmatales	1.271	1.462	91.00	1.30E+03	27.6	2
Sphingomonadales	0.844	0.970	91.97	71.3	499	172
Micrococcales	0.633	0.728	92.70	48	385	244

SIMPER statistics according to the ant nesting mode

15 most influential bacterial orders	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)		
				Ground	Rotten	Tree
Rickettsiales	25.68	29.68	29.68	1.03E+04	2.08E+04	1.08E+04
Rhizobiales	12.76	14.74	44.42	3.64E+03	85.1	1.08E+04
Enterobacteriales	8.057	9.311	53.73	1.37E+03	1.30E+04	18.9
Acetobacteriales	7.861	9.085	62.81	1.88E+03	5.37E+03	2.55E+03
Lactobacillales	5.164	5.967	68.78	2.66E+03	413	2.90E+03
Burkholderiales	3.838	4.435	73.22	990	56.4	3.02E+03
Xanthomonadales	3.547	4.099	77.31	716	22.1	2.99E+03
Erysipelotrichales	3.113	3.598	80.91	1.57E+03	1	1.54+03
Flavobacteriales	2.488	2.875	83.79	1.92E+03	27.9	480
Pseudomonadales	2.219	2.564	86.35	1.46E+03	54	98.4
Corynebacteriales	1.951	2.254	88.61	868	57	655
Entomoplasmatales	1.062	1.227	89.83	258	0	807
Opitutales	1.053	1.217	91.05	18.5	3.71	1.07E+03
Sphingomonadales	0.854	0.988	92.04	613	21.4	25.8
Micrococcales	0.732	0.846	92.88	114	18.7	508

SIMPER statistics according to the ant habitat

15 most influential bacterial orders	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)	
				Rainforest	City
Rickettsiales	24.64	28.92	28.92	1.41E+04	9.94E+03
Rhizobiales	12.82	15.05	43.97	8.26E+03	3.21E+03
Acetobacteriales	7.124	8.363	52.33	1.15E+03	4.23E+03
Enterobacteriales	6.424	7.541	59.87	904	4.33E+03
Lactobacillales	5.658	6.642	66.51	706	4.21E+03
Burkholderiales	3.747	4.398	70.91	1.62E+03	1.65E+03
Erysipelotrichales	3.564	4.184	75.09	2.60E+03	1.17
Xanthomonadales	3.498	4.106	79.20	1.46E+03	1.51E+03
Flavobacteriales	2.863	3.361	82.56	401	1.80E+03
Pseudomonadales	2.450	2.876	85.44	1.14E+03	295
Corynebacteriales	2.019	2.370	87.81	679	652
Entomoplasmatales	1.115	1.308	89.12	713	135
Opitutales	1.004	1.178	90.29	282	562
Sphingomonadales	0.964	1.132	91.43	84.3	525
Micrococcales	0.721	0.847	92.27	385	111

Supplementary File 6. SIMPER statistics for each factor in the forest/city dataset.

SIMPER statistics according to the ant subfamily

15 most influential bacterial orders	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)					
				Dolichoderinae	Ectatomminae	Formicinae	Myrmicinae	Paraponerinae	Ponerinae
Rhizobiales	16.09	19.02	19.02	2.29E+03	5.71E+03	1.41E+03	1.69E+04	1.53E+03	1.08E+04
Rickettsiales	14.92	17.64	36.66	19	25	1.19E+04	23.2	3	2.15E+04
Acetobacterales	8.182	9.673	46.34	664	325	5.94E+03	7.30E+03	216	2.58E+03
Lactobacillales	5.737	6.782	53.12	890	311	789	8.19E+03	397	2.33E+03
Flavobacteriales	5.481	6.480	59.60	1.62E+04	4.12E+03	115	134	65	16
Burkholderiales	4.410	5.214	64.81	2.18E+03	5.81E+03	1.74E+03	3.33E+03	1.31E+03	163
Enterobacteriales	4.088	4.833	69.65	75.5	0	2.77E+03	37.8	197	3.21E+03
Pseudomonadales	3.513	4.153	73.80	633	21.5	771	680	7.41E+03	110
Corynebacteriales	3.058	3.615	77.41	3.31E+03	275	273	2.27E+03	263	176
Xanthomonadales	3.045	3.600	81.01	2.23E+03	5.73E+03	917	907	2.10E+03	141
Sphingomonadales	1.720	2.033	83.05	3.33E+03	115	1.06E+03	71.3	142	76.8
Opitutales	1.452	1.716	84.76	11.5	5.14E+03	53	181	24	16
Entomoplasmatales	1.344	1.588	86.35	0	0	0	2.31E+03	1	3.8
Micrococcales	1.254	1.483	87.33	190	562	202	1.34E+03	142	59.4
Erysipelotrichales	0.547	0.647	87.98	4	0	0	0	4.5	831

SIMPER statistics according to the ant genus

15 most influential bacterial orders	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)									
				<i>Atta</i>	<i>Azteca</i>	<i>Camponotus</i>	<i>Cephalotes</i>	<i>Daceton</i>	<i>Ectatomma</i>	<i>Gigantiops</i>	<i>Neoponera</i>	<i>Odontomachus</i>	<i>Paraponera</i>
Rhizobiales	16.92	19.62	19.62	1.07E+04	129	58	1.61E+04	3.67E+04	4.27E+03	3.50E+03	1.65E+04	308	1.53E+03
Rickettsiales	15.74	18.26	37.87	22	8	0	15	24.5	38	5	14	5.18E+04	3
Acetobacterales	8.652	10.030	47.90	1.46E+03	1.73E+04	9.93E+03	40.5	3.44E+03	500	8.46E+03	30.5	33	216
Lactobacillales	6.245	7.241	55.15	5.96E+03	2.44E+04	2	266	2.5	168	124	793	993	397
Flavobacteriales	4.938	5.726	60.87	304	52.5	4	4.06E+03	8.5	279	76.5	1.62E+04	13.7	65
Burkholderiales	4.575	5.305	66.18	693	388	44.5	1.42E+04	63.5	3.15E+03	2.19E+03	940	160	1.31E+03
Enterobacteriales	4.314	5.003	71.18	188	22	1.31E+04	0	0	39	104	0	169	197
Pseudomonadales	3.265	3.786	74.97	1.37E+03	597	67.5	0	1.14E+02	1.42E+03	585	163	128	7.41E+03
Corynebacteriales	3.119	3.617	78.58	1.70E+03	5.40E+03	134	0	5.00E+00	518	3.45E+03	116	39	263
Xanthomonadales	2.917	3.382	81.97	320	381	36	7.45E+03	282	1.89E+03	1.05E+03	1.42E+03	20.3	2.10E+03
Sphingomonadales	1.598	1.853	83.82	255	59.5	48	0	10.5	2.05E+03	3.36E+03	142	28	142
Entomoplasmatales	1.372	1.590	87.15	135	3.58E+03	30	814	9.5	342	254	125	19.3	142
Micrococcales	1.348	1.563	88.71	4.5	16.5	0	5.65E+03	10.5	107	2	15.5	21	24
Opitutales	1.503	1.742	85.56	0	5.5	0	4	6.94E+03	0	0	0	6.33	1
Erysipelotrichales	0.569	0.660	89.37	0	0	0	0	0	0	4	2.08E+03	0	4.5

SIMPER statistics according to the ant diet

15 most influential bacterial orders	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)		
				Carnivorous	Omnivorous	Herbivorous
Rhizobiales	16.84	19.96	19.96	1.24E+04	2.71E+03	1.16E+04
Rickettsiales	13.57	16.09	36.05	1.20E+04	5.97E+03	15.3
Acetobacteriales	8.946	10.61	46.65	1.73E+03	3.22E+03	1.04E+04
Lactobacillales	7.148	8.474	55.13	1.42E+03	695	1.22E+04
Burkholderiales	4.935	5.851	60.98	515	2.86E+03	4.70E+03
Flavobacteriales	4.748	5.629	66.61	89.2	5.13E+03	53.3
Enterobacteriales	3.71	4.398	71.01	1.85E+03	1.40E+03	11
Corynebacteriales	3.359	3.982	74.99	470	1.03E+03	2.70E+03
Pseudomonadales	2.98	3.532	78.52	2.02E+03	548	325
Xanthomonadales	2.839	3.366	81.89	563	2.45E+03	1.32E+03
Micrococcales	1.678	1.989	83.88	76.4	289	1.98E+03
Sphingomonadales	1.600	1.730	85.77	107	1.39E+03	33.5
Entomoplasmatales	1.459	1.530	89.03	1.20E+03	0	774
Opitutales	1.291	0.877	89.91	16.6	1.31E+03	266
Erysipelotrichales	0.494	0.586	90.50	463	1	0

SIMPER statistics according to the ant nesting mode

15 most influential bacterial orders	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)		
				Ground	Rotten	Tree
Rickettsiales	16.83	19.81	19.81	6.79E+03	2.02E+04	24.3
Rhizobiales	16.71	19.67	39.48	6.58E+03	5.27E+03	1.44E+04
Acetobacteriales	8.265	9.731	49.21	3.29E+03	2.12E+03	6.37E+03
Lactobacillales	7.172	8.444	57.66	616	3.61E+03	7.66E+03
Flavobacteriales	5.416	6.377	64.03	816	8.75	5.51E+03
Burkholderiales	4.438	5.225	69.26	2.03E+03	113	3.92E+03
Enterobacteriales	4.095	4.822	74.08	1.49E+03	2.78E+03	55.7
Corynebacteriales	2.968	3.494	77.57	1.20E+03	176	1.57E+03
Pseudomonadales	2.77	3.262	80.84	1.78E+03	41.3	691
Xanthomonadales	2.591	3.050	83.89	1.88E+03	150	1.45E+03
Sphingomonadales	1.638	1.929	85.82	453	61.5	1.16E+03
Entomoplasmatales	1.522	1.792	87.61	281	0	1.80E+03
Micrococcales	1.070	1.260	88.87	859	51.8	211
Opitutales	1.035	1.218	90.09	965	4.25	181
Erysipelotrichales	0.422	0.497	90.59	379	0	1.33

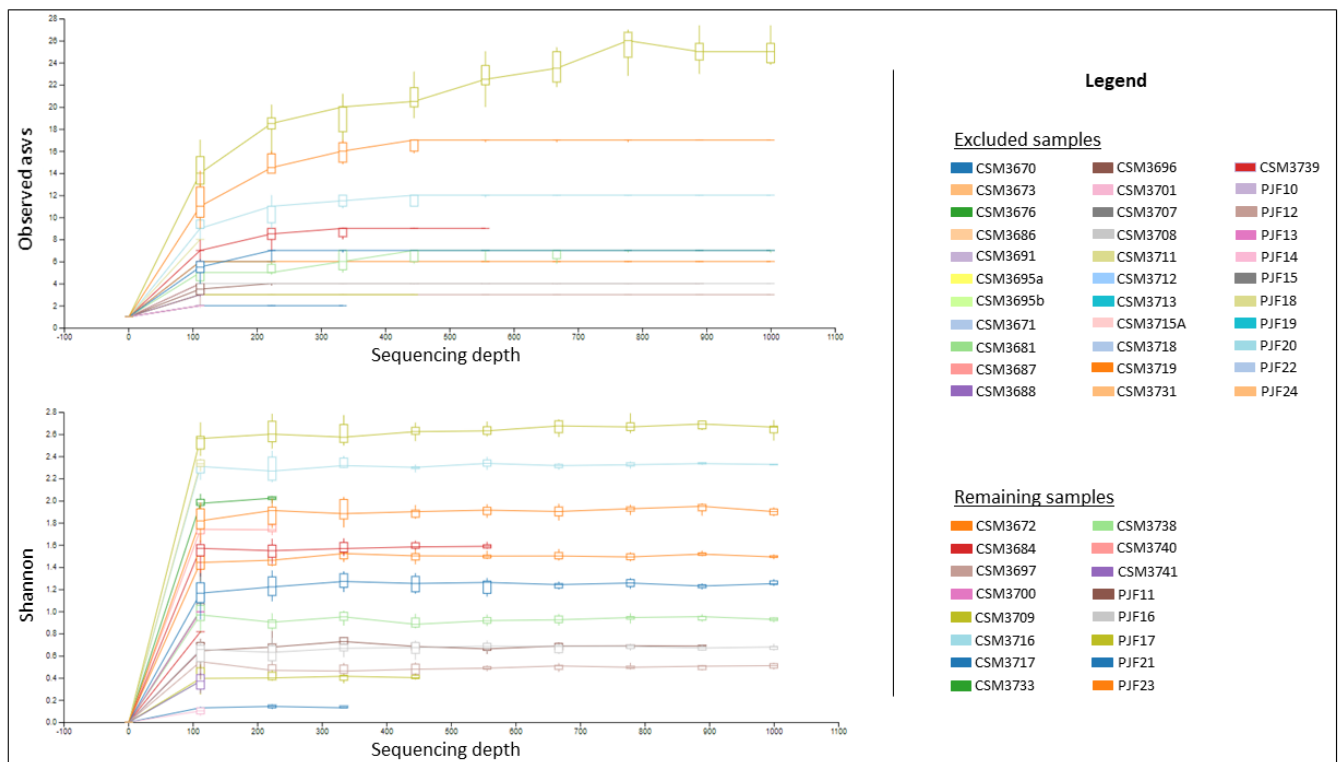
SIMPER statistics according to the ant habitat

15 most influential bacterial orders	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)	
				Rainforest	City
Rhizobiales	16.65	19.99	19.99	1.25E+04	4.99E+03
Rickettsiales	14.90	17.89	37.89	7.47E+03	7.35E+03
Acetobacteriales	8.168	9.806	47.69	2.66E+03	5.11E+03
Lactobacillales	6.059	7.274	54.97	623	5.54E+03
Flavobacteriales	4.841	5.812	60.78	127	3.71E+03
Burkholderiales	4.323	5.19	65.97	2.48E+03	1.96E+03
Enterobacteriales	3.970	4.766	70.73	1.12E+03	1.52E+03
Pseudomonadales	3.331	3.999	74.73	1.92E+03	422
Corynebacteriales	3.011	3.614	78.35	1.51E+03	750
Xanthomonadales	2.790	3.350	81.70	964	1.84E+03
Sphingomonadales	1.636	1.964	83.66	128	994
Entomoplasmatales	1.444	1.734	85.40	1.08E+03	281
Micrococcales	1.338	1.606	87.00	881	192
Opitutales	1.278	1.534	88.54	116	960
Erysipelotrichales	0.571	0.685	89.23	416	0.727

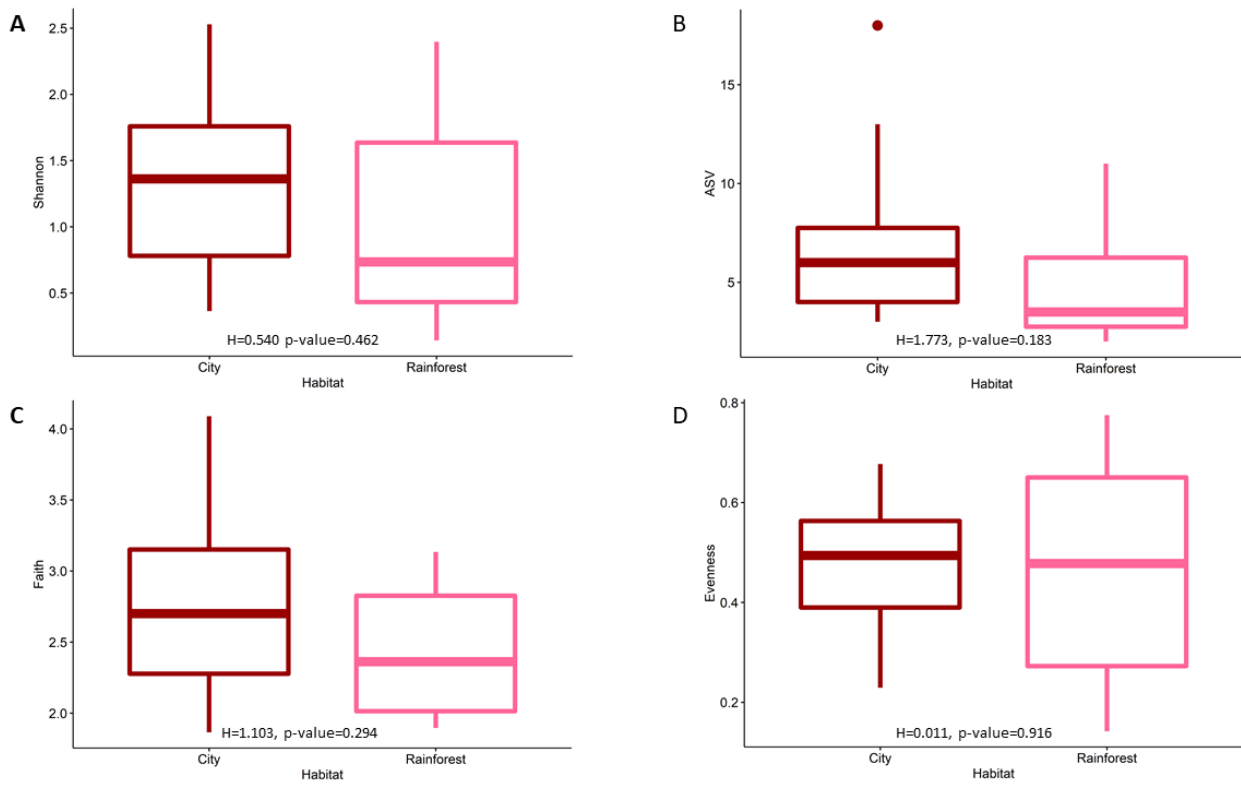
Supplementary File 7. SIMPER comparisons for each factor. Only the comparisons in which the bacterial relative abundance were statistically different (p-value < 0.05) are shown here. The statistical p-values were obtained with pairwise Kruskal-Wallis.

Factor	Comparison	SIMPER	Bacterial order	p-value	fdr_correction	Left.mean.abund	Left.stdev	Right.mean.abund	Right.stdev
Diet	Carnivorous_Herbivorous	0.053	Erysipelotrichales	0.011	0.157	0.099	0.195	0.000	0.000
	Carnivorous_Omnivorous	0.059	Erysipelotrichales	0.002	0.094	0.099	0.195	5.35E-05	1.18E-04
	Herbivorous_Omnivorous	0.157	Rhizobiales	0.007	0.157	0.236	0.332	0.036	0.061
	Carnivorous_Herbivorous	0.015	Entomoplasmatales	0.014	0.157	0.031	0.065	0.000	0.000
Nest	Rotten_Ground	0.075	Lactobacillales	0.033	0.297	0.008	0.021	0.077	0.168
	Rotten_Ground	0.031	Erysipelotrichales	0.044	0.297	5.50E-05	1.46E-04	0.052	0.153
	Tree_Ground	0.041	Erysipelotrichales	0.015	0.242	0.022	0.093	0.052	0.153
	Rotten_Ground	0.131	Rhizobiales	0.016	0.242	0.003	0.003	0.107	0.212
	Tree_Rotten	0.249	Rickettsiales	0.049	0.297	0.302	0.433	0.208	0.355
Habitat	City_Rainforest	0.061	Acetobacterales	0.016	0.265	0.114	0.207	0.000	0.000

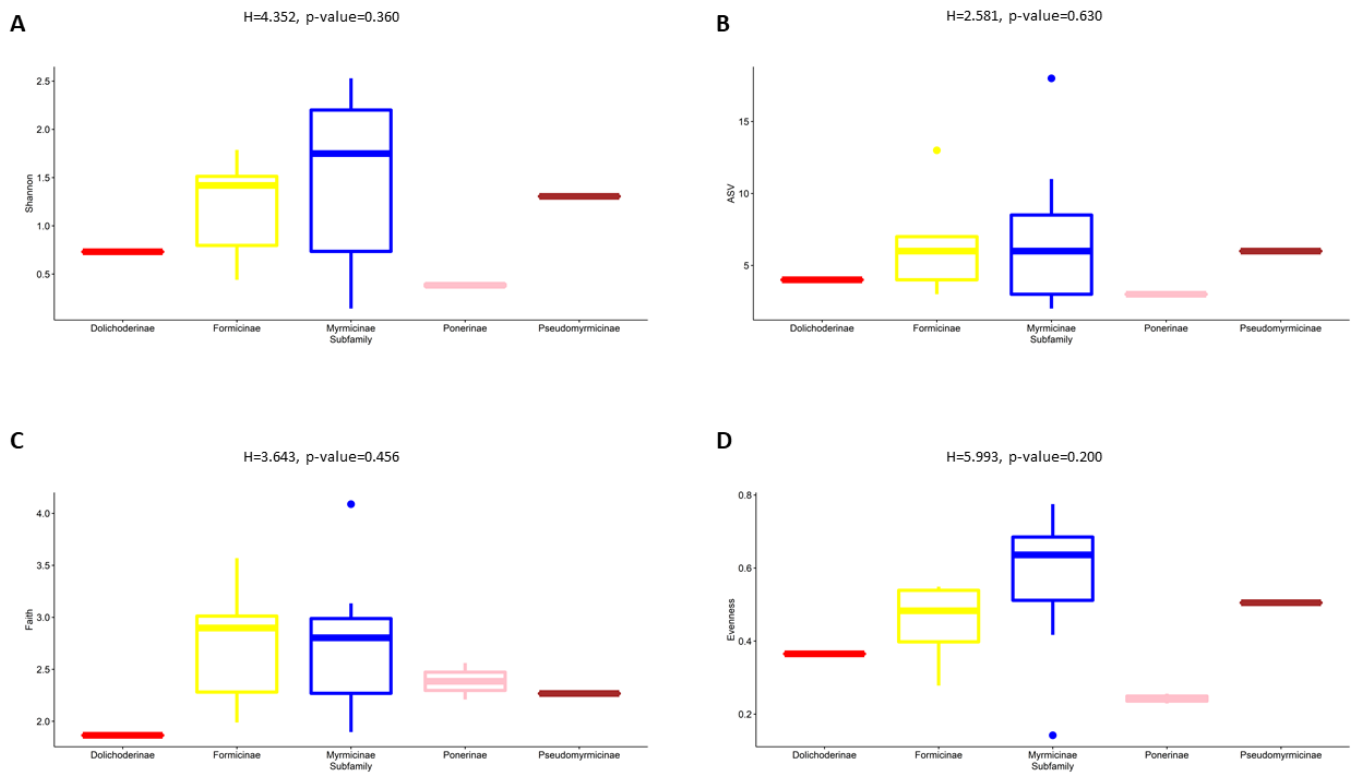
Annexe 2. Supporting Information de l'article Potential impact of host nesting mode in structuring their associated eukaryote communities



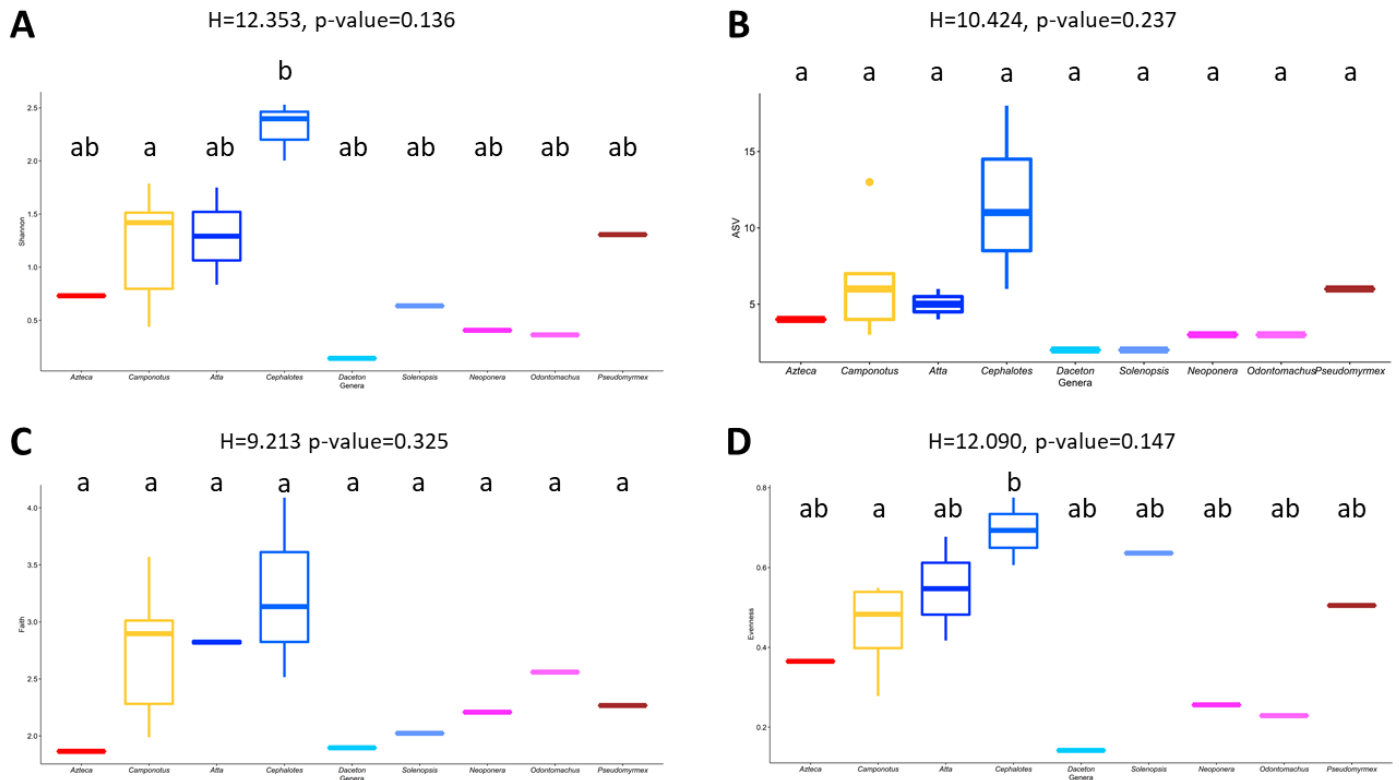
Supplementary Figure 12. Rarefaction curves used to estimate number of observed ASVs and richness in the Shannon index. The vertical axis shows the bacterial ASVs observed (A) or the Shannon index (B) and the number of sequences per sample is shown on the horizontal axis. At our chosen rarefaction threshold (199), all the remaining samples have reached a plateau for the Shannon index (B) and the number of observed ASVs does not increase a lot after the chosen threshold in most of the remaining samples (A), indicating that we seem to have recover the majority of the eukaryote diversity.



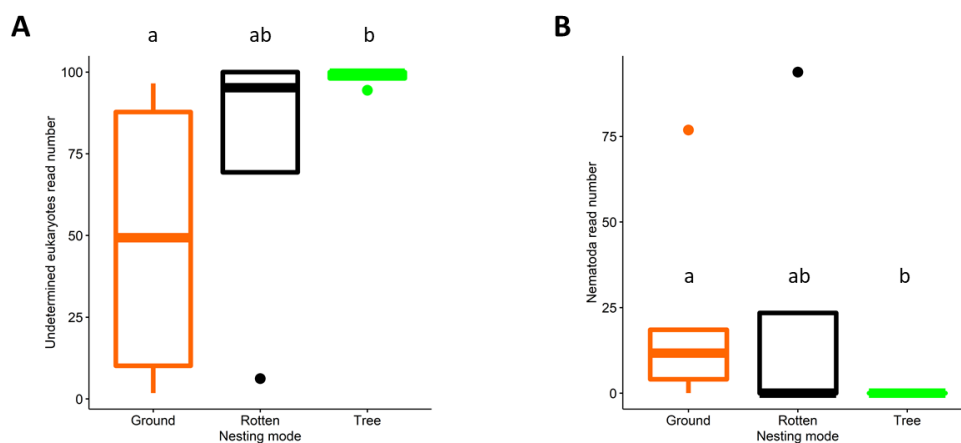
Supplementary Figure 13. Differences in eukaryote alpha diversity by ant habitat. (A) Mean (\pm standard error) Shannon diversity. (B) Mean (\pm standard error) ASV count. (C) Mean (\pm standard error) Faith's phylogenetic diversity. (D) mean (\pm standard error) Pielou's evenness. The statistical p-values were obtained with pairwise Kruskal-Wallis. P-values for this figure are found in Supplementary File 9.



Supplementary Figure 14. Differences in eukaryote alpha diversity by ant subfamily. (A) Mean (\pm standard error) Shannon diversity. (B) Mean (\pm standard error) ASV count. (C) Mean (\pm standard error) Faith's phylogenetic diversity. (D) mean (\pm standard error) Pielou's evenness. The statistical p-values were obtained with pairwise Kruskal-Wallis. P-values for this figure are found in Supplementary File 9.



Supplementary Figure 15. Differences in eukaryote alpha diversity by ant genus. (A) Mean (\pm standard error) Shannon diversity. (B) Mean (\pm standard error) ASV count. (C) Mean (\pm standard error) Faith's phylogenetic diversity. (D) mean (\pm standard error) Pielou's evenness. Different letters at the top of the figure illustrate genera with significant differences ($P < 0.05$) in this alpha diversity metric. The statistical p-values were obtained with pairwise Kruskal-Wallis. P-values for this figure are found in Supplementary File 9.

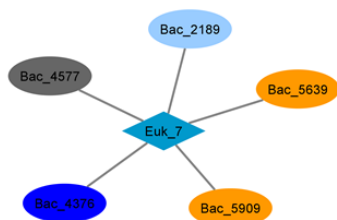


Supplementary Figure 16. SIMPER analysis using ASV abundance of different eukaryote phyla by nesting mode. (A) Mean (\pm standard error) undetermined eukaryotes read number. (B) Mean (\pm standard error) Nematoda read number. Different letters at the top of the figure illustrate significant differences ($P < 0.05$) determined by Kruskal-Wallis. P-values for this figure are found in Supplementary File 11.

Hosts :

Camponotus sp. 1: city ; omnivorous ; ground

Camponotus rapax: rainforest ; omnivorous ; ground



Hosts :

Atta cephalotes: city ; herbivorous ; ground

Atta cephalotes: rainforest ; herbivorous ; ground

Odontomachus haematodus: city ; carnivorous ; rotten



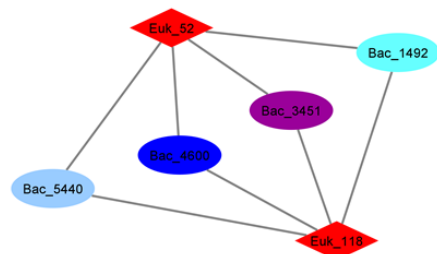
Network legend

Bacterial order

- Acetobacteriales
- Corynebacteriales
- Cytophagales
- Enterobacteriales
- Pseudomonadales
- Rhizobiales
- Undetermined

Eukaryote subphyla

- ◆ Eugregarinorida
- ◆ Nematoda



Hosts :

Camponotus sp. 2: city ; omnivorous ; rotten

Camponotus blandus: city ; omnivorous ; ground

Supplementary Figure 17. Correlation network between bacterial and eukaryote ASVs associated with Amazonian ants. The taxonomic assignments are indicated in the legend to the order level for the bacteria ASVs and to the subphyla level for the eukaryote ASVs. Statistics for the correlation network with a p-value lower than 0.05 are in Supplementary File 15

Supplementary File 8. List of collected samples.

Sample ID	Barcode sequence	Linker primer sequence	Barcode plate	Subfamily	Genus	Species	Sample type	Habitat mode	Habitat	Collection location	Diet	Nest	Nesting mode	Qubit 1ng/1uL	Well
CSM3670	AAGGCTG GGCAC	GTGTGYC AGCMGCC GCGGTAA	10	Ponerinae	Odontomachus	haematodus	Sample	City	City	Baduel	Carnivorous	Rotten log	Rotten	1.1113	B1
CSM3671	CGATAGG CCTTA	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Cephalotes	minutus	Sample	City	City	Baduel	Herbivorous	Twig	Tree	0.1729	B2
CSM3672	TAAGATGC AGTC	GTGTGYC AGCMGCC GCGGTAA	10	Formicinae	Camponotus	sp1	Sample	City	City	Baduel	Omnivorous	NA	Ground	1.4113	B3
CSM3673	TCCGCAAC CTGA	GTGTGYC AGCMGCC GCGGTAA	10	Paraponerinae	Paraponera	clavata	Sample	City	City	Baduel	Carnivorous	Ground	Ground	8.7923	B4
CSM3676	AGCCGAC TCTGT	GTGTGYC AGCMGCC GCGGTAA	10	Ectatomminae	Ectatomma	brunneum	Sample	City	City	Baduel	Omnivorous	Ground	Ground	1.32	B5
CSM3681	GTGTTCCC AGAA	GTGTGYC AGCMGCC GCGGTAA	10	Ponerinae	Odontomachus	hastatus	Sample	Rainforest	Rainforest	Camp	Carnivorous	Palm	Tree	2.2866	B6
CSM3684	CTCGAGC GTACTION	GTGTGYC AGCMGCC GCGGTAA	10	Formicinae	Camponotus	rapax	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Ground	Ground	1.5243	B7
CSM3686	GCTGTGAT TCGA	GTGTGYC AGCMGCC GCGGTAA	10	Ectatomminae	Ectatomma	tuberculatum	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Ground	Ground	1.3842	B8
CSM3687	CTGCCCTC GATG	GTGTGYC AGCMGCC GCGGTAA	10	Ponerinae	Odontomachus	scalptus	Sample	Rainforest	Rainforest	Petit plateau	Carnivorous	Ground	Ground	1.7947	B9
CSM3688	TTGGGTAC ACGT	GTGTGYC AGCMGCC GCGGTAA	10	Dolichoderinae	Dolichoderus	bispinosus	Sample	Rainforest	Rainforest	Petit plateau	Herbivorous	Palm	Tree	0.588	B10

CSM3691	GATATCGA TGAA	GTGTGYC AGCMGCC GCGGTAA	10	Dolichoderinae	Dolichoderus	attelaboides	Sample	Rainforest	Rainforest	Petit plateau	Herbivorous	Palm	Tree	1.5329	B11
CSM3695 a	AACGCGA AATTC	GTGTGYC AGCMGCC GCGGTAA	10	Formicinae	Camponotus	femoratus	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Ant garden	Tree	3.7	B12
CSM3695 b	GTGAGCA CGCAG	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Crematogaster	levior	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Ant garden	Tree	2.1	C1
CSM3696	CGATCGA ACACT	GTGTGYC AGCMGCC GCGGTAA	10	Dolichoderinae	Azteca	sp	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Tree	Tree	0.1	C2
CSM3697	TTCGCTAA CCTT	GTGTGYC AGCMGCC GCGGTAA	10	Formicinae	Camponotus	atriceps	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Rotten log	Rotten	2.1096	C3
CSM3700	GCAGGGT CGAAC	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Solenopsis	virulens	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	Rotten log	Rotten	0.2161	C4
CSM3701	TAGAGGC GTAGG	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Crematogaster	limata	Sample	Rainforest	Rainforest	Petit plateau	Omnivorous	NA	Ground	0.1	C5
CSM3707	CAGGTAAT AATG	GTGTGYC AGCMGCC GCGGTAA	10	Dorylinae	Eciton	rapax	Sample	Rainforest	Rainforest	Grand plateau	Carnivorous	Bivouac	Ground	0.1599	C6
CSM3708	GCCGAAC GCCGA	GTGTGYC AGCMGCC GCGGTAA	10	Paraponerinae	Paraponera	clavata	Sample	Rainforest	Rainforest	Grand plateau	Carnivorous	Ground	Ground	1.407	C7
CSM3709	GTCTAGTA TTTC	GTGTGYC AGCMGCC GCGGTAA	10	Ponerinae	Neoponera	commutata	Sample	Rainforest	Rainforest	Parare	Carnivorous	Ground	Ground	3.9063	C8
CSM3711	AAGCATTG AGAT	GTGTGYC AGCMGCC GCGGTAA	10	Pseudomyrmicinae	Pseudomyrmex	sp1	Sample	Rainforest	Rainforest	Terrace	Omnivorous	NA	Tree	0.0219	C9
CSM3712	CCGCTGAT GTCA	GTGTGYC AGCMGCC GCGGTAA	10	Ponerinae	Anochetus	emarginatus	Sample	Rainforest	Rainforest	Terrace	Carnivorous	Dead twig above ground	Tree	1.0661	C10

CSM3713	ACCCTCAG CCCA	GTGTGYC AGCMGCC GCGGTAA	10	Pseudomyrmicinae	Pseudomyrmex	gracilis	Sample	Rainforest	Rainforest camp	Camp	Omnivorous	Dead twig above ground	Tree	2.355	C11
CSM3715 A	CCGTCAA GATGT	GTGTGYC AGCMGCC GCGGTAA	10	Formicinae	Gigantiops	destructor	Sample	Rainforest	Rainforest	Camp	Omnivorous	Ground	Ground	0.1	D1
CSM3716	GGTGACT AGTTC	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Cephalotes	atratus	Sample	Rainforest	Rainforest camp	Camp	Herbivorous	Tree	Tree	1.3056	C12
CSM3717	GGAGTGC CCGAA	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Daceton	armigerum	Sample	Rainforest	Rainforest rocky savana	Mont	Omnivorous	Tree	Tree	0.8339	D3
CSM3718	CATGACG CCTCC	GTGTGYC AGCMGCC GCGGTAA	10	Dorylinae	Labidus	coecus	Sample	Rainforest	Rainforest rocky savana	Mont	Carnivorous	Bivouac	Ground	0.2935	D4
CSM3719	GTATTATG ACTT	GTGTGYC AGCMGCC GCGGTAA	10	Ponerinae	Odontomachus	haematodus	Sample	Rainforest	Rainforest rocky savana	Mont	Carnivorous	Ground	Ground	1.5994	D5
CSM3731	CATAATCT TGCT	GTGTGYC AGCMGCC GCGGTAA	10	Dolichoderinae	Azteca	chartifex	Sample	Rainforest	Rainforest camp	Camp	Omnivorous	Carton	Tree	0.0815	D6
CSM3733	TGCGTGG GTGGA	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Cephalotes	pavonii	Sample	Rainforest	Rainforest	Terrace	Herbivorous	Twig	Tree	0.3362	D7
CSM3738	ACTATATT AGCT	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Atta	cephalotes	Sample	Rainforest	Rainforest	Parare	Herbivorous	Ground	Ground	0.73	D8
CSM3739	TTCTGGTC TTGT	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Daceton	armigerum	Sample	City	Forest city	Loyola	Omnivorous	Tree	Tree	0.7073	D9
CSM3740	TCCAGGG CTATA	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Atta	cephalotes	Sample	City	Forest city	Montabo	Herbivorous	Ground	Ground	1.8164	D10
CSM3741	GACTATAA TGCC	GTGTGYC AGCMGCC GCGGTAA	10	Ponerinae	Odontomachus	haematodus	Sample	City	Forest city	Montabo	Carnivorous	Rotten log	Rotten	2.1916	D11

PJF10	CTATCATC CTCA	GTGTGYC AGCMGCC GCGGTAA	13	Myrmicinae	Solenopsis	geminata	Sample	City	City	Pasteur	Omnivorous	Ground	Ground	0.1	G9
PJF11	AATTGAAT TTAC	GTGTGYC AGCMGCC GCGGTAA	10	Dolichoderinae	Azteca	chartifex	Sample	City	City	Pasteur	Omnivorous	Carton	Tree	0.1	E1
PJF12	AGGAAAG CCAGA	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Atta	sexdens	Sample	City	City	Pasteur	Herbivorous	Ground	Ground	0.4812	E2
PJF13	ATGGCTGT CAGT	GTGTGYC AGCMGCC GCGGTAA	10	Formicinae	Camponotus	atriceps	Sample	City	City	Pasteur	Omnivorous	Rotten log	Rotten	1.3325	E3
PJF14	GAGGGTT GTCGG	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Nylanderia	sp	Sample	City	City	Pasteur	Omnivorous	Rotten log	Rotten	0.1	E4
PJF15	CCACGTAC GTAA	GTGTGYC AGCMGCC GCGGTAA	10	Ponerinae	Odontomachus	bori	Sample	City	City	Pasteur	Carnivorous	Ground	Ground	0.9291	E5
PJF16	AGTAAGTC GATA	GTGTGYC AGCMGCC GCGGTAA	10	Formicinae	Camponotus	sp2	Sample	City	Forest city	Montabo	Omnivorous	Rotten log	Rotten	5.0985	E6
PJF17	TTGGGCC ACATA	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Cephalotes	atratus	Sample	City	Forest city	Loyola	Herbivorous	Tree	Tree	4.1124	E7
PJF18	TCTATCTG GCTT	GTGTGYC AGCMGCC GCGGTAA	10	Formicinae	Gigantiops	destructor	Sample	City	Forest city	Loyola	Omnivorous	Ground	Ground	1.1611	E8
PJF19	CCGTGGA CCAGG	GTGTGYC AGCMGCC GCGGTAA	10	Ponerinae	Neoponera	commutata	Sample	City	Forest city	Montabo	Carnivorous	Ground	Ground	2.8585	E12
PJF20	GTCCTAT TATC	GTGTGYC AGCMGCC GCGGTAA	10	Ectatomminae	Ectatomma	tuberculatum	Sample	City	Forest city	Loyola	Omnivorous	Ground	Ground	0.1235	F1
PJF21	GCCTGCA GTA	GTGTGYC AGCMGCC GCGGTAA	10	Pseudomyrmicinae	Pseudomyrmex	sp2	Sample	City	City	Pasteur	Omnivorous	Twig	Tree	4.71	E10

PJF22	ATACCTAA ATAT	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Solenopsis	sp	Sample	City	City	Pasteur	Omnivorous	Twig	Tree	0.887	E9
PJF23	ATGGTTCA CCCG	GTGTGYC AGCMGCC GCGGTAA	10	Formicinae	Camponotus	blandus	Sample	City	City	Pasteur	Omnivorous	Ground	Ground	28	E11
PJF24	GCGACGC GGCAT	GTGTGYC AGCMGCC GCGGTAA	10	Myrmicinae	Crematogaster	sp2	Sample	City	City	Pasteur	Omnivorous	Twig	Tree	0.1	F2
negative1	CAGAGCT AATTG	GTGTGYC AGCMGCC GCGGTAA	18	Control	Control	Control	Control	Control	Control	Control	Control	Control	Control	0.101	D1
negative2	ATATATAG TATC	GTGTGYC AGCMGCC GCGGTAA	15	Control	Control	Control	Control	Control	Control	Control	Control	Control	Control	0.1017	G5

Supplementary File 9. Statistic calculations of alpha diversity metrics for each factor. The four tested alpha diversity metrics are Shannon diversity, Pielou's evenness, Faith's PD and ASV richness. The cells colored in red in each table represent p-values < 0.05. The statistical p-values were obtained with pairwise Kruskal-Wallis.

Shannon diversity

Factor	Group 1	Group 2	H	p-value	q-value
Subfamily	Dolichoderinae (n=1)	Formicinae (n=5)	0.771	0.380	0.633
		Myrmicinae (n=7)	0.429	0.513	0.641
		Ponerinae (n=2)	1.500	0.221	0.552
		Pseudomyrmicinae (n=1)	1.000	0.317	0.633
	Formicinae (n=5)	Myrmicinae (n=7)	0.534	0.465	0.641
		Ponerinae (n=2)	3.750	0.053	0.528
		Pseudomyrmicinae (n=1)	0.086	0.770	0.827
	Myrmicinae (n=7)	Ponerinae (n=2)	2.143	0.143	0.552
		Pseudomyrmicinae (n=1)	0.048	0.827	0.827
	Ponerinae (n=2)	Pseudomyrmicinae (n=1)	1.500	0.221	0.552
Genus	<i>Atta</i> (n=2)	<i>Azteca</i> (n=1)	1.500	0.221	0.368
		<i>Camponotus</i> (n=5)	0.150	0.699	0.740
		<i>Cephalotes</i> (n=3)	3.000	0.083	0.368
		<i>Daceton</i> (n=1)	1.500	0.221	0.368
		<i>Neoponera</i> (n=1)	1.500	0.221	0.368
		<i>Odontomachus</i> (n=1)	1.500	0.221	0.368
		<i>Pseudomyrmex</i> (n=1)	0.000	1.000	1.000
		<i>Solenopsis</i> (n=1)	1.500	0.221	0.368
	<i>Azteca</i> (n=1)	<i>Camponotus</i> (n=5)	0.771	0.380	0.414
		<i>Cephalotes</i> (n=3)	1.800	0.180	0.368
		<i>Daceton</i> (n=1)	1.000	0.317	0.368
		<i>Neoponera</i> (n=1)	1.000	0.317	0.368
		<i>Odontomachus</i> (n=1)	1.000	0.317	0.368
		<i>Pseudomyrmex</i> (n=1)	1.000	0.317	0.368
		<i>Solenopsis</i> (n=1)	1.000	0.317	0.368
	<i>Camponotus</i> (n=5)	<i>Cephalotes</i> (n=3)	5.000	0.025	0.368
		<i>Daceton</i> (n=1)	2.143	0.143	0.368
		<i>Neoponera</i> (n=1)	2.143	0.143	0.368
		<i>Odontomachus</i> (n=1)	2.143	0.143	0.368
		<i>Pseudomyrmex</i> (n=1)	0.086	0.770	0.792
		<i>Solenopsis</i> (n=1)	0.771	0.380	0.414
	<i>Cephalotes</i> (n=3)	<i>Daceton</i> (n=1)	1.800	0.180	0.368
		<i>Neoponera</i> (n=1)	1.800	0.180	0.368
		<i>Odontomachus</i> (n=1)	1.800	0.180	0.368
		<i>Pseudomyrmex</i> (n=1)	1.800	0.180	0.368
		<i>Solenopsis</i> (n=1)	1.800	0.180	0.368
	<i>Daceton</i> (n=1)	<i>Neoponera</i> (n=1)	1.000	0.317	0.368
		<i>Odontomachus</i> (n=1)	1.000	0.317	0.368

		<i>Pseudomyrmex</i> (n=1)	1.000	0.317	0.368
		<i>Solenopsis</i> (n=1)	1.000	0.317	0.368
	<i>Neoponera</i> (n=1)	<i>Odontomachus</i> (n=1)	1.000	0.317	0.368
		<i>Pseudomyrmex</i> (n=1)	1.000	0.317	0.368
		<i>Solenopsis</i> (n=1)	1.000	0.317	0.368
	<i>Odontomachus</i> (n=1)	<i>Pseudomyrmex</i> (n=1)	1.000	0.317	0.368
		<i>Solenopsis</i> (n=1)	1.000	0.317	0.368
	<i>Pseudomyrmex</i> (n=1)	<i>Solenopsis</i> (n=1)	1.000	0.317	0.368
Nest	Ground (n=6)	Rotten (n=4)	3.682	0.045	0.165
		Tree (n=6)	0.231	0.631	0.631
	Rotten (n=4)	Tree (n=6)	2.227	0.136	0.203
Habitat	Rainforest (n=8)	City (n=8)	0.540	0.462	0.462

Pielou's evenness

Factor	Group 1	Group 2	H	p-value	q-value
Subfamily	<i>Dolichoderinae</i> (n=1)	Formicinae (n=5)	0.771	0.380	0.475
		Myrmicinae (n=7)	1.190	0.275	0.453
		Ponerinae (n=2)	1.500	0.221	0.441
		<i>Pseudomyrmicinae</i> (n=1)	1.000	0.317	0.453
	<i>Formicinae</i> (n=5)	Myrmicinae (n=7)	2.380	0.123	0.441
		Ponerinae (n=2)	3.750	0.053	0.441
		<i>Pseudomyrmicinae</i> (n=1)	0.086	0.770	0.770
	<i>Myrmicinae</i> (n=7)	Ponerinae (n=2)	2.143	0.143	0.441
		<i>Pseudomyrmicinae</i> (n=1)	0.429	0.513	0.570
	<i>Ponerinae</i> (n=2)	<i>Pseudomyrmicinae</i> (n=1)	1.500	0.221	0.441
Genus	<i>Atta</i> (n=2)	<i>Azteca</i> (n=1)	1.500	0.221	0.381
		<i>Camponotus</i> (n=5)	0.600	0.439	0.493
		<i>Cephalotes</i> (n=3)	1.333	0.248	0.381
		<i>Daceton</i> (n=1)	1.500	0.221	0.381
		<i>Neoponera</i> (n=1)	1.500	0.221	0.381
		<i>Odontomachus</i> (n=1)	1.500	0.221	0.381
		<i>Pseudomyrmex</i> (n=1)	0.000	1.000	1.000
		<i>Solenopsis</i> (n=1)	0.000	1.000	1.000
	<i>Azteca</i> (n=1)	<i>Camponotus</i> (n=5)	0.771	0.380	0.441
		<i>Cephalotes</i> (n=3)	1.800	0.180	0.381
		<i>Daceton</i> (n=1)	1.000	0.317	0.381
		<i>Neoponera</i> (n=1)	1.000	0.317	0.381
		<i>Odontomachus</i> (n=1)	1.000	0.317	0.381
		<i>Pseudomyrmex</i> (n=1)	1.000	0.317	0.381
		<i>Solenopsis</i> (n=1)	1.000	0.317	0.381
	<i>Camponotus</i> (n=5)	<i>Cephalotes</i> (n=3)	5.000	0.025	0.381
		<i>Daceton</i> (n=1)	2.143	0.143	0.381
<i>Neoponera</i> (n=1)		2.143	0.143	0.381	
<i>Odontomachus</i> (n=1)		2.143	0.143	0.381	

		<i>Pseudomyrmex</i> (n=1)	0.086	0.770	0.815
		<i>Solenopsis</i> (n=1)	2.143	0.143	0.381
	<i>Cephalotes</i> (n=3)	<i>Daceton</i> (n=1)	1.800	0.180	0.381
		<i>Neoponera</i> (n=1)	1.800	0.180	0.381
		<i>Odontomachus</i> (n=1)	1.800	0.180	0.381
		<i>Pseudomyrmex</i> (n=1)	1.800	0.180	0.381
		<i>Solenopsis</i> (n=1)	0.200	0.655	0.714
		<i>Daceton</i> (n=1)	<i>Neoponera</i> (n=1)	1.000	0.317
	<i>Odontomachus</i> (n=1)		1.000	0.317	0.381
	<i>Pseudomyrmex</i> (n=1)		1.000	0.317	0.381
	<i>Solenopsis</i> (n=1)		1.000	0.317	0.381
	<i>Neoponera</i> (n=1)	<i>Odontomachus</i> (n=1)	1.000	0.317	0.381
		<i>Pseudomyrmex</i> (n=1)	1.000	0.317	0.381
		<i>Solenopsis</i> (n=1)	1.000	0.317	0.381
	<i>Odontomachus</i> (n=1)	<i>Pseudomyrmex</i> (n=1)	1.000	0.317	0.381
		<i>Solenopsis</i> (n=1)	1.000	0.317	0.381
	<i>Pseudomyrmex</i> (n=1)	<i>Solenopsis</i> (n=1)	1.000	0.317	0.381
Nest	Ground (n=6)	Rotten (n=4)	1.136	0.286	0.591
		Tree (n=6)	0.231	0.631	0.631
	Rotten (n=4)	Tree (n=6)	0.727	0.394	0.591
Habitat	Rainforest (n=8)	City (n=8)	0.011	0.916	0.916

Faith's PD

Factor	Group 1	Group 2	H	p-value	q-value
Subfamily	Dolichoderinae (n=1)	Formicinae (n=5)	2.143	0.143	0.698
		Myrmicinae (n=7)	2.333	0.127	0.698
		Ponerinae (n=2)	1.500	0.221	0.698
		Pseudomyrmicinae (n=1)	1.000	0.317	0.698
	Formicinae (n=5)	Myrmicinae (n=7)	0.059	0.808	0.897
		Ponerinae (n=2)	0.600	0.439	0.698
		Pseudomyrmicinae (n=1)	0.770	0.380	0.698
	Myrmicinae (n=7)	Ponerinae (n=2)	0.343	0.558	0.698
		Pseudomyrmicinae (n=1)	0.429	0.513	0.698
	Ponerinae (n=2)	Pseudomyrmicinae (n=1)	0.000	1.000	1.000
Genus	<i>Atta</i> (n=2)	<i>Azteca</i> (n=1)	1.500	0.221	0.394
		<i>Camponotus</i> (n=5)	0.150	0.699	0.718
		<i>Cephalotes</i> (n=3)	0.333	0.564	0.615
		<i>Daceton</i> (n=1)	1.500	0.221	0.394
		<i>Neoponera</i> (n=1)	1.500	0.221	0.394
		<i>Odontomachus</i> (n=1)	1.500	0.221	0.394
		<i>Pseudomyrmex</i> (n=1)	1.500	0.221	0.394
		<i>Solenopsis</i> (n=1)	1.500	0.221	0.394
	<i>Azteca</i> (n=1)	<i>Camponotus</i> (n=5)	2.143	0.143	0.394
		<i>Cephalotes</i> (n=3)	1.800	0.180	0.394

		<i>Daceton (n=1)</i>	1.000	0.317	0.394
		<i>Neoponera (n=1)</i>	1.000	0.317	0.394
		<i>Odontomachus (n=1)</i>	1.000	0.317	0.394
		<i>Pseudomyrmex (n=1)</i>	1.000	0.317	0.394
		<i>Solenopsis (n=1)</i>	1.000	0.317	0.394
	<i>Camponotus (n=5)</i>	<i>Cephalotes (n=3)</i>	1.089	0.143	0.394
		<i>Daceton (n=1)</i>	2.143	0.180	0.394
		<i>Neoponera (n=1)</i>	0.771	0.317	0.427
		<i>Odontomachus (n=1)</i>	0.086	0.770	0.770
		<i>Pseudomyrmex (n=1)</i>	0.771	0.380	0.427
		<i>Solenopsis (n=1)</i>	0.771	0.380	0.427
	<i>Cephalotes (n=3)</i>	<i>Daceton (n=1)</i>	1.800	0.180	0.394
		<i>Neoponera (n=1)</i>	1.800	0.180	0.394
		<i>Odontomachus (n=1)</i>	0.200	0.655	0.693
		<i>Pseudomyrmex (n=1)</i>	1.800	0.180	0.394
		<i>Solenopsis (n=1)</i>	1.800	0.180	0.394
	<i>Daceton (n=1)</i>	<i>Neoponera (n=1)</i>	1.000	0.317	0.394
		<i>Odontomachus (n=1)</i>	1.000	0.317	0.394
		<i>Pseudomyrmex (n=1)</i>	1.000	0.317	0.394
		<i>Solenopsis (n=1)</i>	1.000	0.317	0.394
	<i>Neoponera (n=1)</i>	<i>Odontomachus (n=1)</i>	1.000	0.317	0.394
		<i>Pseudomyrmex (n=1)</i>	1.000	0.317	0.394
		<i>Solenopsis (n=1)</i>	1.000	0.317	0.394
	<i>Odontomachus (n=1)</i>	<i>Pseudomyrmex (n=1)</i>	1.000	0.317	0.394
		<i>Solenopsis (n=1)</i>	1.000	0.317	0.394
	<i>Pseudomyrmex (n=1)</i>	<i>Solenopsis (n=1)</i>	1.000	0.317	0.394
Nest	Ground (n=6)	Rotten (n=4)	4.545	0.033	0.099
		Tree (n=6)	0.641	0.423	0.635
	Rotten (n=4)	Tree (n=6)	0.045	0.831	0.831
Habitat	Rainforest (n=8)	City (n=8)	1.103	0.294	0.294

ASV richness

Factor	Group 1	Group 2	H	p-value	q-value
Subfamily	Dolichoderinae (n=1)	Formicinae (n=5)	0.353	0.552	0.913
		Myrmicinae (n=7)	0.198	0.657	0.913
		Ponerinae (n=2)	2.000	0.157	0.524
		Pseudomyrmicinae (n=1)	1.000	0.317	0.747
	Formicinae (n=5)	Myrmicinae (n=7)	0.108	0.743	0.913
		Ponerinae (n=2)	2.585	0.108	0.524
		Pseudomyrmicinae (n=1)	0.000	1.000	1.000
	Myrmicinae (n=7)	Ponerinae (n=2)	0.791	0.374	0.747
		Pseudomyrmicinae (n=1)	0.051	0.822	0.913
	Ponerinae (n=2)	Pseudomyrmicinae (n=1)	2.000	0.157	0.524
Genus	<i>Atta (n=2)</i>	<i>Azteca (n=1)</i>	0.500	0.480	0.526

		<i>Camponotus (n=5)</i>	0.156	0.693	0.714
		<i>Cephalotes (n=3)</i>	2.193	0.139	0.385
		<i>Daceton (n=1)</i>	1.500	0.221	0.385
		<i>Neoponera (n=1)</i>	1.500	0.221	0.385
		<i>Odontomachus (n=1)</i>	1.500	0.221	0.385
		<i>Pseudomyrmex (n=1)</i>	0.500	0.480	0.526
		<i>Solenopsis (n=1)</i>	1.500	0.221	0.385
	<i>Azteca (n=1)</i>	<i>Camponotus (n=5)</i>	0.353	0.552	0.587
		<i>Cephalotes (n=3)</i>	1.800	0.180	0.385
		<i>Daceton (n=1)</i>	1.000	0.317	0.385
		<i>Neoponera (n=1)</i>	1.000	0.317	0.385
		<i>Odontomachus (n=1)</i>	1.000	0.317	0.385
		<i>Pseudomyrmex (n=1)</i>	1.000	0.317	0.385
		<i>Solenopsis (n=1)</i>	1.000	0.317	0.385
	<i>Camponotus (n=5)</i>	<i>Cephalotes (n=3)</i>	1.439	0.230	0.385
		<i>Daceton (n=1)</i>	2.143	0.143	0.385
		<i>Neoponera (n=1)</i>	1.412	0.235	0.385
		<i>Odontomachus (n=1)</i>	1.412	0.235	0.385
		<i>Pseudomyrmex (n=1)</i>	0.000	1.000	1.000
		<i>Solenopsis (n=1)</i>	2.143	0.143	0.385
	<i>Cephalotes (n=3)</i>	<i>Daceton (n=1)</i>	1.800	0.180	0.385
		<i>Neoponera (n=1)</i>	1.800	0.180	0.385
		<i>Odontomachus (n=1)</i>	1.800	0.180	0.385
		<i>Pseudomyrmex (n=1)</i>	0.889	0.346	0.405
		<i>Solenopsis (n=1)</i>	1.800	0.180	0.385
	<i>Daceton (n=1)</i>	<i>Neoponera (n=1)</i>	1.000	0.317	0.385
		<i>Odontomachus (n=1)</i>	1.000	0.317	0.385
		<i>Pseudomyrmex (n=1)</i>	1.000	0.317	0.385
		<i>Solenopsis (n=1)</i>	NA	NA	NA
	<i>Neoponera (n=1)</i>	<i>Odontomachus (n=1)</i>	NA	NA	NA
		<i>Pseudomyrmex (n=1)</i>	1.000	0.317	0.385
		<i>Solenopsis (n=1)</i>	1.000	0.317	0.385
	<i>Odontomachus (n=1)</i>	<i>Pseudomyrmex (n=1)</i>	1.000	0.317	0.385
		<i>Solenopsis (n=1)</i>	1.000	0.317	0.385
	<i>Pseudomyrmex (n=1)</i>	<i>Solenopsis (n=1)</i>	1.000	0.317	0.385
Nest	Ground (n=6)	Rotten (n=4)	4.257	0.039	0.002
		Tree (n=6)	0.007	0.935	0.410
	Rotten (n=4)	Tree (n=6)	2.981	0.084	0.039
Habitat	Rainforest (n=8)	City (n=8)	1.773	0.183	0.183

Supplementary File 10. Statistic calculations of beta diversity metrics for each factor. The two tested beta diversity metrics are Bray-Curtis and Weighted unifrac. The cells colored in red in each table represent p-values < 0.05. The statistical p-values were obtained with PERMANOVAs.

Bray-Curtis distance

Factor	Group 1	Group 2	pseudo-F	p-value	q-value
Subfamily	Formicinae (n=5)	Myrmicinae (n=7)	1.135	0.067	0.402
		NA (n=2)	1.061	0.209	0.418
		Ponerinae (n=2)	1.146	0.153	0.418
	Myrmicinae (n=7)	NA (n=2)	1.035	0.599	0.719
		Ponerinae (n=2)	0.823	0.727	0.727
	NA (n=2)	Ponerinae (n=2)	1.092	0.317	0.476
Genus	<i>Atta</i> (n=2)	<i>Camponotus</i> (n=5)	1.396	0.064	0.138
		<i>Cephalotes</i> (n=3)	1.345	0.112	0.224
		NA (n=6)	0.989	0.441	0.487
	<i>Camponotus</i> (n=5)	<i>Cephalotes</i> (n=3)	1.068	0.247	0.371
		NA (n=6)	1.099	0.052	0.138
	<i>Cephalotes</i> (n=3)	NA (n=6)	1.014	0.487	0.487
Nest	Ground (n=6)	Rotten (n=4)	0.814	0.809	1.000
		Tree (n=6)	1.121	0.048	0.252
	Rotten (n=4)	Tree (n=6)	1.000	1.000	1.000
Habitat	Rainforest (n=8)	City (n=8)	1.042	0.340	0.340

Weighted unifrac distance

Factor	Group 1	Group 2	pseudo-F	p-value	q-value
Subfamily	<i>Formicinae</i> (n=5)	<i>Myrmicinae</i> (n=7)	0.691	0.607	0.705
		NA (n=2)	0.606	0.458	0.705
		<i>Ponerinae</i> (n=2)	0.996	0.421	0.705
	Myrmicinae (n=7)	NA (n=2)	0.578	0.705	0.705
		<i>Ponerinae</i> (n=2)	0.535	0.677	0.705
	NA (n=2)	<i>Ponerinae</i> (n=2)	1.300	0.334	0.705
Genus	<i>Atta</i> (n=2)	<i>Camponotus</i> (n=5)	2.506	0.091	0.228
		<i>Cephalotes</i> (n=3)	7.822	0.114	0.228
		NA (n=6)	4.691	0.056	0.228
	<i>Camponotus</i> (n=5)	<i>Cephalotes</i> (n=3)	0.897	0.540	0.540
		NA (n=6)	1.098	0.346	0.415
	<i>Cephalotes</i> (n=3)	NA (n=6)	0.984	0.346	0.415
Nest	Ground (n=6)	Rotten (n=4)	0.800	0.446	0.446
		Tree (n=6)	3.143	0.029	0.087
	Rotten (n=4)	Tree (n=6)	1.164	0.347	0.446
Habitat	Rainforest (n=8)	City (n=8)	1.295	0.194	0.194

Supplementary File 11. SIMPER statistics for each factor.

SIMPER statistics according to the ant subfamily

11 most influential subphyla	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)				
				Dolichoderinae	Formicinae	Myrmicinae	Ponerinae	Pseudomyrmicinae
Undetermined	18.46	44	44.00	100.00	72.70	73.90	51.40	99.30
Nematoda	12.05	28.71	72.70	0.000	4.920	13.50	48.60	0.000
Eugregarinorida	6.205	14.79	87.49	0.000	19.60	0.493	0.000	0.000
Chytridiomycota	4.045	9.368	97.13	0.000	0.000	11.30	0.000	0.000
Mortierellomycotina	0.301	0.717	97.84	0.000	0.962	0.000	0.000	0.000
Saccharomycotina	0.168	0.401	98.24	0.000	0.323	0.094	0.000	0.655
Ustilaginomycotina	0.158	0.376	98.62	0.000	0.455	0.056	0.000	0.000
Pezizomycotina	0.084	0.201	98.82	0.000	0.000	0.235	0.000	0.000
Mucoromycotina	0.070	0.168	98.99	0.000	0.225	0.000	0.000	0.000
Basidiobolomycetes	0.069	0.164	99.15	0.000	0.000	0.193	0.000	0.000
Arthropoda	0.041	0.097	99.25	0.000	0.000	0.113	0.000	0.000

SIMPER statistics according to the ant genus

11 most influential subphyla	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)								
				<i>Azteca</i>	<i>Camponotus</i>	<i>Atta</i>	<i>Cephalotes</i>	<i>Daceton</i>	<i>Solenopsis</i>	<i>Neoponera</i>	<i>Odontomachus</i>	<i>Pseudomyrmex</i>
Undetermined	19.88	45.09	45.09	100.00	72.70	12.50	97.40	100.00	100.00	96.50	6.220	99.30
Nematoda	12.65	28.70	73.79	0.000	4.920	47.20	0.000	0.000	0.000	3.470	93.80	0.000
Chytridiomycota	5.217	11.83	97.46	0.000	19.60	0.000	1.150	0.000	0.000	0.000	0.000	0.000
Eugregarinorida	5.223	11.85	85.63	0.000	0.000	39.50	0.000	0.000	0.000	0.000	0.000	0.000
Mortierellomycotina	0.250	0.567	98.03	0.000	0.962	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Saccharomycotina	0.152	0.345	98.38	0.000	0.323	0.000	0.219	0.000	0.000	0.000	0.000	0.655
Ustilaginomycotina	0.139	0.314	98.69	0.000	0.455	0.000	0.132	0.000	0.000	0.000	0.000	0.000
Pezizomycotina	0.098	0.223	98.91	0.000	0.000	0.101	0.481	0.000	0.000	0.000	0.000	0.000
Basidiobolomycetes	0.058	0.132	99.25	0.000	0.225	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mucoromycotina	0.089	0.202	99.11	0.000	0.000	0.674	0.000	0.000	0.000	0.000	0.000	0.000
Arthropoda	0.049	0.110	99.36	0.000	0.000	0.000	0.264	0.000	0.000	0.000	0.000	0.000

SIMPER statistics according to the ant nesting mode

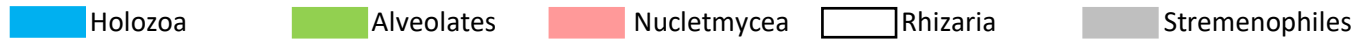
11 most influential subphyla	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)		
				Ground	Rotten	Tree
Undetermined	21.11	46.42	46.42	49.10	74.20	98.60
Nematoda	12.53	27.56	73.98	20.40	23.40	0.000
Eugregarinorida	5.925	13.03	87.01	14.70	2.400	0.576
Chytridiomycota	4.708	10.35	97.36	13.20	0.000	0.000
Mortierellomycotina	0.286	0.630	97.99	0.802	0.000	0.000
Saccharomycotina	0.159	0.349	98.34	0.270	0.000	0.219
Ustilaginomycotina	0.154	0.339	98.68	0.379	0.000	0.066
Pezizomycotina	0.093	0.205	98.88	0.034	0.000	0.241
Mucoromycotina	0.08	0.177	99.06	0.225	0.000	0.000
Basidiobolomycetes	0.067	0.147	99.21	0.188	0.000	0.000
Arthropoda	0.047	0.104	99.31	0.000	0.000	0.132

SIMPER statistics according to the ant habitat

11 most influential subphyla	Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)	
				City	Rainforest
Undetermined	20.02	44.76	44.76	61.90	86.00
Nematoda	12.26	27.41	72.17	23.70	3.360
Eugrargarinorida	6.270	14.02	86.19	12.50	0.186
Chytridiomycota	4.944	11.05	97.25	0.000	9.890
Mortierellomycotina	0.281	0.629	97.88	0.523	0.078
Saccharomycotina	0.164	0.366	98.24	0.292	0.074
Ustilaginomycotina	0.161	0.359	98.60	0.284	0.049
Pezizomycotina	0.094	0.209	98.81	0.131	0.075
Basidiobolomycetes	0.070	0.157	99.16	0.141	0.000
Mucoromycotina	0.084	0.189	99.00	0.000	0.169
Arthropoda	0.050	0.111	99.27	0.099	0.000

Supplementary File 12. SIMPER statistics for each factor after removing the undetermined eukaryotes. The color of the cells correspond to the eukaryote infrakingdom.

Tables color legend



SIMPER statistics according to the ant subfamily

Most influential phyla			Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)			
Subphylum	Infrakingdom	Phylum				Formicinae	Myrmicinae	Ponerinae	Pseudomyrmicinae
Nematoda	Holozoa	Metazoa	30.47	35.53	35.53	63.00	115.00	108.00	0.00
Gregarinasina	Alveolata	Apicomplexa	28.77	33.55	69.08	342.00	60.80	0.000	0.000
Chytridiomycota	Nucleomyces	Fungi	11.73	13.68	82.76	0.000	293.00	0.000	0.000
Pezizomycotina	Nucleomyces	Fungi	3.140	3.661	86.42	0.000	22.30	0.000	0.000
Saccharomycotina	Nucleomyces	Fungi	2.879	3.358	89.78	5.750	2.750	0.000	9.000
Mortierellomycotina	Nucleomyces	Fungi	2.145	2.502	92.28	13.50	0.000	0.000	0.000
Ustilaginomycotina	Nucleomyces	Fungi	1.757	2.049	94.33	0.000	15.50	0.000	0.000
Arthropoda	Holozoa	Metazoa	1.729	2.017	96.35	6.500	1.000	0.000	0.000
Dinophyceae	Alveolata	Dinoflagellata	1.320	1.539	97.89	0.000	2.750	0.000	0.000
Phyllopharyngea	Alveolata	Ciliophora	0.453	0.529	98.42	0.000	4.000	0.000	0.000
Colpodea	Alveolata	Ciliophora	0.340	0.397	98.81	0.000	3.000	0.000	0.000
Basidiobolomycetes	Nucleomyces	Fungi	0.200	0.233	99.05	0.000	5.000	0.000	0.000
Rhizaspidae	Rhizaria	Cercozoa	0.198	0.231	99.28	0.000	1.750	0.000	0.000
Oligohymenophorea	Alveolata	Ciliophora	0.171	0.200	99.48	4.000	0.000	0.000	0.000
Mucoromycotina	Nucleomyces	Fungi	0.170	0.198	99.68	0.000	1.500	0.000	0.000
Chrysophyceae	Stremenophiles	Ochrophyta	0.118	0.138	99.81	2.750	0.000	0.000	0.000
Annelida	Holozoa	Metazoa	0.096	0.113	99.93	2.250	0.000	0.000	0.000
Puccinimycotina	Nucleomyces	Fungi	0.064	0.075	100.00	1.500	0.000	0.000	0.000

SIMPER statistics according to the ant genus

Most influential phyla			Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)					
Subphylum	Infrakingdom	Phylum				<i>Camponotus</i>	<i>Atta</i>	<i>Cephalotes</i>	<i>Neoponera</i>	<i>Odontomachus</i>	<i>Pseudomyrmex</i>
Nematoda	Holozoa	Metazoa	32.22	37.08	37.08	63.00	230.00	0.000	19.00	196.00	0.000
Gregarinasina	Alveolata	Apicomplexa	26.83	30.88	67.96	342.00	0.000	122.00	0.000	0.000	0.000
Chytridiomycota	Nucleomycea	Fungi	13.49	15.52	83.48	0.000	587.00	0.000	0.000	0.000	0.000
Pezizomycotina	Nucleomycea	Fungi	3.210	3.694	87.18	0.000	1.500	43.00	0.000	0.000	0.000
Saccharomycotina	Nucleomycea	Fungi	2.661	3.063	90.24	5.750	0.000	5.500	0.000	0.000	9.000
Mortierellomycotina	Nucleomycea	Fungi	1.917	2.207	92.45	13.50	0.000	0.000	0.000	0.000	0.000
Arthropoda	Holozoa	Metazoa	1.848	2.127	94.58	0.000	0.000	31.00	0.000	0.000	0.000
Ustilaginomycotina	Nucleomycea	Fungi	1.589	1.829	96.41	6.500	0.000	2.000	0.000	0.000	0.000
Dinophyceae	Alveolata	Dinoflagellata	1.268	1.460	97.87	0.000	0.000	5.500	0.000	0.000	0.000
Phyllopharyngea	Alveolata	Ciliophora	0.477	0.549	98.41	0.000	0.000	8.000	0.000	0.000	0.000
Colpodea	Alveolata	Ciliophora	0.358	0.412	98.83	0.000	0.000	6.000	0.000	0.000	0.000
Basidiobolomycetes	Nucleomycea	Fungi	0.230	0.265	99.09	0.000	10.00	0.000	0.000	0.000	0.000
Rhizaspidae	Rhizaria	Cercozoa	0.209	0.240	99.33	0.000	0.000	3.500	0.000	0.000	0.000
Oligohymenophorea	Alveolata	Ciliophora	0.179	0.206	99.54	0.000	0.000	3.000	0.000	0.000	0.000
Mucoromycotina	Nucleomycea	Fungi	0.153	0.176	99.71	4.000	0.000	0.000	0.000	0.000	0.000
Chrysophyceae	Stremenophiles	Ochrophyta	0.105	0.121	99.83	2.750	0.000	0.000	0.000	0.000	0.000
Annelida	Holozoa	Metazoa	0.087	0.099	99.93	2.250	0.000	0.000	0.000	0.000	0.000
Puccinimycotina	Nucleomycea	Fungi	0.057	0.067	100.00	1.500	0.000	0.000	0.000	0.000	0.000

SIMPER statistics according to the ant nesting mode

Most influential phyla			Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)		
Subphylum	Infrakingdom	Phylum				Ground	Rotten	Tree
Nematoda	Holozoa	Metazoa	34.13	38.14	38.14	122.00	98.00	0.00
Gregarinasina	Alveolata	Apicomplexa	30.21	33.77	71.91	209.00	57.50	81.00
Chytridiomycota	Nucleomycea	Fungi	10.190	11.39	83.30	196.00	0.000	0.000
Saccharomycotina	Nucleomycea	Fungi	3.321	3.711	87.01	0.500	0.000	28.70
Pezizomycotina	Nucleomycea	Fungi	3.129	3.497	90.51	3.830	0.000	6.670
Arthropoda	Holozoa	Metazoa	2.024	2.262	92.77	0.000	0.000	20.70
Mortierellomycotina	Nucleomycea	Fungi	1.952	2.182	94.95	9.000	0.000	0.000
Ustilaginomycotina	Nucleomycea	Fungi	1.487	1.662	96.61	4.330	0.000	1.330
Dinophyceae	Alveolata	Dinoflagellata	1.119	1.251	97.86	0.000	0.000	3.670
Phyllopharyngea	Alveolata	Ciliophora	0.522	0.584	98.45	0.000	0.000	5.330
Colpodea	Alveolata	Ciliophora	0.392	0.438	98.88	0.000	0.000	4.000
Rhizaspidae	Rhizaria	Cercozoa	0.229	0.255	99.14	0.000	0.000	2.330
Oligohymenophorea	Alveolata	Ciliophora	0.196	0.219	99.36	0.000	0.000	2.000
Basidiobolomycetes	Nucleomycea	Fungi	0.174	0.194	99.55	3.330	0.000	0.000
Mucoromycotina	Nucleomycea	Fungi	0.153	0.17	99.72	2.670	0.000	0.000
Chrysophyceae	Stremenophiles	Ochrophyta	0.105	0.117	99.84	1.830	0.000	0.000
Annelida	Holozoa	Metazoa	0.086	0.096	99.94	1.500	0.000	0.000
Puccinimycotina	Nucleomycea	Fungi	0.057	0.064	100.00	1.000	0.000	0.000

SIMPER statistics according to the ant habitat

Most influential phyla			Average dissimilarity	Contribution to difference (%)	Cumulative (%)	Mean read number per group (%)	
Subphylum	Infrakingdom	Phylum				City	Rainforest
Nematoda	Holozoa	Metazoa	32.52	36.11	36.11	87.10	79.30
Gregarinasina	Alveolata	Apicomplexa	27.35	30.37	66.48	228.00	3.750
Chytridiomycota	Nucleotmycea	Fungi	16.72	18.57	85.06	0.000	293.00
Pezizomycotina	Nucleotmycea	Fungi	2.766	3.072	88.13	11.70	1.750
Saccharomycotina	Nucleotmycea	Fungi	2.729	3.031	91.16	5.290	1.500
Mortierellomycotina	Nucleotmycea	Fungi	1.935	2.149	93.31	7.140	1.000
Ustilaginomycotina	Nucleotmycea	Fungi	1.609	1.786	95.10	3.710	1.000
Arthropoda	Holozoa	Metazoa	1.545	1.716	96.81	8.860	0.000
Dinophyceae	Alveolata	Dinoflagellata	1.177	1.307	98.12	0.714	1.500
Phyllopharyngea	Alveolata	Ciliophora	0.399	0.443	98.56	2.290	0.000
Colpodea	Alveolata	Ciliophora	0.299	0.332	98.89	1.710	0.000
Basidiobolomycetes	Nucleotmycea	Fungi	0.285	0.317	99.21	0.000	5.000
Rhizaspididae	Rhizaria	Cercozoa	0.174	0.194	99.40	1.000	0.000
Oligohymenophorea	Alveolata	Ciliophora	0.150	0.166	99.57	0.857	0.000
Mucoromycotina	Nucleotmycea	Fungi	0.148	0.164	99.73	2.290	0.000
Chrysophyceae	Stremenophiles	Ochrophyta	0.102	0.113	99.85	1.570	0.000
Annelida	Holozoa	Metazoa	0.083	0.092	99.94	1.290	0.000
Puccinimycotina	Nucleotmycea	Fungi	0.055	0.062	100.00	0.857	0.000

Supplementary File 13. SIMPER comparisons for each factor. Only the comparisons in which the bacterial relative abundance were statistically different (p-value < 0.05) are shown here. The statistical p-values were obtained with pairwise Kruskal-Wallis.

Factor	Comparison	SIMPER	Subphylum	p-value	fdr_correction	Left.mean.abund	Left.stdev	Right.mean.abund	Right.stdev
Nest	Ground_Tree	0.201	Nematoda	0.007	0.041	0.204	0.287	0.000	0.000
	Ground_Tree	0.489	Undetermined	0.006	0.041	0.491	0.437	0.986	0.022

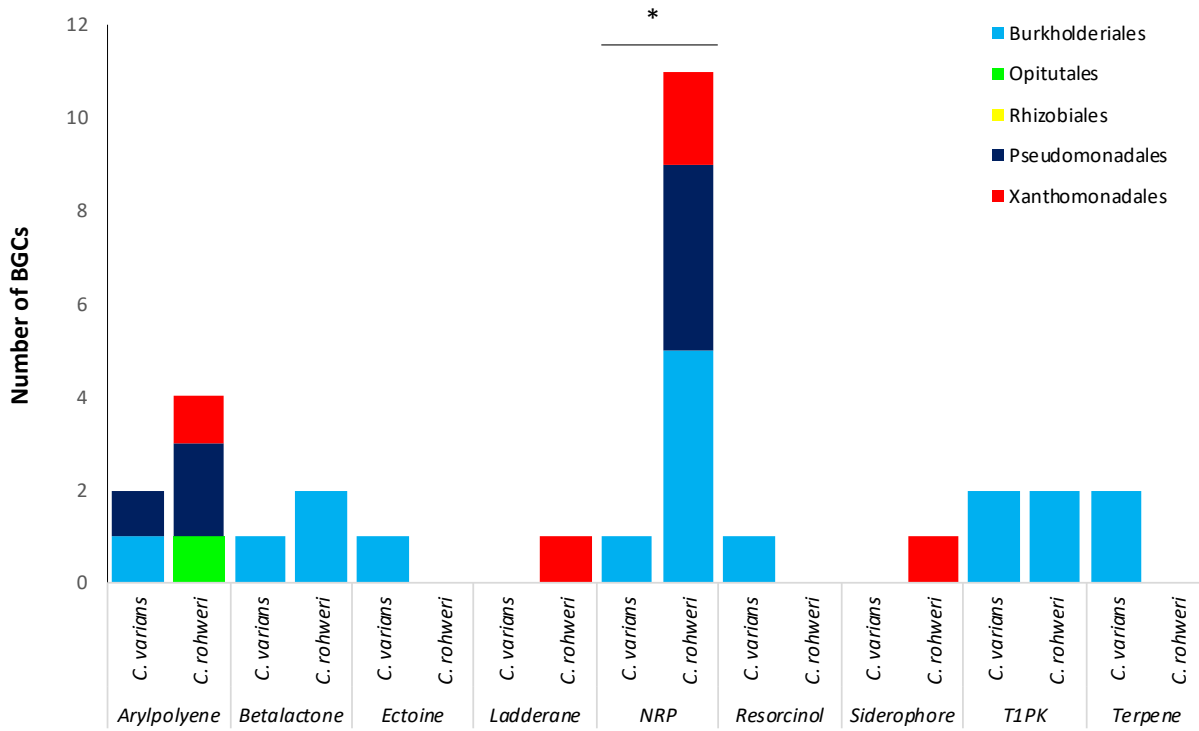
Supplementary File 14. Description of the eukaryote subphyla identified in our samples. The genera in bold are genera which were found in associations with our Amazonian ant samples.

Eukaryote subphyla	Description	Genus examples
Nematoda	Diverse subphylum containing a large number of parasitic species, some of them being entomopathogenic	<i>Heterorhabditis</i> ; <i>Steinernema</i>
Gregarinasina	A subphylum of alveolates inhabiting the intestines of many invertebrates	<i>Exoschizon</i> ; <i>Gregarina</i>
Saccharomycotina	A subphylum containing most of the ascomycete yeasts, with some of its members being associated with insects	<i>Saccharomyces</i> ; <i>Starmerella</i>
Chytridiomycota	Zoosporic fungi with some members found on scale of insects	<i>Batrachochytrium</i> ; <i>Myiophagus</i>
Pezizomycotina	Filamentous and lichenized fungi with some members closely associated with insects	<i>Ascobolus</i> ; <i>Urceolaria</i>
Dinophyceae	Dinoflagellates living in marine environment, with some members being parasitic of marine organisms	<i>Ceratocorys</i> ; <i>Ostreopsis</i>
Ustilaginomycotina	A subphylum of fungi, composed of mostly parasites of vascular plants and associated with insects	<i>Malassezia</i> ; <i>Ustilago</i>
Arthropoda	Invertebrate animals	<i>Achipteria</i> ; <i>Sarcoptes</i>
Mortierellomycotina	Subphylum composed of soil fungi, with some members being weak insect parasites	<i>Mortierella</i>
Phyllopharyngea	Ciliates living in marine environments, with some members being ectosymbionts of crustaceans	<i>Chilodonella</i> ; <i>Vasichona</i>
Colpodea	Ciliates common in freshwater and soil habitats	<i>Colpoda</i> ; <i>Kreyella</i>
Rhizaspidae	Protists living in marine environments	<i>Capsellina</i> ; <i>Rhogostoma</i>
Oligohymenophorea	Ciliates living in marine environments, with some members being parasites of aquatic organisms	<i>Cothurnia</i> ; <i>Trichodina</i>
Basidiobolomycetes	Fungi found in the intestines of cold-blooded vertebrates and on decaying fruits and soil	<i>Basidiobolus</i> ; <i>Schizangiella</i>
Mucoromycotina	Fungi which can be insect parasitic	<i>Endogone</i> ; <i>Mucor</i>
Chrysophyceae	Algae found mostly in freshwater	<i>Hydrurus</i> ; <i>Stichogloea</i>
Annelida	Segmented worms living in aquatic or terrestrial moist environments	<i>Haplotaxis</i> ; <i>Palola</i>
Pucciniomycotina	Subphylum of entomopathogenic fungi	<i>Agaricostibum</i> ; <i>Puccinia</i>

Supplementary File 15. Statistics for the co-occurrence network between bacterial ASVs and eukaryote ASVs. Only correlations between two ASVs with a p-value lower than 0.05 are taken into account.

Source	Target	Interaction	p-value	R	p-Adj	Directed
Bac_5639	Euk_7	interacts with	0.004436	0.875	0.163657	FALSE
Bac_5440	Euk_52	interacts with	0.036984	0.736989	0.565005	FALSE
Bac_3451	Euk_52	interacts with	0.009167	0.839286	0.276679	FALSE
Bac_3451	Euk_118	interacts with	0.009167	0.839286	0.276679	FALSE
Bac_1492	Euk_52	interacts with	0.032104	0.75	0.565005	FALSE
Bac_1492	Euk_118	interacts with	0.032104	0.75	0.565005	FALSE
Bac_5440	Euk_118	interacts with	3.70E-02	0.736989	0.565005	FALSE
Bac_4577	Euk_7	interacts with	0	1	0	FALSE
Bac_4376	Euk_7	interacts with	0	1	0	FALSE
Bac_2189	Euk_7	interacts with	0	1	0	FALSE
Bac_5909	Euk_7	interacts with	0	1	0	FALSE
Bac_4600	Euk_52	interacts with	0	1	0	FALSE
Bac_4600	Euk_118	interacts with	0	1	0	FALSE
Bac_101	Euk_72	interacts with	8.14E-04	0.929981	0.038595	FALSE

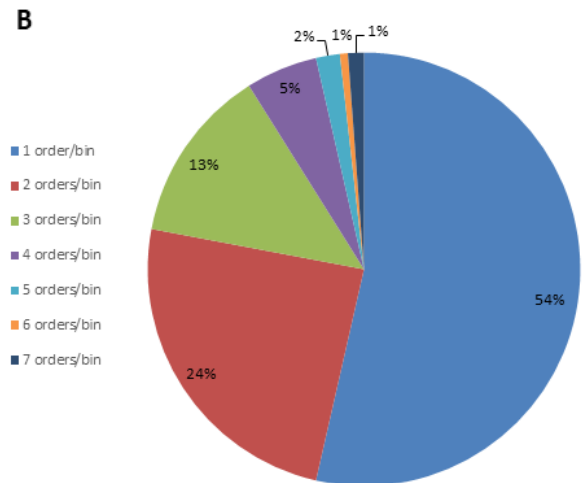
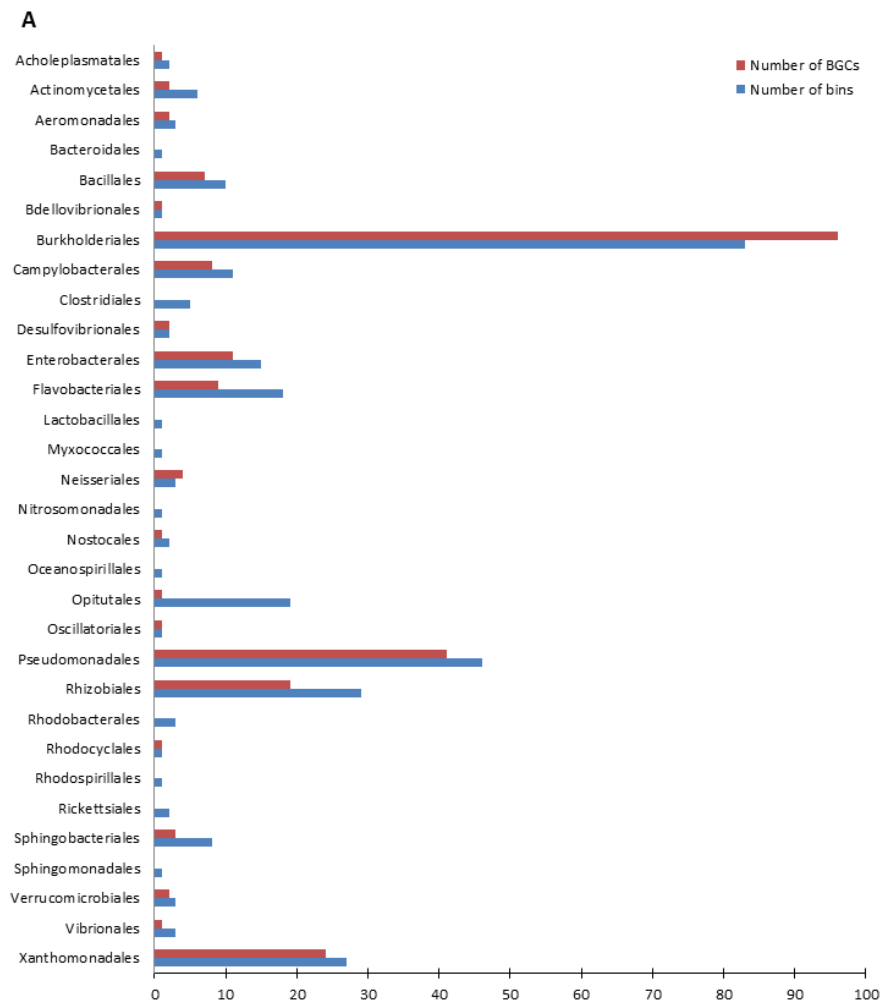
Annexe 3. Supporting Information de l'article Assessing biosynthetic gene cluster diversity of specialized metabolites in the conserved gut symbionts of herbivorous turtle ants



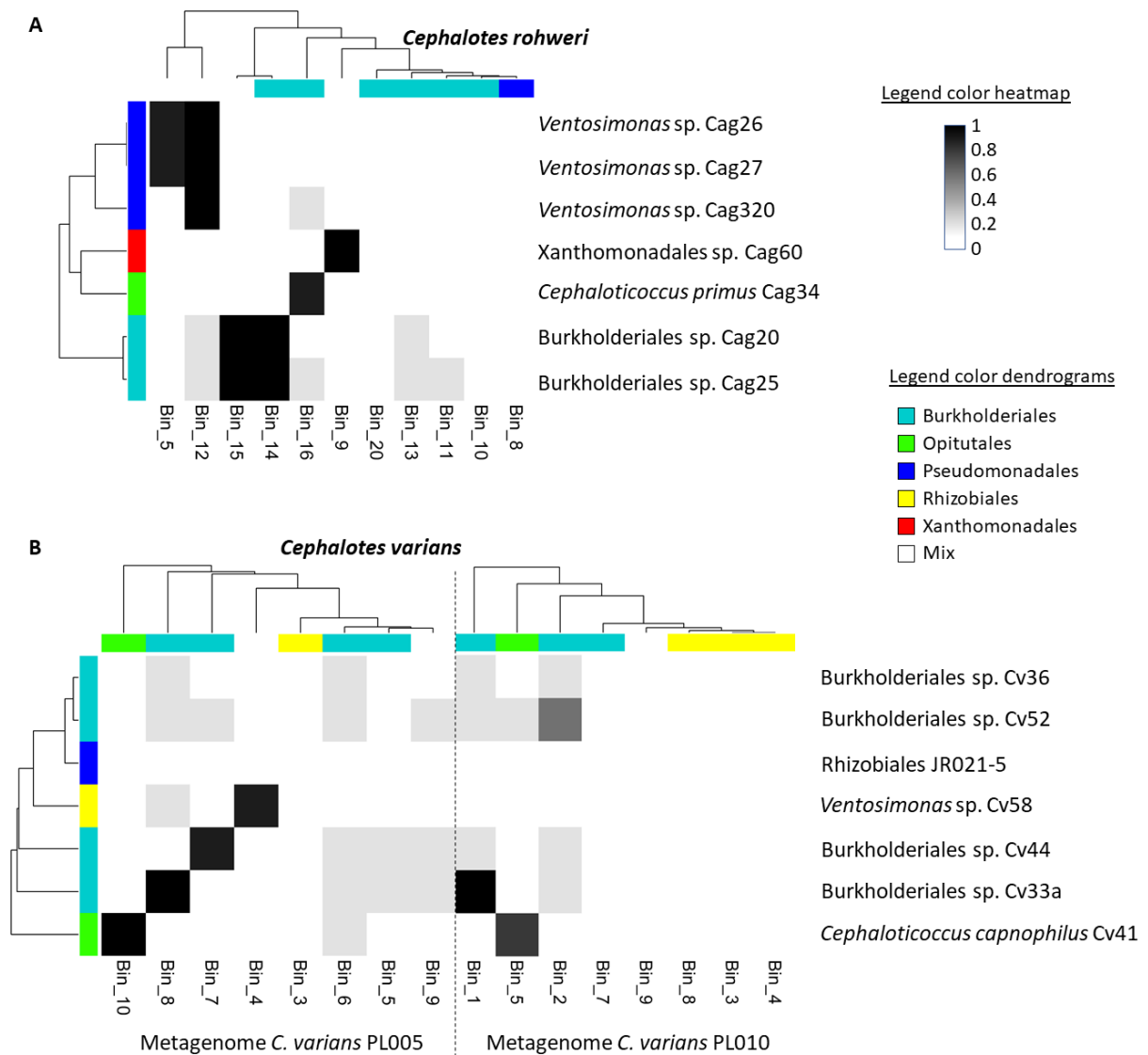
Supplementary Figure 18. Assessment of the differences between the BGCs in the genome analysis.

T-test showing the statistical differences between the bacterial BGCs across isolated bacteria species

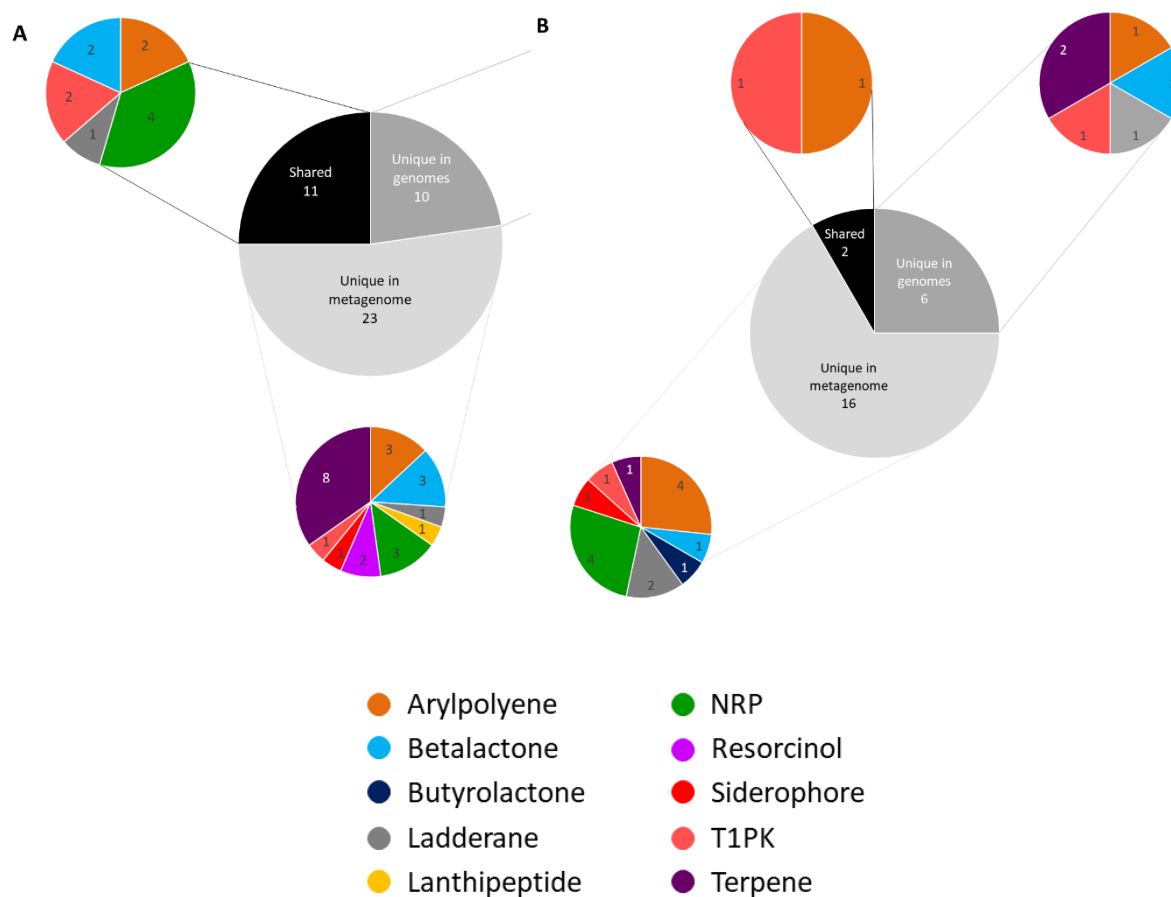
and Cephalotes species. The symbol * represents a p-value lower than 0.05.



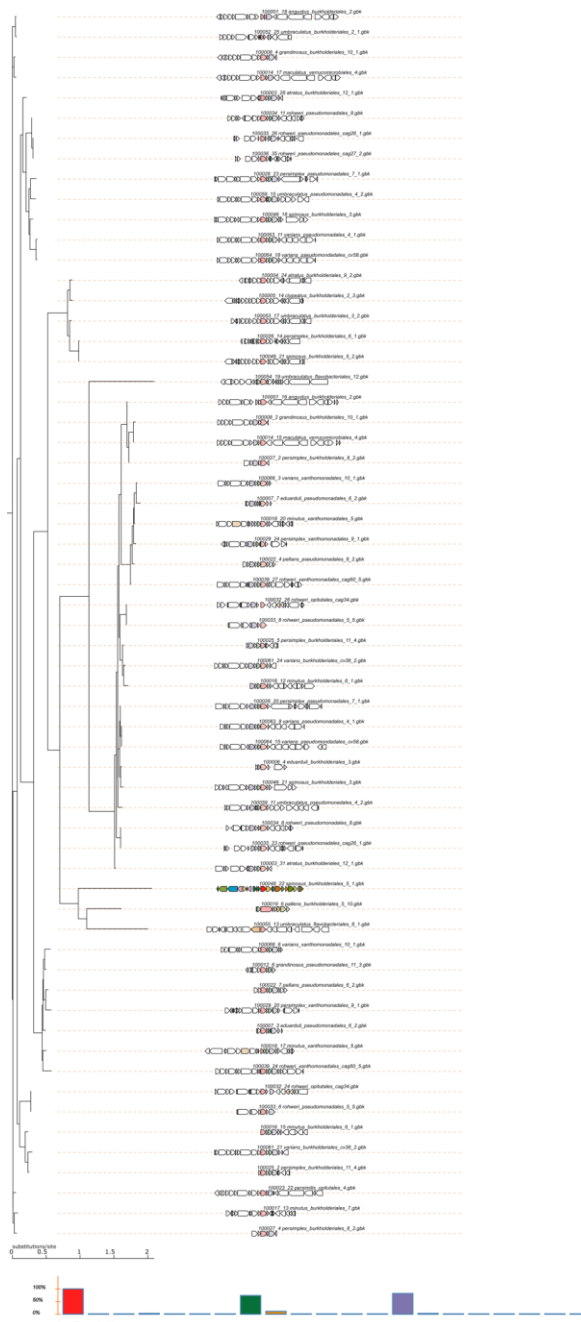
Supplementary Figure 19. Assessment of the bacterial composition of the metagenomic bins. (A) List of bacterial orders identified in the bins identified in metagenomes. The blue bars represent the number of bins in which each bacterial order was found. The red bars represent the number of BGCs possessed by each bacterial order. (B) The proportion of different bacterial orders found in the bin.



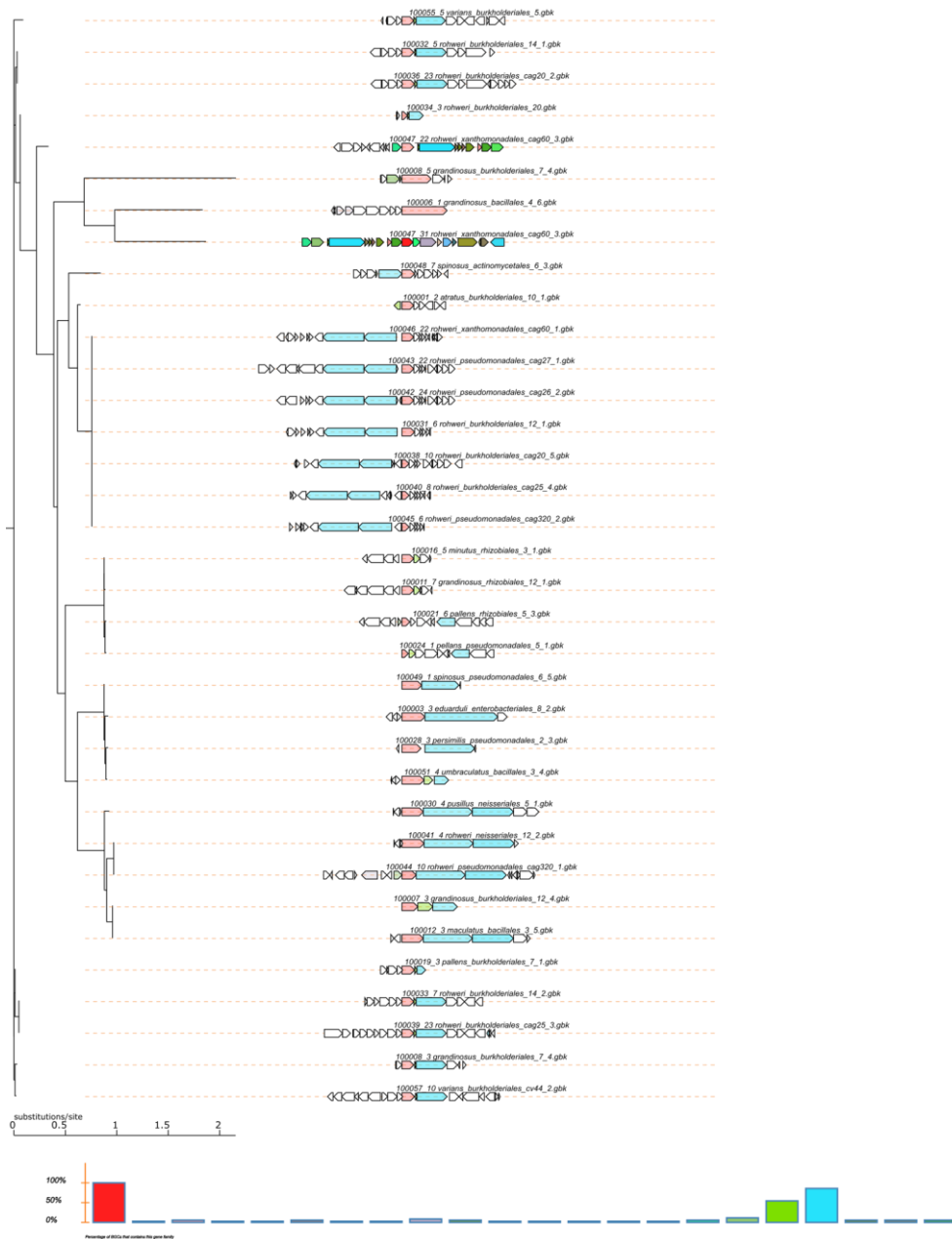
Supplementary Figure 20. Heat map of relationships between isolated genomes and metagenomic bins for *C. rohweri* (A) and *C. varians* (B). The scale bar of the heatmap is shaded from black to white according to identity on the basis of calculated values (gANI x AF). Color bars in the dendrograms indicate bacterial orders identified either in the metagenomic bins or in the isolated genomes. The “mix” term in the color dendrograms legend refers to metagenomic bins in which more than one different bacterial order was identified.



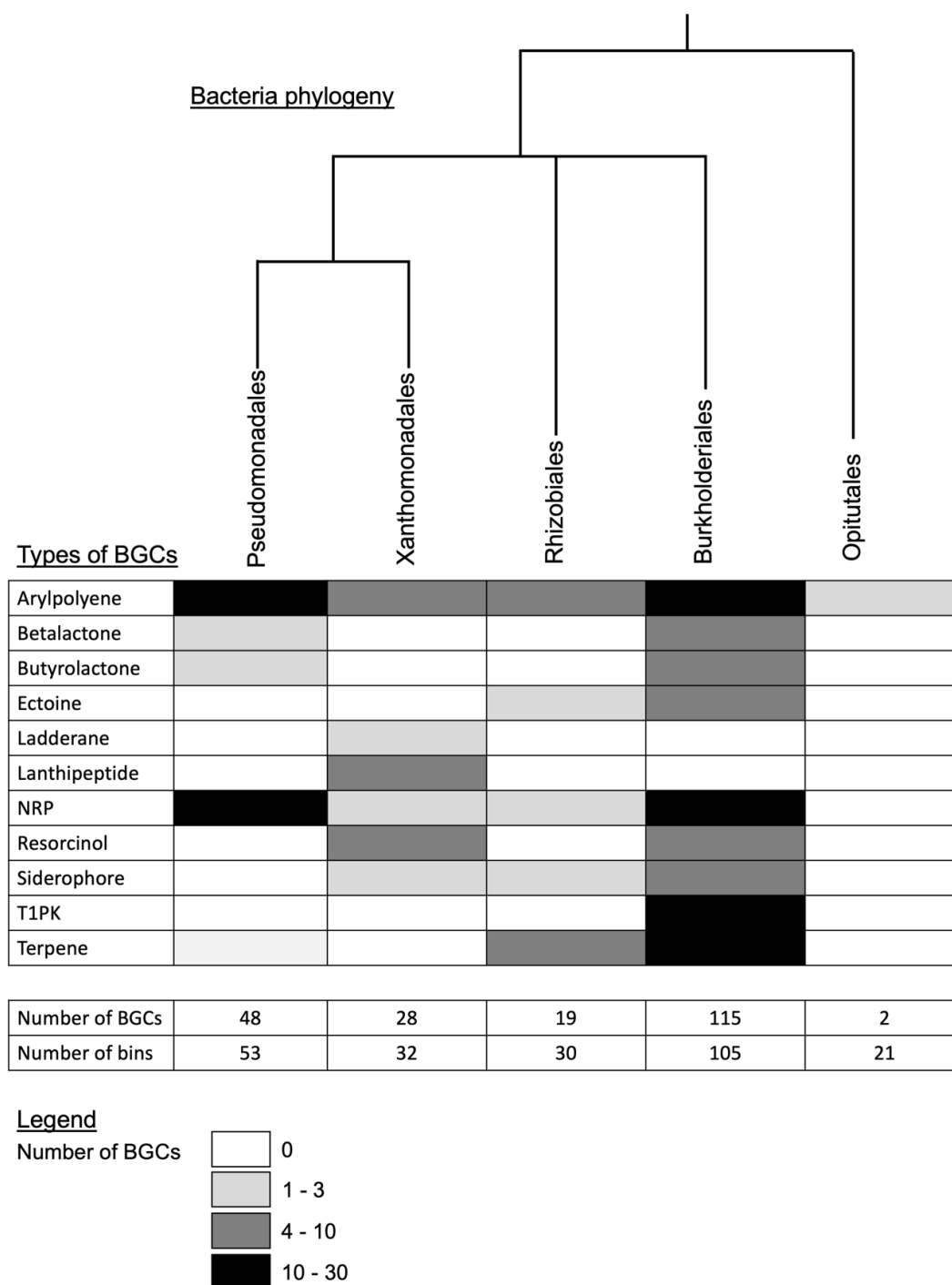
Supplementary Figure 21. Number of shared BGCs found in the metagenomic bins and isolated genomes, for *C. rohweri* and *C. varians*. The central pies represent the number of BGCs shared between the genomes and the metagenomes, and the number of BGCs which were only found in the genomes or only in the metagenomes. The peripheric pies represent the type and number of BGCs for each category (shared, unique in genomes or unique in metagenome). The color corresponding to each type of BGCs can be found in the legendscale bar of the heatmap is shaded from black to white according to identity on the basis of calculated values ($gANI \times AF$). Color bars in the dendrograms indicate bacterial orders identified either in the metagenomic bins or in the isolated genomes. The “mix” term in the color dendrograms legend refers to metagenomic bins in which more than one different bacterial order was identified.



Supplementary Figure 22. Genomic core analysis and phylogenetic tree of the arylpolyene BGCs from the *Cephalotes* bacterial genomes and metagenomes. This phylogeny was implemented through CORASON by uploading all the arylpolyene BGCs retrieved from the *Cephalotes* associated genomes and metagenomes into a single database and by comparing their sequences and architectures to a reference BGC and query protein as detailed in Supplementary Table 8. The histogram represents the percentage of BGC that contains this gene family.



Supplementary Figure 23. Genomic core analysis and phylogenetic tree of the NRP BGCs from the *Cephalotes* bacterial genomes and metagenomes. This phylogeny was implemented through CORASON by uploading all the NRP BGCs retrieved from the *Cephalotes* associated genomes and metagenomes into a single database and by comparing their sequences and architectures to a reference BGC and query protein as detailed in Supplementary Table 8. The histogram represents the percentage of BGC that contains this gene family.



Supplementary Figure 24. Number and type of BGCs identified in the genomes and metagenomes of the *Cephalotes* gut core symbionts. The black to white shading filling in the table represent the number of BGCs of each type identified in the genomes and metagenomes of the 5 main bacterial order associated with the *Cephalotes* gut. The blue shading filling represent the number of bins in which each bacterial order has been found.

Supplementary Table 1. Assembly statistics of genomes. This table was extracted from the

Supplementary Information of Hu *et al.*, 2018.

Genome name	Taxonomic affiliation	Total length (bp)	Scaffold number	N50 (bp)	GC% mean	IMG project ID	Ant host
<i>Cephalotococcus primus</i> Cag34	<u>Order:</u> Opitutales <u>Class:</u> Opitutae <u>Phylum:</u> Verrucomicrobia	2,353,466	23	207,028	62.43%	Gp0154034	<i>C. rohweri</i>
<i>Cephalotococcus canophilus</i> Cv41	<u>Order:</u> Opitutales <u>Class:</u> Opitutae <u>Phylum:</u> Verrucomicrobia	2,094,663	34	169,871	59.29%	Gp0110136	<i>C. varians</i>
Rhizobiales sp. JR021-5	<u>Order:</u> Rhizobiales <u>Class:</u> Alphaproteobacteria <u>Phylum:</u> Proteobacteria	1,943,462	1	1,943,462	57.28%	Gp0155340	<i>C. varians</i>
Xanthomonadales sp. Cag60	<u>Order:</u> Xanthomonadales <u>Class:</u> Gammaproteobacteria <u>Phylum:</u> Proteobacteria	3,424,135	179	44,406	56.17%	Gp0127963	<i>C. rohweri</i>
<i>Ventosimonas gracilis</i> Cv58	<u>Order:</u> Pseudomonadales <u>Class:</u> Gammaproteobacteria <u>Phylum:</u> Proteobacteria	2,623,100	221	58,779	53.40%	Gp0110146	<i>C. varians</i>
<i>Ventosimonas</i> sp. Cag26	<u>Order:</u> Pseudomonadales <u>Class:</u> Gammaproteobacteria <u>Phylum:</u> Proteobacteria	2,729,219	92	79,155	54.82%	Gp0154031	<i>C. rohweri</i>
<i>Ventosimonas</i> sp. Cag27	<u>Order:</u> Pseudomonadales <u>Class:</u> Gammaproteobacteria <u>Phylum:</u> Proteobacteria	2,719,975	89	79,155	54.89%	Gp0154032	<i>C. rohweri</i>
<i>Ventosimonas</i> sp. Cag320	<u>Order:</u> Pseudomonadales <u>Class:</u> Gammaproteobacteria <u>Phylum:</u> Proteobacteria	2,825,500	94	108,424	55.67%	Gp0154033	<i>C. rohweri</i>

Burkholderiales <i>sp. Cag20</i>	<u>Order:</u> Burkholderiales <u>Class:</u> Betaproteobacteria <u>Phylum:</u> Proteobacteria	3,101,960	121	69,296	58.26%	Gp0154021	<i>C. rohweri</i>
Burkholderiales <i>sp. Cag25</i>	<u>Order:</u> Burkholderiales <u>Class:</u> Betaproteobacteria <u>Phylum:</u> Proteobacteria	3,191,349	112	142,682	58.13%	Gp0154030	<i>C. rohweri</i>
Burkholderiales <i>sp. Cv44</i>	<u>Order:</u> Burkholderiales <u>Class:</u> Betaproteobacteria <u>Phylum:</u> Proteobacteria	3,159,338	601	33,742	60.40%	Gp0110143	<i>C. varians</i>
Burkholderiales <i>sp. Cv33a</i>	<u>Order:</u> Burkholderiales <u>Class:</u> Betaproteobacteria <u>Phylum:</u> Proteobacteria	2,469,676	150	48,592	58.91%	Gp0110144	<i>C. varians</i>
Burkholderiales <i>sp. Cv26</i>	<u>Order:</u> Burkholderiales <u>Class:</u> Betaproteobacteria <u>Phylum:</u> Proteobacteria	2,921,795	234	54,338	59.95%	Gp0110137	<i>C. varians</i>
Burkholderiales <i>sp. Cv52</i>	<u>Order:</u> Burkholderiales <u>Class:</u> Betaproteobacteria <u>Phylum:</u> Proteobacteria	2,935,707	203	93,689	60.36%	Gp0110145	<i>C. varians</i>

Supplementary Table 2. Assembly statistics of metagenomes. This table was extracted from the Supplementary Information of : Hu, Y. *et al.* (2018) Herbivorous turtle ants obtain essential nutrients from a highly conserved nitrogen-recycling gut microbiome. *Nat. Commun.* 9, 964.

Ant species	Bins number	Reads numbers	Total length Mbp	Total length Scaffold >=1Kbp (Mbp)	Scaffold number	Scaffold >=1Kbp number	N50 for scaffold length (bp)	GC% mean	Read coverage mean	IMG project ID
<i>C. varians</i> PL005W	11	40,804,316	184.98	142.43	281,884	129,829	1,074	39.18%	16.60	Gp0095985
<i>C. varians</i> PL010W	11	135,924,092	436.39	340.3	647,079	91,083	10,750	41.95%	41.75	Gp009598
<i>C. angustus</i>	2	3,534,316	21.5	8.22	30,858	3,531	1,118	50.80%	6.79	Gp0125961
<i>C. atratus</i>	15	34,205,408	218.88	141.37	198,281	67,767	1,488	38.53%	9.20	Gp0125962
<i>C. clypeatus</i>	6	24,093,876	104.18	41.7	135,743	16,952	1,031	37.42%	8.62	Gp0125963
<i>C. eduarduli</i>	8	22,505,270	82.63	37.38	100,267	12,793	1,259	39.99%	9.53	Gp0125964
<i>C. grandinosus</i>	12	37,871,910	156.24	60.81	220,221	22,090	1,215	39.68%	10.06	Gp0125967
<i>C. maculatus</i>	6	40,739,970	144.43	42.17	230,315	18,066	921	39.60%	7.52	Gp0125968
<i>C. minutus</i>	8	31,576,424	115.64	37.69	180,629	15,101	1,038	38.38%	8.10	Gp0125969
<i>C. pallens</i>	10	28,073,250	148.44	55.05	211,667	21,214	1,095	39.56%	7.92	Gp0125970
<i>C. pellans</i>	7	33,836,572	142.82	43.13	224,021	18,268	934	38.32%	8.51	Gp0126569
<i>C. persimilis</i>	6	28,044,242	106.24	44.84	133,984	19,037	1,077	40.41%	10.09	Gp0126571
<i>C. persimplex</i>	11	40,440,804	231.49	102.43	291,903	53,617	1,090	37.94%	10.14	Gp0126580
<i>C. pusillus</i>	5	18,252,188	61.89	24.03	93,263	9,713	1,389	42.35%	6.67	Gp0126572
<i>C. rohweri</i>	22	58,943,942	305.84	271.58	142,961	78,950	3,982	40.47%	22.40	Gp0126573
<i>C. similimus</i>	6	30,304,070	92.55	37.04	122,379	16,294	1,030	37.75%	13.09	Gp0126574
<i>C. spinosus</i>	8	19,456,806	94.81	41	117,242	13,728	1,206	38.79%	8.66	Gp0126575
<i>C. umbraculatus</i>	14	37,344,984	256.42	193.69	191,339	84,591	1,984	37.76%	14.53	Gp0126577

Supplementary Table 3. Statistic assessment of the metagenomic bins created. The metagenomic bins were created by uploading each metagenome sequences into the Anvi'o version 5.5. This table contains for each *Cephalotes* metagenome the number of bins created, and for each bin it shows the percentage of completeness, the percentage of contamination, the percentage of strain heterogeneity, the number of contigs, the number of genes, and the number of BGCs found. The percentage of completeness, contamination and strain heterogeneity were obtained through the CheckM software. The number of contigs and the number of genes were obtained through the Anvi'o software. The number of BGC was obtained through the AntiSMASH software.

Host species	Number of bins	Completeness bin (%)	CheckM completeness class	Contamination bin (%)	CheckM contamination class	Strain heterogeneity bin (%)	Number contigs bin	Number genes bin	Length bin (Mb)	Number of BGCs present	
<i>C. angustus</i>	2	1	10.19	Partial	1.1	Low	0	316	1,506	1.23	0
		2	95.69	Near	0	Low	0	66	2,075	2.55	1
<i>C. atratus</i>	15	1	98.28	Near	2.77	Low	0	3,112	8,646	11.92	1
		2	96.55	Near	3.76	Low	0	2,760	8,783	10.34	1
		3	89.83	Substantial	28.68	Very high	31.94	323	2,509	2.22	0
		4	100	Near	0	Low	0	38	1,402	1.52	0
		5	100	Near	0	Low	0	176	2,053	2.21	0
		6	84.09	Substantial	15.99	Very high	0	729	5,094	4.73	0
		7	81.58	Substantial	2.07	Low	0	275	2,115	1.89	1
		8	96.55	Near	2.77	Low	0	85	2,053	2.14	0
		9	81.43	Substantial	25.49	Very high	65	455	3,733	3.5	2
		10	68.97	Moderate	31.56	Very high	4.44	910	4,862	4.23	2
		11	80.06	Substantial	32.37	Very high	38.46	798	4,765	4.39	2
		12	60.34	Moderate	21.22	Very high	18	523	2,575	2.21	1
		13	92.24	Near	0	Low	0	136	2,006	2.42	0
		14	82.52	Substantial	2.59	Low	0	425	2,743	2.77	0
		15	3.45	Partial	0	Low	0	194	423	0.542	0
<i>C. clypeatus</i>	6	1	98.58	Near	70.47	Very high	0	525	4,276	4.48	0
		2	94.83	Near	31.9	Very high	0	632	5,449	5.14	4
		3	99.5	Near	49.95	Very high	31.58	717	4,724	4.39	2
		4	98.28	Near	0	Low	0	143	1,824	2.03	0
		5	80.41	Substantial	1.72	Low	0	508	2,165	2.26	2
		6	98.28	Near	0.16	Low	0	805	3,940	4.53	0
<i>C. eduarduli</i>	8	1	96.55	Near	0	Low	0	448	2,497	2.8	1
		2	83.7	Substantial	0.86	Low	100	299	2,013	1.76	0

		3	98.28	Near	6.9	Medium	0	181	3,023	3.06	1
		4	86.41	Substantial	14.81	High	54.55	640	4,304	3.73	3
		5	98.28	Near	0	Low	0	71	2,152	2.61	0
		6	79	Substantial	27.9	Very high	52.63	588	4,057	3.9	3
		7	65.43	Moderate	20.85	Very high	53.33	526	2,793	2.28	0
		8	14.58	Partial	1.72	Low	100	736	4,462	3.88	6
<i>C. grandinosus</i>	12	1	98.28	Near	0.86	Low	100	79	1,817	1.99	0
		2	68.5	Moderate	0.31	Low	0	523	2,225	2.23	0
		3	96.55	Near	0	Low	0	90	2,357	2.46	1
		4	82.45	Substantial	24.48	Very high	10.53	871	5,889	5.54	8
		5	88.79	Substantial	57.76	Very high	12.5	672	4,083	4.01	2
		6	76.18	Substantial	33.62	Very high	8	490	3,821	3.76	1
		7	66.93	Moderate	39.66	Very high	20.69	525	3,808	3.56	4
		8	33.35	Partial	2.35	Low	50	222	1,103	0.94	0
		9	43.1	Partial	19.84	Very high	35.71	338	1,693	1.47	0
		10	83.62	Substantial	62.93	Very high	92.86	398	3,704	3.99	2
		11	73.34	Substantial	21.32	Very high	8	832	4,229	4.03	3
		12	41.5	Partial	19.75	Very high	28	635	4,010	3.33	3
<i>C. maculatus</i>	6	1	98.28	Near	3.45	Low	0	472	2,911	3.57	1
		2	96.55	Near	1.72	Low	0	290	3,308	3.44	1
		3	64.58	Moderate	32.05	Very high	30.77	1,300	6,793	6.04	4
		4	98.28	Near	0	Low	0	26	2,147	2.54	1
		5	64.5	Moderate	26.03	Very high	9.09	766	3,583	3.12	1
		6	92.63	Near	6.38	Medium	20	176	2,383	2.4	0
<i>C. minutus</i>	8	1	96.55	Near	0	Low	0	352	2,530	2.37	0
		2	98.28	Near	3.45	Low	0	390	2,671	3.01	1
		3	8.97	Partial	0	Low	0	328	1,740	1.44	3
		4	66.64	Moderate	0.16	Low	0	344	2,329	2.1	0
		5	96.55	Near	0	Low	0	74	2,393	2.5	1
		6	61.15	Moderate	0	Low	0	436	2,528	2.28	1
		7	97.41	Near	0	Low	0	116	2,170	2.51	1
		8	78.03	Substantial	40.99	Very high	9.26	1,072	5,110	4.38	2
<i>C. pallens</i>	10	1	94.83	Near	0	Low	0	253	1,840	1.97	0
		2	91.38	Near	0	Low	0	533	2,760	2.94	2
		3	96.55	Near	3.45	Low	0	162	2,736	2.79	1
		4	70.14	Substantial	20.69	Very high	22.22	993	4,940	4.99	5
		5	89.03	Substantial	53.94	Very high	23.73	1,300	6,840	6.16	8
		6	88.87	Substantial	28.45	Very high	4.76	472	3,815	3.76	3
		7	60.27	Moderate	10.61	High	36.36	451	2,463	2.12	2
		8	85.11	Substantial	0	Low	0	281	2,444	2.33	1

		9	69.09	Moderate	7.42	Medium	16.67	533	2,918	2.63	0
		10	83.62	Substantial	0.16	Low	0	322	2,522	2.84	0
<i>C. pellans</i>	7	1	100	Near	0	Low	0	403	2,303	2.37	1
		2	62.63	Moderate	20.73	Very high	5	918	2,904	3.1	5
		3	98.28	Near	0	Low	0	131	2,172	2.6	2
		4	97.41	Near	0	Low	0	36	4,504	4.12	1
		5	53.92	Moderate	16.09	Very high	11.54	825	2,801	2.58	2
		6	56.97	Moderate	25.86	Very high	33.33	845	4,280	3.92	1
		7	65.52	Moderate	3.45	Low	0	454	4,452	4.02	2
<i>C. persimilis</i>	6	1	98.28	Near	0.31	Low	0	475	5,164	5.07	1
		2	100	Near	13.95	High	0	623	3,314	3.54	7
		3	54.31	Moderate	18.55	Very high	7.14	722	2,148	2.52	0
		4	95.69	Near	0	Low	0	22	4,195	3.49	1
		5	69.76	Moderate	35.85	Very high	46.88	1,261	3,784	3.03	2
		6	78.68	Substantial	28.76	Very high	24.14	810	6,564	5.66	2
<i>C. persimplex</i>	11	1	98.28	Near	0	Low	0	121	1,836	1.94	0
		2	80.53	Substantial	24.87	Very high	0	1,052	3,906	4.1	0
		3	92.16	Near	5.17	Medium	0	1,293	4,905	5.9	0
		4	91.95	Near	0	Low	0	189	1,467	1.33	0
		5	100	Near	0	Low	0	192	2,020	2.11	0
		6	91.38	Near	0	Low	0	207	2,856	2.81	3
		7	100	Near	19.12	Very high	0	497	4,608	4.34	4
		8	81.03	Substantial	5.8	Medium	14.29	437	2,707	2.75	1
		9	81.35	Substantial	32.51	Very high	55.81	825	4,921	4.78	2
		10	54.39	Moderate	10.42	High	0	556	2,557	2.28	0
		11	84.48	Substantial	26.18	Very high	3.85	605	3,858	3.56	4
<i>C. pusillus</i>	5	1	54.94	Moderate	0	Low	0	281	3,727	3.59	0
		2	98.28	Near	6.9	Medium	0	429	2,381	2.73	4
		3	95.69	Near	0	Low	0	165	3,194	2.82	1
		4	78.45	Substantial	18.04	Very high	0	635	3,009	2.54	0
		5	62.54	Moderate	13.09	High	0	635	1,228	1.15	1
<i>C. rohweri</i>	22	1	98.28	Near	0	Low	0	29	1,583	1.64	0
		2	38.76	Partial	38.76	Very high	0	921	6,274	8.68	0
		3	96.55	Near	0	Low	0	47	1,729	1.82	0
		4	67.24	Moderate	12.07	High	4.35	3,378	18,614	31.63	2
		5	98.28	Near	10.97	High	51.06	147	3,052	3.22	6
		6	93.1	Near	0	Low	0	59	2,164	2.42	1
		7	87.3	Substantial	2.3	Low	50	108	2,148	2.24	1
		8	87.62	Substantial	76.65	Very high	29.17	313	4,518	4.46	1
		9	79.31	Substantial	53.45	Very high	20.45	424	5,174	5.36	4

		14	30.39	Partial	0.73	Low	0	5,345	12,816	17.88	0
<i>C. varians</i> PL005W	11	1	98.28	Near	2.38	Low	0	1,020	4,203	4.6	1
		2	93.1	Near	19.44	Very high	86.89	211	2,449	2.52	0
		3	85.58	Substantial	57.76	Very high	45.24	459	3,650	3.6	1
		4	98.28	Near	1.72	Low	100	253	2,957	2.9	4
		5	89.66	Substantial	0	Low	0	67	1,793	1.84	1
		6	88.97	Substantial	48.28	Very high	52.5	219	3,237	3.42	2
		7	75.08	Substantial	41.69	Very high	58.82	224	2,656	2.74	1
		8	98.28	Near	26.8	Very high	3.85	120	2,542	2.65	1
		9	79.31	Substantial	38.54	Very high	42.61	422	3,599	3.55	3
		10	87.93	Substantial	34.48	Very high	6.9	583	3,581	3.55	2
		11	46.08	Partial	31.03	Very high	64	480	3,651	3.53	2
<i>C. varians</i> PL010W	11	1	86.21	Substantial	0.34	Low	0	142	1,237	1.65	0
		2	53.29	Moderate	0	Low	0	136	1,842	1.7	0
		3	50	Moderate	0	Low	0	92	1,438	2.01	0
		4	42.95	Partial	0	Low	0	62	1,983	2.01	0
		5	97.41	Near	0	Low	0	45	2,475	2.04	0
		6	100	Near	1.72	Low	0	126	3,585	2.91	0
		7	98.28	Near	0	Low	0	112	1,586	2.3	2
		8	96.55	Near	0	Low	0	101	2,542	2.16	0
		9	94.83	Near	0	Low	0	84	2,079	1.99	1
		10	94.83	Near	0	Low	0	87	1,972	1.85	0
		11	87.93	Substantial	0	Low	0	63	1,845	1.49	1

Supplementary Table 4. BGCs types identified in the bacterial genomes. This table contains the number of BGCs detected in each bacteria isolated from *C. varians* and *C. rohweri*, as well as the length and number of genes composing these BGCs. The colors fill of the cell represent the type of BGCs.

Host	Symbiont	Type of cluster	Number of genes	Size of cluster (nt)
<i>C. varians</i>	Burkholderiales_Cv33a			
<i>C. varians</i>	Burkholderiales_Cv36_1	Terpene	20	21712
<i>C. varians</i>	Burkholderiales_Cv36_2	Arylpoyene	27	25774
<i>C. varians</i>	Burkholderiales_Cv44_1	T1PKS	42	47950
<i>C. varians</i>	Burkholderiales_Cv44_2	NRP	27	33193
<i>C. varians</i>	Burkholderiales_Cv44_3	Betalactone	16	15437
<i>C. varians</i>	Burkholderiales_Cv52_1	Terpene	22	21745
<i>C. varians</i>	Burkholderiales_Cv52_2	Resorcinol	31	41911
<i>C. varians</i>	Burkholderiales_Cv52_3	T1PKS	28	34252
<i>C. varians</i>	Burkholderiales_Cv52_4	Ectoïne	8	9669
<i>C. varians</i>	<i>Cephaloticoccus_canophilus</i> _Cv41			
<i>C. varians</i>	Rhizobiales JR021-5			
<i>C. varians</i>	<i>Ventrosimonas_gracilis</i> _Cv58	Arylpoyene	37	43580
<i>C. rohweri</i>	Burkholderiales_Cag20_1	T1PKS	36	36938
<i>C. rohweri</i>	Burkholderiales_Cag20_2	NRP	27	30210
<i>C. rohweri</i>	Burkholderiales_Cag20_3	Betalactone	20	22669
<i>C. rohweri</i>	Burkholderiales_Cag20_4	NRP	7	9722
<i>C. rohweri</i>	Burkholderiales_Cag20_5	NRP	34	39296
<i>C. rohweri</i>	Burkholderiales_Cag25_1	Betalactone	20	28119
<i>C. rohweri</i>	Burkholderiales_Cag25_2	T1PKS	34	47910
<i>C. rohweri</i>	Burkholderiales_Cag25_3	NRP	41	45968
<i>C. rohweri</i>	Burkholderiales_Cag25_4	NRP	32	27559
<i>C. rohweri</i>	<i>Cephaloticoccus_primus</i> _Cag34	Arylpoyene	47	42932
<i>C. rohweri</i>	<i>Ventrosimonas</i> _Cag26_1	Arylpoyene	31	26327
<i>C. rohweri</i>	<i>Ventrosimonas</i> _Cag26_2	NRP	33	39361
<i>C. rohweri</i>	<i>Ventrosimonas</i> _Cag27_1	NRP	34	36361

<i>C_rohweri</i>	<i>Ventosimonas_Cag27_2</i>	Arylpoyene	40	26376
<i>C_rohweri</i>	<i>Ventosimonas_Cag320_1</i>	NRP	18	28447
<i>C_rohweri</i>	<i>Ventosimonas_Cag320_2</i>	NRP	28	24149
<i>C_rohweri</i>	<i>Xanthomonadales_Cag60_1</i>	NRP	53	44406
<i>C_rohweri</i>	<i>Xanthomonadales_Cag60_2</i>	Ladderane	28	29577
<i>C_rohweri</i>	<i>Xanthomonadales_Cag60_3</i>	NRP	45	53610
<i>C_rohweri</i>	<i>Xanthomonadales_Cag60_4</i>	Siderophore	9	9390
<i>C_rohweri</i>	<i>Xanthomonadales_Cag60_5</i>	Arylpoyene	43	38815

Supplementary Table 5. BGCs types identified in the bacterial metagenomes. Types and origins of the bacterial BGCs in the metagenome analysis. The colors of the cell represent the type of BGCs. The code color is the same than in Supplementary Figure 22 and Supplementary Table 1.

Geography	Host species	Bins	Completeness bin (%)	CheckM completeness class	Contamination bin (%)	CheckM contamination class	Strain heterogeneity bin (%)	Number contigs bin	Number genes bin	Type gene cluster in bin	Length cluster (nt)	Number of genes in cluster	Symbiont order	Symbiont family	Symbiont genus
Peru	<i>C. angustus</i>	Bin 2	95.69	Near	0.00	Low	0.00	316	1,506.00	Arylpolyene	42,392.00	39	Burkholderiales	Enterobacteriaceae	Serratia
Brazil	<i>C. atratus</i>	Bin 9_2	81.43	Substantial	25.49	Very high	65.00	3,733.00	455	Arylpolyene	23,687.00	31	Burkholderiales	Burkholderiaceae	Ralstonia
Brazil	<i>C. atratus</i>	Bin 2	96.55	Near	3.76	Low	0.00	2,760.00	8,783.00	LAP	6,492.00	9	Bacillales	Listeriaceae	Listeria
Brazil	<i>C. atratus</i>	Bin 11_4	80.06	Substantial	32.37	Very high	38.46	4,765.00	798	Arylpolyene	6,235.00	6	Burkholderiales	Comamonadaceae	Verminephrobacter
Brazil	<i>C. atratus</i>	Bin 12	60.34	Moderate	21.22	Very high	18.00	2,575.00	523	Arylpolyene	24,572.00	35	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. atratus</i>	Bin 1	98.28	Near	2.77	Low	0.00	8,646.00	3,112.00	Bacteriocin	8,308.00	15	Acholeplasmatales	Acholeplasmataceae	Acholeplasma
Brazil	<i>C. atratus</i>	Bin 10_2	68.97	Moderate	31.56	Very high	4.44	4,862.00	910	Hserlactone	6,118.00	8	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. atratus</i>	Bin 11_6	80.06	Substantial	32.37	Very high	38.46	4,765.00	798	Arylpolyene-Ladderane	6,994.00	8	Enterobacteriales	Enterobacteriaceae	Proteus
Brazil	<i>C. atratus</i>	Bin 10_1	68.97	Moderate	31.56	Very high	4.44	4,862.00	910	NRP	6,919.00	7	Burkholderiales	Burkholderiaceae	Burkholderia
Brazil	<i>C. atratus</i>	Bin 7	81.58	Substantial	2.07	Low	0.00	2,115.00	275	Terpene	12,499.00	12	Burkholderiales	Alcaligenaceae	Bordetella
Brazil	<i>C. atratus</i>	Bin 9_1	81.43	Substantial	25.49	Very high	65.00	3,733.00	455	Terpene	12,633.00	12	Burkholderiales	Comamonadaceae	Acidovorax
Brazil	<i>C. clypeatus</i>	Bin 2_3	94.83	Near	31.90	Very high	0.00	632	5,449.00	Arylpolyene	32,535.00	35	Burkholderiales	Comamonadaceae	Melaminivora
Brazil	<i>C. clypeatus</i>	Bin 2_4	94.83	Near	31.90	Very high	0.00	632	5,449.00	LAP	9,615.00	10	Xanthomonadales	Xanthomonadaceae	Lysobacter
Brazil	<i>C. clypeatus</i>	Bin 2_1	94.83	Near	31.90	Very high	0.00	632	5,449.00	Betalactone	25,272.00	28	Burkholderiales	Burkholderiaceae	Cupriavidus
Brazil	<i>C. clypeatus</i>	Bin 3_1	99.50	Near	49.95	Very high	31.58	717	4,724.00	Ladderane	19,231.00	20	Enterobacteriales	Enterobacteriaceae	Serratia

Brazil	<i>C. clypeatus</i>	Bin 5_2	80.41	Substantial	1.72	Low	0.00	508	2,165.00	Lanthipeptide	6,072.00	4	Xanthomonadales	Xanthomonadaceae	Lysobacter
Brazil	<i>C. clypeatus</i>	Bin 3_2	99.50	Near	49.95	Very high	31.58	717	4,724.00	NRP	8,602.00	8	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
Brazil	<i>C. clypeatus</i>	Bin 5_1	80.41	Substantial	1.72	Low	0.00	508	2,165.00	Siderophore	5,204.00	3	Rhizobiales	Rhizobiaceae	Sinorhizobium
Brazil	<i>C. clypeatus</i>	Bin 2_2	94.83	Near	31.90	Very high	0.00	632	5,449.00	Terpene	18,172.00	23	Enterobacteriales	Enterobacteriaceae	Erwinia
Brazil	<i>C. eduarduli</i>	Bin 8_1	14.58	Partial	1.72	Low	100.00	736	4,462.00	Acyl amino acids	11,694.00	11	Rhizobiales	Methylobacteriaceae	Methylobacterium
Brazil	<i>C. eduarduli</i>	Bin 3	98.28	Near	6.90	Medium	0.00	181	3,023.00	Arylpolyene	6,560.00	6	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. eduarduli</i>	Bin 6_2	79.00	Substantial	27.90	Very high	52.63	588	4,057.00	Arylpolyene	5,347.00	9	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. eduarduli</i>	Bin 8_3	14.58	Partial	1.72	Low	100.00	736	4,462.00	Butyrolactone	7,902.00	10	Burkholderiales	Alcaligenaceae	Alcaligenes
Brazil	<i>C. eduarduli</i>	Bin 8_5	14.58	Partial	1.72	Low	100.00	736	4,462.00	Butyrolactone	5,545.00	6	Burkholderiales	Alcaligenaceae	Alcaligenes
Brazil	<i>C. eduarduli</i>	Bin 1	96.55	Near	0.00	Low	0.00	448	2,497.00	Furan	14,566.00	17	Flavobacteriales	Flavobacteriaceae	Elizabethkingia
Brazil	<i>C. eduarduli</i>	Bin 6_1	79.00	Substantial	27.90	Very high	52.63	588	4,057.00	Ladderane	15,012.00	13	Enterobacteriales	Enterobacteriaceae	Serratia
Brazil	<i>C. eduarduli</i>	Bin 8_2	14.58	Partial	1.72	Low	100.00	736	4,462.00	NRP	16,384.00	6	Enterobacteriales	Yersiniaceae	Yersinia
Brazil	<i>C. eduarduli</i>	Bin 4_2	86.41	Substantial	14.81	High	54.55	640	4,304.00	NRP	21,928.00	21	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. eduarduli</i>	Bin 8_4	14.58	Partial	1.72	Low	100.00	736	4,462.00	NRP	5,624.00	6	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. eduarduli</i>	Bin 6_3	79.00	Substantial	27.90	Very high	52.63	588	4,057.00	ResorcinoI	20,058.00	33	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
Brazil	<i>C. eduarduli</i>	Bin 8_6	14.58	Partial	1.72	Low	100.00	736	4,462.00	Siderophore	8,042.00	7	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. eduarduli</i>	Bin 4_1	86.41	Substantial	14.81	High	54.55	640	4,304.00	T1PK	15,929.00	15	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. eduarduli</i>	Bin 4_4	86.41	Substantial	14.81	High	54.55	640	4,304.00	Terpene	10,275.00	14	Rhizobiales	Brucellaceae	Ochrobactrum
Brazil	<i>C. grandinosus</i>	Bin 4_2	82.45	Substantial	24.48	Very high	10.53	425	1,968.00	Arylpolyene	6,092.00	8	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. grandinosus</i>	Bin 10_1	83.62	Substantial	62.93	Very high	92.86	231	2,414.00	Arylpolyene	19,882.00	20	Burkholderiales	Alcaligenaceae	Achromobacter

Brazil	<i>C. grandinosus</i>	Bin 10_2	83.62	Substantial	62.93	Very high	92.86	231	2,414.00	Arylpolyene	23,001.00	31	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. grandinosus</i>	Bin 4_4	82.45	Substantial	24.48	Very high	10.53	425	1,968.00	Arylpolyene	10,418.00	8	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. grandinosus</i>	Bin 11_1	73.34	Substantial	62.93	Very high	92.86	414	2,158.00	Arylpolyene	9,981.00	10	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. grandinosus</i>	Bin 11_3	73.34	Substantial	62.93	Very high	92.86	414	2,158.00	Arylpolyene	6,212.00	9	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. grandinosus</i>	Bin 11_2	73.34	Substantial	62.93	Very high	92.86	173	778	Bacteriocin	6,103.00	6	Xanthomonadales	Xanthomonadaceae	Xanthomonas
Brazil	<i>C. grandinosus</i>	Bin 6_1	76.18	Substantial	33.62	Very high	8.00	234	2,272.00	Betalactone	21,831.00	23	Burkholderiales	Alcaligenaceae	Bordetella
Brazil	<i>C. grandinosus</i>	Bin 7_3	66.93	Moderate	39.66	Very high	20.69	277	2,410.00	Betalactone	13,647.00	15	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. grandinosus</i>	Bin 4_6	82.45	Substantial	24.48	Very high	10.53	425	1,968.00	NRP	15,960.00	11	Bacillales	Paenibacillaceae	Paenibacillus
Brazil	<i>C. grandinosus</i>	Bin 7_4	66.93	Moderate	39.66	Very high	20.69	277	2,410.00	NRP	9,572.00	8	Burkholderiales	Alcaligenaceae	Alcaligenes
Brazil	<i>C. grandinosus</i>	Bin 12_4	41.50	Partial	19.75	Very high	28.00	311	1,926.00	NRP	7,491.00	3	Burkholderiales	Burkholderiaceae	Burkholderia
Brazil	<i>C. grandinosus</i>	Bin 4_3	82.45	Substantial	24.48	Very high	10.53	425	1,968.00	NRP	19,174.00	7	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. grandinosus</i>	Bin 4_9	82.45	Substantial	24.48	Very high	10.53	425	1,968.00	NRP	7,884.00	4	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. grandinosus</i>	Bin 12_1	41.50	Partial	19.75	Very high	28.00	311	1,926.00	NRP	11,820.00	11	Rhizobiales	Brucellaceae	Brucella
Brazil	<i>C. grandinosus</i>	Bin 4_1	82.45	Substantial	24.48	Very high	10.53	425	1,968.00	Phenazine	5,441.00	5	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. grandinosus</i>	Bin 7_2	66.93	Moderate	39.66	Very high	20.69	248	1,398.00	Resorcinol	11,807.00	11	Burkholderiales	Alcaligenaceae	Achromobacter

Brazil	<i>C. grandinosus</i>	Bin 3	96.55	Near	0.00	Low	0.00	90	2,357.00	Resorcino I	32,027.00	33	Xanthomonadales	Xanthomonadaceae	Xanthomonas
Brazil	<i>C. grandinosus</i>	Bin 5_1	88.79	Substantial	57.76	Very high	12.50	267	2,015.00	T1PK	9,635.00	7	Burkholderiales	Alcaligenaceae	Bordetella
Brazil	<i>C. grandinosus</i>	Bin 12_2	41.50	Partial	19.75	Very high	28.00	324	2,084.00	T1PK	30,486.00	26	Burkholderiales	Alcaligenaceae	Bordetella
Brazil	<i>C. grandinosus</i>	Bin 5_5	88.79	Substantial	57.76	Very high	12.50	267	2,015.00	Terpene	12,750.00	16	Burkholderiales	Comamonadaceae	Comamonas
Brazil	<i>C. grandinosus</i>	Bin 7_5	66.93	Moderate	39.66	Very high	20.69	248	1,398.00	Terpene	7,469.00	8	Burkholderiales	Alcaligenaceae	Bordetella
Brazil	<i>C. grandinosus</i>	Bin 4_12	82.45	Substantial	24.48	Very high	10.53	425	1,968.00	LAP	9,681.00	12	Xanthomonadales	Xanthomonadaceae	Xanthomonas
Brazil	<i>C. grandinosus</i>	Bin 4_10	82.45	Substantial	24.48	Very high	10.53	425	1,968.00	Terpene	6,123.00	5	Pseudomonadales	Moraxellaceae	Acinetobacter
Brazil	<i>C. maculatus</i>	Bin 4	98.28	Near	0.00	Low	0.00	26	2,147.00	Arylpolyene	42,670.00	38	Verrucomicrobiales	Opiritaceae	Opiritus
Brazil	<i>C. maculatus</i>	Bin 3_6	64.58	Moderate	32.05	Very high	30.77	1,300.00	6,793.00	Arylpolyene	5,021.00	6	Xanthomonadales	Xanthomonadaceae	Xanthomonas
Brazil	<i>C. maculatus</i>	Bin 3_5	64.58	Moderate	32.05	Very high	30.77	1,300.00	6,793.00	NRP	18,724.00	7	Bacillales	Paenibacillaceae	Brevibacillus
Brazil	<i>C. maculatus</i>	Bin 5_3	64.50	Moderate	26.03	Very high	9.09	766	3,583.00	NRP	6,888.00	2	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. maculatus</i>	Bin 3_1	64.58	Moderate	32.05	Very high	30.77	1,300.00	6,793.00	NRP-T1PK	21,264.00	11	Burkholderiales	Burkholderiaceae	Ralstonia/Burkholderia
Brazil	<i>C. maculatus</i>	Bin 2	96.55	Near	1.72	Low	0.00	290	3,308.00	Resorcino I	32,905.00	36	Burkholderiales	Alcaligenaceae	Alcaligenes
Brazil	<i>C. maculatus</i>	Bin 3_2	64.58	Moderate	32.05	Very high	30.77	1,300.00	6,793.00	LAP	14,213.00	13	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
Brazil	<i>C. maculatus</i>	Bin 1	98.28	Near	3.45	Low	0.00	472	2,911.00	Terpene	20,923.00	17	Rhizobiales	Bartonellaceae	Bartonella
Brazil	<i>C. minutus</i>	Bin 6_1	61.15	Moderate	0.00	Low	0.00	436	2,528.00	Arylpolyene	14,647.00	15	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. minutus</i>	Bin 7	97.41	Near	0.00	Low	0.00	116	2,170.00	Arylpolyene	25,728.00	39	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. minutus</i>	Bin 5	96.55	Near	0.00	Low	0.00	74	2,393.00	Arylpolyene	43,072.00	57	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
Brazil	<i>C. minutus</i>	Bin 8_4	78.03	Substantial	40.99	Very high	9.26	1,072.00	5,110.00	Butyrolactone	9,664.00	13	Pseudomonadales	Pseudomonadaceae	Pseudomonas

Brazil	<i>C. minutus</i>	Bin 3_2	8.97	Partial	0.00	Low	0.00	328	1,740.00	Lanthipeptide	5,678.00	5	Xanthomonadales	Xanthomonadaceae	Stenotropomonas
Brazil	<i>C. minutus</i>	Bin 3_3	8.97	Partial	0.00	Low	0.00	328	1,740.00	LAP	8,413.00	9	Enterobacteriales	Enterobacteriaceae	Serratia
Brazil	<i>C. minutus</i>	Bin 2	98.28	Near	3.45	Low	0.00	390	2,671.00	NRP	7,176.00	5	Desulfovibrionales	Desulfovibrionaceae	Desulfovibrio
Brazil	<i>C. minutus</i>	Bin 8_2	78.03	Substantial	40.99	Very high	9.26	1,072.00	5,110.00	NRP	7,334.00	6	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. minutus</i>	Bin 3_1	8.97	Partial	0.00	Low	0.00	328	1,740.00	NRP	9,249.00	8	Rhizobiales	Brucellaceae	Brucella
Brazil	<i>C. pallens</i>	Bin 5_10	89.03	Substantial	53.94	Very high	23.73	1,300.00	6,840.00	Arylpolyene	7,002.00	8	Burkholderiales	Comamonadaceae	Verminephrobacter
Brazil	<i>C. pallens</i>	Bin 5_11	89.03	Substantial	53.94	Very high	23.73	1,300.00	6,840.00	Arylpolyene	5,011.00	4	Burkholderiales	Comamonadaceae	Verminephrobacter
Brazil	<i>C. pallens</i>	Bin 2_2	91.38	Near	0.00	Low	0.00	533	2,760.00	Arylpolyene	9,255.00	13	Campylobacterales	Campylobacteraceae	Arcobacter
Brazil	<i>C. pallens</i>	Bin 6_4	88.87	Substantial	28.45	Very high	4.76	472	3,815.00	Ectoine	7,791.00	7	Rhizobiales	Brucellaceae	Ochrobactrum
Brazil	<i>C. pallens</i>	Bin 4_1	70.14	Substantial	20.69	Very high	22.22	993	4,940.00	Lanthipeptide	19,732.00	25	Sphingobacteriales	Sphingobacteriaceae	Sphingobacterium
Brazil	<i>C. pallens</i>	Bin 5_2	89.03	Substantial	53.94	Very high	23.73	1,300.00	6,840.00	NRP	21,156.00	17	Burkholderiales	Burkholderiaceae	Burkholderia
Brazil	<i>C. pallens</i>	Bin 5_7	89.03	Substantial	53.94	Very high	23.73	1,300.00	6,840.00	NRP	14,425.00	6	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. pallens</i>	Bin 7_1	60.27	Moderate	10.61	High	36.36	451	2,463.00	NRP	6,271.00	7	Burkholderiales	Alcaligenaceae	Alcaligenes
Brazil	<i>C. pallens</i>	Bin 5_5	89.03	Substantial	53.94	Very high	23.73	1,300.00	6,840.00	NRP	8,496.00	3	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. pallens</i>	Bin 5_3	89.03	Substantial	53.94	Very high	23.73	1,300.00	6,840.00	NRP	19,586.00	17	Rhizobiales	Brucellaceae	Brucella
Brazil	<i>C. pallens</i>	Bin 4_4	70.14	Substantial	20.69	Very high	22.22	993	4,940.00	Phenazine	11,250.00	11	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. pallens</i>	Bin 4_5	70.14	Substantial	20.69	Very high	22.22	993	4,940.00	Phenazine	5,071.00	7	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. pallens</i>	Bin 5_6	89.03	Substantial	53.94	Very high	23.73	1,300.00	6,840.00	Resorcino	10,397.00	9	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. pallens</i>	Bin 2_1	91.38	Near	0.00	Low	0.00	533	2,760.00	Resorcino	9,029.00	11	Campylobacterales	Campylobacteraceae	Arcobacter
Brazil	<i>C. pallens</i>	Bin 3	96.55	Near	3.45	Low	0.00	162	2,736.00	Resorcino	28,165.00	28	Xanthomonadales	Xanthomonadaceae	Pseudoxanthomonas
Brazil	<i>C. pallens</i>	Bin 4_6	70.14	Substantial	20.69	Very high	22.22	993	4,940.00	Siderophore	8,875.00	9	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. pallens</i>	Bin 5_4	89.03	Substantial	53.94	Very high	23.73	1,300.00	6,840.00	T1PK	9,854.00	2	Burkholderiales	Burkholderiaceae	Ralstonia

Brazil	<i>C. pallens</i>	Bin 7_2	60.27	Moderate	10.61	High	36.36	451	2,463.00	T1PK	6,178.00	2	Burkholderiales	Alcaligenaceae	Bordetella
Brazil	<i>C. pallens</i>	Bin 6_1	88.87	Substantial	23.45	Very high	4.76	472	3,815.00	Terpene	10,683.00	12	Burkholderiales	Alcaligenaceae	Bordetella
Brazil	<i>C. pallens</i>	Bin 6_2	88.87	Substantial	23.45	Very high	4.76	472	3,815.00	Terpene	7,225.00	9	Burkholderiales	Alcaligenaceae	Bordetella
Brazil	<i>C. pallens</i>	Bin 8	85.11	Substantial	0.00	Low	0.00	281	2,444.00	Terpene	13,242.00	13	Vibrionales	Vibrionaceae	Vibrio
Brazil	<i>C. pallens</i>	Bin 4_2	70.14	Substantial	20.69	Very high	22.22	993	4,940.00	Thiopeptide	6,791.00	7	Bacillales	Bacillaceae	Bacillus
Brazil	<i>C. pellans</i>	Bin 6_2	56.97	Moderate	25.86	Very high	33.33	845	4,452.00	Arylpolyene	6,881.00	9	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. pellans</i>	Bin 1	100.00	Near	0.00	Low	0.00	403	2,904.00	Arylpolyene-Resorcino	45,905.00	49	Campylobacterales	Campylobacteraceae	Arcobacter
Brazil	<i>C. pellans</i>	Bin 2_5	62.63	Moderate	20.73	Very high	5.00	918	4,504.00	Butyrolactone	5,231.00	7	Actinomycetales	Streptomycetales	Streptomyces
Brazil	<i>C. pellans</i>	Bin 2_4	62.63	Moderate	20.73	Very high	5.00	918	4,504.00	Butyrolactone	5,550.00	8	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. pellans</i>	Bin 7_1	65.52	Moderate	3.45	Low	0.00	454	2,801.00	Ectoine	6,917.00	8	Rhizobiales	Brucellaceae	Ochrobactrum
Brazil	<i>C. pellans</i>	Bin 5_2	53.92	Moderate	16.09	Very high	11.54	825	4,280.00	NRP	12,865.00	4	Neisseriales	Neisseriaceae	Chromobacterium
Brazil	<i>C. pellans</i>	Bin 2_1	62.63	Moderate	20.73	Very high	5.00	918	4,504.00	NRP	5,644.00	4	Pseudomonadales	Moraxellaceae	Acinetobacter
Brazil	<i>C. pellans</i>	Bin 5_1	53.92	Moderate	16.09	Very high	11.54	825	4,280.00	NRP	13,491.00	12	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. pellans</i>	Bin 4	97.41	Near	0.00	Low	0.00	36	2,172.00	Phenazine	20,495.00	19	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. pellans</i>	Bin 3_1	98.28	Near	0.00	Low	0.00	131	2,303.00	Resorcino	8,811.00	10	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
Brazil	<i>C. pellans</i>	Bin 3_2	98.28	Near	0.00	Low	0.00	131	2,303.00	Resorcino	9,855.00	9	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
Brazil	<i>C. pellans</i>	Bin 2_7	62.63	Moderate	20.73	Very high	5.00	918	4,504.00	Siderophore	9,526.00	11	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. pellans</i>	Bin 7_2	65.52	Moderate	3.45	Low	0.00	454	2,801.00	Terpene	6,474.00	8	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. pellans</i>	Bin 2_3	62.63	Moderate	20.73	Very high	5.00	918	4,504.00	Thiopeptide	11,018.00	8	Bacillales	Bacillaceae	Bacillus
Brazil	<i>C. persimilis</i>	Bin 4	95.69	Near	0.00	Low	0.00	22	2.148	Arylpolyene	42,873.00	42	Opitutales	Opitutaceae	Cephalotococcus
Brazil	<i>C. persimilis</i>	Bin 5_1	69.76	Moderate	35.85	Very high	46.88	1,261.00	6.564	Arylpolyene	9,360.00	12	Pseudomonadales	Pseudomonadaceae	Pseudomonas

Brazil	<i>C. persimilis</i>	Bin 1	98.28	Near	0.31	Low	0.00	475	3.314	Arylpolyene-Betalactone-Resorcino l	63,930.00	68	Campylobacterales	Campylobacteraceae	Arcobacter
Brazil	<i>C. persimilis</i>	Bin 6_2	76.68	Substantial	28.76	Very high	24.14	810	4.195	Ectoioine	10,333.00	9	Burkholderiales	Alcaligenaceae	Bordetella
Brazil	<i>C. persimilis</i>	Bin 2_5	100.00	Near	13.95	High	0.00	623	5.164	Lanthipeptide	14,029.00	14	Flavobacteriales	Flavobacteriaceae	Kordia
Brazil	<i>C. persimilis</i>	Bin 5_3	69.76	Moderate	35.85	Very high	46.88	1,261.00	6.564	NRP	10,502.00	8	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. persimilis</i>	Bin 2_1	100.00	Near	13.95	High	0.00	623	5.164	LAP	11,913.00	13	Xanthomonadales	Xanthomonadaceae	Xanthomonas
Brazil	<i>C. persimilis</i>	Bin 2_2	100.00	Near	13.95	High	0.00	623	5.164	NRP	14,699.00	13	Nostocales	Nostocaceae	Cylindrospermum
Brazil	<i>C. persimilis</i>	Bin 2_6	100.00	Near	13.95	High	0.00	623	5.164	NRP	11,940.00	5	Oscillatoriales	Microcoleaceae	Microcoleus
Brazil	<i>C. persimilis</i>	Bin 2_3	100.00	Near	13.95	High	0.00	623	5.164	NRP	10,617.00	4	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. persimilis</i>	Bin 2_4	100.00	Near	13.95	High	0.00	623	5.164	Resorcino l	22,365.00	21	Xanthomonadales	Xanthomonadaceae	Xanthomonas
Brazil	<i>C. persimilis</i>	Bin 2_8	100.00	Near	13.95	High	0.00	623	5.164	Siderophore	9,495.00	8	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. persimilis</i>	Bin 6_1	78.68	Substantial	28.76	Very high	24.14	810	4.195	T1PK	6,869.00	7	Burkholderiales	Alcaligenaceae	Achromobacter
Peru	<i>C. persimplex</i>	Bin 6_1	91.38	Near	0.00	Low	0.00	207	2,856.00	Arylpolyene	25,746.00	42	Burkholderiales	Comamonadaceae	Alicyclophilus
Peru	<i>C. persimplex</i>	Bin 8_2	81.03	Substantial	5.80	Medium	14.29	437	2,707.00	Arylpolyene	5,178.00	6	Burkholderiales	Alcaligenaceae	Achromobacter
Peru	<i>C. persimplex</i>	Bin 11_4	84.48	Substantial	26.18	Very high	3.85	605	3.56	Arylpolyene	6,553.00	10	Burkholderiales	Alcaligenaceae	Achromobacter
Peru	<i>C. persimplex</i>	Bin 7_1	100.00	Near	19.12	Very high	0.00	497	4,608.00	Arylpolyene	35,413.00	36	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Peru	<i>C. persimplex</i>	Bin 9_1	81.35	Substantial	32.51	Very high	55.81	825	4,921.00	Arylpolyene	22,142.00	29	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
Peru	<i>C. persimplex</i>	Bin 7_2	100.00	Near	19.12	Very high	0.00	497	4,608.00	Arylpolyene-Ladderane	45,003.00	40	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Peru	<i>C. persimplex</i>	Bin 9_5	81.35	Substantial	32.51	Very high	55.81	825	4,921.00	Arylpolyene-Ladderane	5,725.00	4	Pseudomonadales	Pseudomonadaceae	Pseudomonas

Peru	<i>C. persimplex</i>	Bin 7_4	100.00	Near	19.12	Very high	0.00	497	4,608.00	Bacteriocin	7,647.00	7	Flavobacteriales	Flavobacteriaceae	Elizabethkingia
Peru	<i>C. persimplex</i>	Bin 7_5	100.00	Near	19.12	Very high	0.00	497	4,608.00	NRP	5,145.00	4	Desulfovibrionales	Desulfovibrionaceae	Desulfovibrio
Peru	<i>C. persimplex</i>	Bin 11_3	84.48	Substantial	26.18	Very high	3.85	605	3.56	NRP-T1PK	16,353.00	4	Enterobacteriales	Pectobacteriaceae	Brenneria
Peru	<i>C. persimplex</i>	Bin 11_1	84.48	Substantial	26.18	Very high	3.85	605	3.56	ResorcinoI	8,229.00	9	Burkholderiales	Alcaligenaceae	Achromobacter
Peru	<i>C. persimplex</i>	Bin 6_2	91.38	Near	0.00	Low	0.00	207	2,856.00	Terpene	11,095.00	10	Burkholderiales	Comamonadaceae	Alicyclophilus
Peru	<i>C. persimplex</i>	Bin 6_3	91.38	Near	0.00	Low	0.00	207	2,856.00	Terpene	10,829.00	10	Burkholderiales	Comamonadaceae	Acidovorax
Peru	<i>C. persimplex</i>	Bin 11_2	84.48	Substantial	26.18	Very high	3.85	605	3.56	Terpene	17,690.00	19	Burkholderiales	Alcaligenaceae	Achromobacter
Brazil	<i>C. pusillus</i>	Bin 2_1	98.28	Near	6.90	Medium	0.00	429	3,727.00	Arylpolyene	20,582.00	25	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Brazil	<i>C. pusillus</i>	Bin 3	95.69	Near	0.00	Low	0.00	165	2,381.00	Arylpolyene	24,174.00	23	Verrucomicrobiales	Opitutaceae	Lacunisphaera
Brazil	<i>C. pusillus</i>	Bin 2_2	98.28	Near	6.90	Medium	0.00	429	3,727.00	Butyrolactone	5,470.00	7	Rhodocyclales	Zoogloeaceae	Thauera
Brazil	<i>C. pusillus</i>	Bin 2_3	98.28	Near	6.90	Medium	0.00	429	3,727.00	Furan	5,003.00	9	Sphingobacteriales	Sphingobacteriaceae	Sphingobacterium
Brazil	<i>C. pusillus</i>	Bin 2_4	98.28	Near	6.90	Medium	0.00	429	3,727.00	Lanthipeptide	8,874.00	6	Sphingobacteriales	Sphingobacteriaceae	Pedobacter
Brazil	<i>C. pusillus</i>	Bin 5_1	62.54	Moderate	13.09	High	0.00	635	3,009.00	NRP	19,615.00	8	Neisseriales	Neisseriaceae	Chromobacterium
USA	<i>C. rohweri</i>	Bin 5_5	98.28	Near	10.97	High	51.06	3,052.00	147	Arylpolyene	8,431.00	8	Pseudomonadales	Ventrosimonadaceae	Ventrosimonas
USA	<i>C. rohweri</i>	Bin 8	87.62	Substantial	76.65	Very high	29.17	4,518.00	313	Arylpolyene	24,839.00	31	Pseudomonadales	Pseudomonadaceae	Pseudomonas
USA	<i>C. rohweri</i>	Bin 4_2	67.24	Moderate	12.07	High	4.35	18,614	3,378.00	Arylpolyene	6,739.00	7	Rhizobiales	Bartonellaceae	Bartonella
USA	<i>C. rohweri</i>	Bin 5_2	98.28	Near	10.97	High	51.06	3,052.00	147	Arylpolyene	29,483.00	27	Rhizobiales	Bartonellaceae	Bartonella
USA	<i>C. rohweri</i>	Bin 13	93.10	Near	3.45	Low	50.00	2,033.00	88	Betalactone	12,898.00	17	Burkholderiales	Alcaligenaceae	Achromobacter
USA	<i>C. rohweri</i>	Bin 15_3	69.83	Moderate	21.47	Very high	68.42	2,775.00	131	Betalactone	17,844.00	13	Burkholderiales	Alcaligenaceae	Achromobacter
USA	<i>C. rohweri</i>	Bin 12_3	85.06	Substantial	22.88	Very high	39.13	3,421.00	202	Betalactone	16,867.00	19	Pseudomonadales	Pseudomonadaceae	Pseudomonas
USA	<i>C. rohweri</i>	Bin 15_2	69.83	Moderate	21.47	Very high	68.42	2,775.00	131	Betalactone	22,283.00	21	Pseudomonadales	Pseudomonadaceae	Pseudomonas
USA	<i>C. rohweri</i>	Bin 9_2	79.31	Substantial	53.45	Very high	20.45	5,174.00	424	Ladderane	13,287.00	9	Xanthomonadales	Xanthomonadaceae	Xanthomonas

USA	<i>C. rohweri</i>	Bin 9_3	79.31	Substantial	53.45	Very high	20.45	5,174.00	424	Ladderane	18,057.00	18	Enterobacteriales	Enterobacteriaceae	Serratia
USA	<i>C. rohweri</i>	Bin 5_6	98.28	Near	10.97	High	51.06	3,052.00	147	Lanthipeptide	6,741.00	6	Xanthomonadales	Xanthomonadales	Stenotrophomonas
USA	<i>C. rohweri</i>	Bin 5_3	98.28	Near	10.97	High	51.06	3,052.00	147	NRP	11,362.00	8	Burkholderiales	Burkholderiaceae	Burkholderia
USA	<i>C. rohweri</i>	Bin 12_1	85.06	Substantial	22.88	Very high	39.13	3,421.00	202	NRP	23,818.00	25	Burkholderiales	Burkholderiaceae	Burkholderia
USA	<i>C. rohweri</i>	Bin 14_1	63.79	Moderate	17.82	Very high	91.89	2,045.00	121	NRP	16,730.00	11	Burkholderiales	Alcaligenaceae	Alcaligenes
USA	<i>C. rohweri</i>	Bin 14_2	63.79	Moderate	17.82	Very high	91.89	2,045.00	121	NRP	15,915.00	13	Burkholderiales	Alcaligenaceae	Alcaligenes
USA	<i>C. rohweri</i>	Bin 20	22.42	Partial	8.31	Medium	0.00	46,227.00	14,539.00	NRP	8,155.00	5	Burkholderiales	Alcaligenaceae	Alcaligenes
USA	<i>C. rohweri</i>	Bin 12_2	85.06	Substantial	22.88	Very high	39.13	3,421.00	202	NRP	16,717.00	7	Pseudomonadales	Ventrosimonadales	Ventrosimonas
USA	<i>C. rohweri</i>	Bin 10_2	98.28	Near	0.00	Low	0.00	2,295.00	77	ResorcinoI	23,616.00	23	Burkholderiales	Alcaligenaceae	Bordetella
USA	<i>C. rohweri</i>	Bin 9_4	79.31	Substantial	53.45	Very high	20.45	5,174.00	424	ResorcinoI	12,547.00	12	Xanthomonadales	Xanthomonadales	Xanthomonas
USA	<i>C. rohweri</i>	Bin 5_4	98.28	Near	10.97	High	51.06	3,052.00	147	Siderophore	10,804.00	9	Burkholderiales	Alcaligenaceae	Achromobacter
USA	<i>C. rohweri</i>	Bin 10_3	98.28	Near	0.00	Low	0.00	2,295.00	77	T1PK	15,520.00	9	Burkholderiales	Alcaligenaceae	Bordetella
USA	<i>C. rohweri</i>	Bin 15_1	69.83	Moderate	21.47	Very high	68.42	2,775.00	131	T1PK	32,313.00	27	Burkholderiales	Alcaligenaceae	Bordetella
USA	<i>C. rohweri</i>	Bin 15_4	69.83	Moderate	21.47	Very high	68.42	2,775.00	131	T1PK	16,292.00	8	Burkholderiales	Alcaligenaceae	Bordetella
USA	<i>C. rohweri</i>	Bin 9_1	79.31	Substantial	53.45	Very high	20.45	5,174.00	424	Terpene	16,223.00	17	Burkholderiales	Alcaligenaceae	Bordetella
USA	<i>C. rohweri</i>	Bin 10_1	98.28	Near	0.00	Low	0.00	2,295.00	77	Terpene	20,894.00	22	Burkholderiales	Alcaligenaceae	Bordetella
USA	<i>C. rohweri</i>	Bin 11	100.00	Near	0.00	Low	0.00	2,004.00	31	Terpene	15,849.00	13	Burkholderiales	Alcaligenaceae	Bordetella
USA	<i>C. rohweri</i>	Bin 16	88.09	Substantial	0.00	Low	0.00	2,073.00	90	Terpene	12,767.00	10	Burkholderiales	Alcaligenaceae	Bordetella
USA	<i>C. rohweri</i>	Bin 4_1	67.24	Moderate	12.07	High	4.35	18,614.00	3,378.00	Terpene	10,327.00	9	Rhizobiales	Bartonellaceae	Bartonella
USA	<i>C. rohweri</i>	Bin 5_1	98.28	Near	10.97	High	51.06	3,052.00	147	Terpene	12,715.00	12	Rhizobiales	Brucellaceae	Ochrobactrum
USA	<i>C. rohweri</i>	Bin 6	93.10	Near	0.00	Low	0.00	2,164.00	59	Terpene	18,554.00	18	Rhizobiales	Rhizobiaceae	Rhizobium
USA	<i>C. rohweri</i>	Bin 7	87.30	Substantial	2.30	Low	50.00	2,148.00	108	Terpene	19,007.00	18	Rhizobiales	Bartonellaceae	Bartonella

Peru	<i>C. similimus</i>	Bin 5_3	60.27	Moderate	21.00	Very high	12.50	803	4,703.00	Arylpolyene	5,313.00	5	Neisseriales	Neisseriaceae	Neisseria
Peru	<i>C. similimus</i>	Bin 6_6	75.39	Substantial	24.50	Very high	7.32	826	4,667.00	Arylpolyene	7,127.00	8	Burkholderiales	Alcaligenaceae	Achromobacter
Peru	<i>C. similimus</i>	Bin 5_1	60.27	Moderate	21.00	Very high	12.50	803	4,703.00	Arylpolyene	6,645.00	7	Burkholderiales	Alcaligenaceae	Achromobacter
Peru	<i>C. similimus</i>	Bin 6_4	75.39	Substantial	24.50	Very high	7.32	826	4,667.00	Arylpolyene	7,790.00	9	Xanthomonadales	Xanthomonadaceae	Xanthomonas
Peru	<i>C. similimus</i>	Bin 6_5	75.39	Substantial	24.50	Very high	7.32	826	4,667.00	Butyrolactone	6,364.00	8	Aeromonadales	Aeromonadaceae	Aeromonas
Peru	<i>C. similimus</i>	Bin 6_2	75.39	Substantial	24.50	Very high	7.32	826	4,667.00	Furan	5,440.00	9	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Peru	<i>C. similimus</i>	Bin 5_2	60.27	Moderate	21.00	Very high	12.50	803	4,703.00	Lanthipeptide	8,781.00	8	Flavobacteriales	Flavobacteriaceae	Chryseobacterium
Peru	<i>C. spinosus</i>	Bin 2_2	87.93	Substantial	8.97	Medium	0.00	450	3,359.00	Arylpolyene	16,285.00	16	Burkholderiales	Alcaligenaceae	Bordetella
Peru	<i>C. spinosus</i>	Bin 2_4	87.93	Substantial	8.97	Medium	0.00	450	3,359.00	Arylpolyene	7,544.00	8	Burkholderiales	Alcaligenaceae	Achromobacter
Peru	<i>C. spinosus</i>	Bin 3	77.89	Substantial	18.97	Very high	75.00	267	2,519.00	Arylpolyene	25,079.00	26	Burkholderiales	Alcaligenaceae	Achromobacter
Peru	<i>C. spinosus</i>	Bin 4_2	41.38	Partial	5.17	Medium	33.33	173	2,059.00	Arylpolyene	6,935.00	5	Burkholderiales	Alcaligenaceae	Achromobacter
Peru	<i>C. spinosus</i>	Bin 5_1	98.28	Near	0.00	Low	0.00	81	2,658.00	Arylpolyene	41,203.00	44	Burkholderiales	Comamonadaceae	Acidovorax
Peru	<i>C. spinosus</i>	Bin 5_2	98.28	Near	0.00	Low	0.00	81	2,658.00	Arylpolyene	41,206.00	43	Burkholderiales	Comamonadaceae	Melaminivora
Peru	<i>C. spinosus</i>	Bin 8_1	93.18	Near	0.00	Low	0.00	825	3,903.00	Arylpolyene	5,023.00	8	Campylobacterales	Campylobacteraceae	Arcobacter
Peru	<i>C. spinosus</i>	Bin 7_1	100.00	Near	0.00	Low	0.00	514	4,141.00	Arylpolyene	36,717.00	32	Rhizobiales	Rhizobiaceae	Rhizobium
Peru	<i>C. spinosus</i>	Bin 8_2	93.18	Near	0.00	Low	0.00	825	3,903.00	Betalactone	14,713.00	13	Campylobacterales	Campylobacteraceae	Arcobacter
Peru	<i>C. spinosus</i>	Bin 6_1	54.20	Moderate	18.97	Very high	33.33	1,067.00	5,996.00	Butyrolactone	5,440.00	6	Burkholderiales	Comamonadaceae	Verminephrobacter
Peru	<i>C. spinosus</i>	Bin 2_1	87.93	Substantial	8.97	Medium	0.00	450	3,359.00	Ectoine	9,895.00	11	Burkholderiales	Alcaligenaceae	Pigmentiphaga
Peru	<i>C. spinosus</i>	Bin 6_4	54.20	Moderate	18.97	Very high	33.33	1,067.00	5,996.00	Lanthipeptide	6,371.00	4	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
Peru	<i>C. spinosus</i>	Bin 6_3	54.20	Moderate	18.97	Very high	33.33	1,067.00	5,996.00	NRP	12,639.00	12	Actinomycetales	Streptomycetaceae	Streptomyces
Peru	<i>C. spinosus</i>	Bin 6_5	54.20	Moderate	18.97	Very high	33.33	1,067.00	5,996.00	NRP	7,901.00	2	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Peru	<i>C. spinosus</i>	Bin 6_7	54.20	Moderate	18.97	Very high	33.33	1,067.00	5,996.00	NRP	6,715.00	7	Pseudomonadales	Pseudomonadaceae	Pseudomonas

Peru	<i>C. spinosus</i>	Bin 6_2	54.20	Moderate	18.97	Very high	33.33	1,067.00	5,996.00	Phenazine	14,940.00	14	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Peru	<i>C. spinosus</i>	Bin 7_2	100.00	Near	0.00	Low	0.00	514	4,141.00	Siderophore	7,469.00	4	Rhizobiales	Rhizobiaceae	Sinorhizobium
Peru	<i>C. spinosus</i>	Bin 7_3	100.00	Near	0.00	Low	0.00	514	4,141.00	Siderophore	8,038.00	6	Rhizobiales	Rhizobiaceae	Sinorhizobium
Peru	<i>C. spinosus</i>	Bin 2_3	87.93	Substantial	8.97	Medium	0.00	450	3,359.00	Terpene	21,709.00	21	Burkholderiales	Alcaligenaceae	Bordetella
Peru	<i>C. spinosus</i>	Bin 5_4	98.28	Near	0.00	Low	0.00	81	2,658.00	Terpene	5,320.00	5	Burkholderiales	Comamonadaceae	Acidovorax
Peru	<i>C. umbraculatus</i>	Bin 2_1	63.79	Moderate	30.38	Very high	3.12	664	3,383.00	Arylpolyene	26,451.00	31	Burkholderiales	Alcaligenaceae	Achromobacter
Peru	<i>C. umbraculatus</i>	Bin 5	98.28	Near	21.00	Very high	0.00	4,233.00	11,471.00	CDPS	9,839.00	15	Bdellovibrionales	Bdellovibrionaceae	Micavibrio
Peru	<i>C. umbraculatus</i>	Bin 3_2	93.03	Near	3.45	Low	33.33	332	4,114.00	Arylpolyene	34,176.00	36	Burkholderiales	Comamonadaceae	Delftia
Peru	<i>C. umbraculatus</i>	Bin 8_1	79.31	Substantial	14.55	High	0.00	669	3,307.00	Arylpolyene	32,683.00	27	Flavobacteriales	Crocinitomicaceae	Fluviicola
Peru	<i>C. umbraculatus</i>	Bin 8_2	79.31	Substantial	14.55	High	0.00	669	3,307.00	Arylpolyene	27,994.00	27	Flavobacteriales	Crocinitomicaceae	Fluviicola
Peru	<i>C. umbraculatus</i>	Bin 9_1	94.83	Near	2.35	Low	0.00	753	4,373.00	Arylpolyene	21,416.00	27	Flavobacteriales	Crocinitomicaceae	Fluviicola
Peru	<i>C. umbraculatus</i>	Bin 9_2	94.83	Near	2.35	Low	0.00	753	4,373.00	Arylpolyene	27,532.00	28	Flavobacteriales	Flavobacteriaceae	Elizabethkingia
Peru	<i>C. umbraculatus</i>	Bin 12	76.49	Substantial	50.00	Very high	15.38	374	2,580.00	Arylpolyene	33,424.00	32	Flavobacteriales	Flavobacteriaceae	Chryseobacterium
Peru	<i>C. umbraculatus</i>	Bin 4_2	100.00	Near	0.86	Low	0.00	147	2,493.00	Arylpolyene	26,803.00	30	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Peru	<i>C. umbraculatus</i>	Bin 4_1	100.00	Near	0.86	Low	0.00	147	2,493.00	Arylpolyene-Ladderane	43,125.00	38	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Peru	<i>C. umbraculatus</i>	Bin 3_4	93.03	Near	3.45	Low	33.33	332	4,114.00	NRP	7,797.00	6	Bacillales	Paenibacillaceae	Brevibacillus

Peru	<i>C. umbraculatus</i>	Bin 3_3	93.03	Near	3.45	Low	33.33	332	4,114.00	NRP	12,405.00	14	Pseudomonadales	Moraxellaceae	Acinetobacter
Peru	<i>C. umbraculatus</i>	Bin 2_2	63.79	Moderate	30.38	Very high	3.12	664	3,383.00	Phenazine	7,847.00	12	Xanthomonadales	Xanthomonadaceae	Lysobacter
Peru	<i>C. umbraculatus</i>	Bin 3_1	93.03	Near	3.45	Low	33.33	332	4,114.00	Terpene	10,431.00	11	Burkholderiales	Comamonadaceae	Delftia
USA	<i>C. varians</i> PLO05	Bin 6_4	88.97	Substantial	48.28	Very high	52.50	219	3,237.00	Arylpolycyclic	24,835.00	28	Burkholderiales	Alcaligenaceae	Achromobacter
USA	<i>C. varians</i> PLO05	Bin 4_1	98.28	Near	1.72	Low	100.00	253	2,957.00	Arylpolycyclic	31,597.00	34	Pseudomonadales	Ventrosimonadaceae	Ventrosimonas
USA	<i>C. varians</i> PLO05	Bin 10_2	87.93	Substantial	34.48	Very high	6.90	583	3,581.00	Arylpolycyclic	17,255.00	22	Pseudomonadales	Pseudomonadaceae	Pseudomonas
USA	<i>C. varians</i> PLO05	Bin 3_1	85.58	Substantial	57.76	Very high	45.24	459	3,650.00	Arylpolycyclic	8,577.00	8	Rhizobiales	Bartonellaceae	Bartonella
USA	<i>C. varians</i> PLO05	Bin 10_1	87.93	Substantial	34.48	Very high	6.90	583	3,581.00	Arylpolycyclic	5,347.00	5	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
USA	<i>C. varians</i> PLO05	Bin 1	98.28	Near	2.38	Low	0.00	1,020.00	4,203.00	Betalactone	20,310.00	17	Campylobacterales	Campylobacteraceae	Arcobacter
USA	<i>C. varians</i> PLO05	Bin 4_3	98.28	Near	1.72	Low	100.00	253	2,957.00	Butyrolactone	6,403.00	9	Aeromonadales	Aeromonadaceae	Aeromonas
USA	<i>C. varians</i> PLO05	Bin 11_1	46.08	Partial	31.03	Very high	64.00	480	3,651.00	Ladderane	7,506.00	10	Enterobacteriales	Enterobacteriaceae	Serratia
USA	<i>C. varians</i> PLO05	Bin 11_2	46.08	Partial	31.03	Very high	64.00	480	3,651.00	Ladderane	18,686.00	19	Enterobacteriales	Enterobacteriaceae	Serratia
USA	<i>C. varians</i> PLO05	Bin 9_1	79.31	Substantial	38.54	Very high	42.61	422	3,599.00	NRP	19,629.00	7	Bacillales	Paenibacillaceae	Brevibacillus
USA	<i>C. varians</i> PLO05	Bin 4_4	98.28	Near	1.72	Low	100.00	253	2,957.00	NRP	8,444.00	7	Burkholderiales	Burkholderiaceae	Ralstonia
USA	<i>C. varians</i> PLO05	Bin 5	89.66	Substantial	0.00	Low	0.00	67	1,793.00	NRP	28,568.00	25	Burkholderiales	Alcaligenaceae	Alcaligenes
USA	<i>C. varians</i> PLO05	Bin 6_2	88.97	Substantial	48.28	Very high	52.50	219	3,237.00	NRP	30,680.00	27	Burkholderiales	Alcaligenaceae	Alcaligenes
USA	<i>C. varians</i> PLO05	Bin 4_2	98.28	Near	1.72	Low	100.00	253	2,957.00	Siderophore	9,888.00	11	Burkholderiales	Alcaligenaceae	Achromobacter
USA	<i>C. varians</i> PLO05	Bin 7	75.08	Substantial	41.69	Very high	58.82	224	2,656.00	T1PK	10,882.00	5	Burkholderiales	Alcaligenaceae	Achromobacter
USA	<i>C. varians</i> PLO05	Bin 8	98.28	Near	26.80	Very high	3.85	120	2,542.00	T1PK	16,345.00	8	Burkholderiales	Alcaligenaceae	Achromobacter
USA	<i>C. varians</i> PLO05	Bin 9_2	79.31	Substantial	38.54	Very high	42.61	422	3,599.00	Terpene	18,015.00	13	Burkholderiales	Comamonadaceae	Delftia
USA	<i>C. varians</i> PLO05	Bin 9_3	79.31	Substantial	38.54	Very high	42.61	422	3,599.00	Terpene	14,521.00	11	Burkholderiales	Comamonadaceae	Delftia

USA	<i>C. varians</i> PLO10	Bin_1_1	87.93	Substantial	0.00	Low	0.00	63	1845	Betalactone	20574	17	Campylobacterales	Campyrobacteraceae	Arcobacter
USA	<i>C. varians</i> PLO10	Bin_9_1	94.83	Near	0.00	Low	0.00	84	2079	Resorcino	20727	21	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas
USA	<i>C. varians</i> PLO10	Bin_7_1	98.28	Near	0.00	Low	0.00	136	1586	Terpene	15447	32	Burkholderiales	Comamonadaceae	Acidovorax
USA	<i>C. varians</i> PLO10	Bin_7_2	98.28	Near	0.00	Low	0.00	136	1586	Terpene	11621	10	Burkholderiales	Comamonadaceae	Delftia

Supplementary Table 6. Genomic similarity measurement between metagenomic bins and isolated genomes.

Host	Genome	Metagenome	Order_identified	gANI	AF	%_mapped
<i>C. rohweri</i>	Burkholderiales sp. Cag20	Bin_5	Burkholderiales Opituales Xanthomonadales	0.52	0.06	2.802
		Bin_8	Pseudomonadales	0.73	0.08	0.064
		Bin_9	Burkholderiales Xanthomonadales	0.59	0.09	0.367
		Bin_10	Burkholderiales	0.55	0.15	0.032
		Bin_11	Burkholderiales	0.69	0.09	0.128
		Bin_12	Burkholderiales Pseudomonadales	0.74	0.18	2.747
		Bin_13	Burkholderiales	0.70	0.21	0.926
		Bin_14	Burkholderiales	0.97	0.85	41.860
		Bin_15	Burkholderiales Pseudomonadales	0.98	0.87	53.657
		Bin_16	Burkholderiales	0.74	0.08	0.064
		Bin_20	Burkholderiales	0.00	0.00	0.016
	Burkholderiales sp. Cag25	Bin_5	Burkholderiales Opituales Xanthomonadales	0.51	0.09	0.031
		Bin_8	Pseudomonadales	0.64	0.12	0.031
		Bin_9	Burkholderiales Xanthomonadales	0.51	0.14	0.031
		Bin_10	Burkholderiales	0.61	0.12	0.031
		Bin_11	Burkholderiales	0.52	0.17	0.062
		Bin_12	Burkholderiales Pseudomonadales	0.83	0.12	13.383
		Bin_13	Burkholderiales	0.72	0.23	0.078
		Bin_14	Burkholderiales	0.98	0.87	45.684
		Bin_15	Burkholderiales Pseudomonadales	0.98	0.89	38.410
		Bin_16	Burkholderiales	0.70	0.16	0.047
		Bin_20	Burkholderiales	0.00	0.00	0.000
	<i>Cephalotococcus primus</i> Cag34	Bin_5	Burkholderiales Opituales Xanthomonadales	0.88	0.02	2.123
		Bin_8	Pseudomonadales	0.00	0.00	0.000
		Bin_9	Burkholderiales Xanthomonadales	0.00	0.00	0.018
		Bin_10	Burkholderiales	0.00	0.00	0.000
		Bin_11	Burkholderiales	0.00	0.00	0.000
		Bin_12	Burkholderiales Pseudomonadales	0.00	0.00	0.000

		Bin_13	Burkholderiales	0.00	0.00	0.000
		Bin_14	Burkholderiales	0.71	0.01	0.000
		Bin_15	Burkholderiales Pseudomonadales	0.00	0.00	0.000
		Bin_16	Burkholderiales	0.96	0.82	35.595
		Bin_20	Burkholderiales	0.00	0.00	0.064
	<i>Ventossimonas sp.</i> Cag26	Bin_5	Burkholderiales Opitutales Xanthomonadales	0.96	0.80	11.309
		Bin_8	Pseudomonadales	0.78	0.05	2.635
		Bin_9	Burkholderiales Xanthomonadales	0.71	0.04	0.000
		Bin_10	Burkholderiales	0.67	0.10	0.000
		Bin_11	Burkholderiales	0.72	0.08	0.000
		Bin_12	Burkholderiales Pseudomonadales	0.99	0.89	65.371
		Bin_13	Burkholderiales	0.00	0.00	0.000
		Bin_14	Burkholderiales	0.70	0.03	1.145
		Bin_15	Burkholderiales Pseudomonadales	0.00	0.00	0.000
		Bin_16	Burkholderiales	0.54	0.09	0.000
		Bin_20	Burkholderiales	0.00	0.00	0.036
		<i>Ventossimonas sp.</i> Cag27	Bin_5	Burkholderiales Opitutales Xanthomonadales	0.96	0.80
	Bin_8		Pseudomonadales	0.78	0.05	2.623
	Bin_9		Burkholderiales Xanthomonadales	0.71	0.04	0.000
	Bin_10		Burkholderiales	0.67	0.10	0.000
	Bin_11		Burkholderiales	0.72	0.08	0.000
	Bin_12		Burkholderiales Pseudomonadales	0.99	0.89	65.496
	Bin_13		Burkholderiales	0.00	0.00	0.000
	Bin_14		Burkholderiales	0.70	0.03	0.146
	Bin_15		Burkholderiales Pseudomonadales	0.00	0.00	0.000
	Bin_16		Burkholderiales	0.54	0.09	0.000
	<i>Ventossimonas sp.</i> Cag320	Bin_5	Burkholderiales Opitutales Xanthomonadales	0.87	0.00	0.000
		Bin_8	Pseudomonadales	0.80	0.07	3.090
		Bin_9	Burkholderiales Xanthomonadales	0.64	0.05	1.035
		Bin_10	Burkholderiales	0.69	0.07	0.000
		Bin_11	Burkholderiales	0.00	0.00	0.000

		Bin_12	Burkholderiales Pseudomonadales	0.99	0.84	56.797		
		Bin_13	Burkholderiales	0.00	0.00	1.105		
		Bin_14	Burkholderiales	0.00	0.00	0.018		
		Bin_15	Burkholderiales Pseudomonadales	0.61	0.04	0.000		
		Bin_16	Burkholderiales	0.66	0.15	0.000		
		Bin_20	Burkholderiales	0.00	0.00	1.193		
	Xanthomonadales sp. Cag60	Bin_5	Burkholderiales Opitutales Xanthomonadales	0.74	0.03	0.966		
		Bin_8	Pseudomonadales	0.69	0.08	0.115		
		Bin_9	Burkholderiales Xanthomonadales	0.97	0.82	33.362		
		Bin_10	Burkholderiales	0.68	0.08	0.115		
		Bin_11	Burkholderiales	0.71	0.04	0.014		
		Bin_12	Burkholderiales Pseudomonadales	0.76	0.05	0.894		
		Bin_13	Burkholderiales	0.64	0.07	0.144		
		Bin_14	Burkholderiales	0.00	0.00	0.072		
		Bin_15	Burkholderiales Pseudomonadales	0.65	0.05	0.029		
		Bin_16	Burkholderiales	0.72	0.05	0.000		
		Bin_20	Burkholderiales	0.00	0.00	0.014		
		<i>C_varians_005</i>	Burkholderiales sp. Cv33a	Bin_3	Rhizobiales	0.66	0.03	0.060
				Bin_4	Burkholderiales Pseudomonadales	0.69	0.08	0.959
Bin_5	Burkholderiales			0.66	0.15	0.100		
Bin_6	Burkholderiales			0.66	0.21	0.160		
Bin_7	Burkholderiales			0.68	0.12	0.040		
Bin_8	Burkholderiales			0.98	0.93	83.267		
Bin_9	Burkholderiales			0.70	0.20	4.153		
Bin_10	Pseudomonadales Xanthomonadales			0.58	0.05	0.060		
Burkholderiales sp. Cv36	Bin_3		Rhizobiales	0.00	0.00	0.000		
	Bin_4		Burkholderiales Pseudomonadales	0.72	0.07	0.588		
	Bin_5		Burkholderiales	0.61	0.14	0.252		
	Bin_6		Burkholderiales	0.72	0.13	0.336		
	Bin_7		Burkholderiales	0.64	0.12	0.605		
	Bin_8		Burkholderiales	0.71	0.24	0.453		
	Bin_9		Burkholderiales	0.78	0.10	1.075		
Bin_10	Pseudomonadales Xanthomonadales		0.66	0.04	0.084			
Burkholderiales sp. Cv44	Bin_3		Rhizobiales	0.00	0.00	0.046		
	Bin_4		Burkholderiales Pseudomonadales	0.73	0.08	1.834		

		Bin_5	Burkholderiales	0.74	0.15	1.136
		Bin_6	Burkholderiales	0.91	0.18	22.255
		Bin_7	Burkholderiales	0.97	0.83	36.860
		Bin_8	Burkholderiales	0.67	0.12	1.303
		Bin_9	Burkholderiales	0.86	0.14	14.084
		Bin_10	Pseudomonadales Xanthomonadales	0.54	0.05	1.046
	Burkholderiales sp. Cv52	Bin_3	Rhizobiales	0.64	0.04	0.000
		Bin_4	Burkholderiales Pseudomonadales	0.80	0.08	0.485
		Bin_5	Burkholderiales	0.61	0.14	0.218
		Bin_6	Burkholderiales	0.62	0.18	0.435
		Bin_7	Burkholderiales	0.74	0.16	0.301
		Bin_8	Burkholderiales	0.75	0.13	0.268
		Bin_9	Burkholderiales	0.77	0.13	1.138
	<i>Cephalotococcus capnophilus</i> Cv41	Bin_3	Rhizobiales	0.00	0.00	0.000
		Bin_4	Burkholderiales Pseudomonadales	0.72	0.02	1.523
		Bin_5	Burkholderiales	0.57	0.12	0.000
		Bin_6	Burkholderiales	0.69	0.18	2.024
		Bin_7	Burkholderiales	0.56	0.09	2.024
		Bin_8	Burkholderiales	0.00	0.00	0.000
		Bin_9	Burkholderiales	0.73	0.04	2.262
	Rhizobiales sp. JR021- 5	Bin_3	Rhizobiales	0.00	0.00	0.000
		Bin_4	Burkholderiales Pseudomonadales	0.00	0.00	0.000
		Bin_5	Burkholderiales	0.00	0.00	0.000
		Bin_6	Burkholderiales	0.00	0.00	0.000
		Bin_7	Burkholderiales	0.00	0.00	0.000
		Bin_8	Burkholderiales	0.00	0.00	0.000
		Bin_9	Burkholderiales	0.00	0.00	0.000
	<i>Ventosimonas gracilis</i> Cv58	Bin_3	Rhizobiales	0.00	0.00	0.000
		Bin_4	Burkholderiales Pseudomonadales	0.96	0.84	49.317
		Bin_5	Burkholderiales	0.00	0.00	0.056
Bin_6		Burkholderiales	0.70	0.06	0.131	
Bin_7		Burkholderiales	0.68	0.06	0.281	
Bin_8		Burkholderiales	0.55	0.19	0.056	
Bin_9		Burkholderiales	0.76	0.08	0.842	
Bin_10	Pseudomonadales Xanthomonadales	0.61	0.05	0.112		

<i>C_varians_010</i>	Burkholderiales sp. Cv33a	Bin_1	Burkholderiales	1.00	0.85	98.488
		Bin_2	Burkholderiales	0.70	0.19	0.045
		Bin_3	Rhizobiales	0.00	0.00	0.045
		Bin_4	Rhizobiales	0.00	0.00	0.000
		Bin_5	Opitutales	0.70	0.11	0.073
		Bin_7	Burkholderiales	0.65	0.10	0.043
		Bin_8	Rhizobiales	0.68	0.02	0.046
		Bin_9	Burkholderiales Pseudomonadales Xanthomonadales	0.67	0.09	0.124
	Burkholderiales sp. Cv36	Bin_1	Burkholderiales	0.70	0.15	0.059
		Bin_2	Burkholderiales	0.78	0.19	0.226
		Bin_3	Rhizobiales	0.00	0.00	0.045
		Bin_4	Rhizobiales	0.00	0.00	0.000
		Bin_5	Opitutales	0.52	0.08	0.024
		Bin_7	Burkholderiales	0.71	0.08	0.000
		Bin_8	Rhizobiales	0.00	0.00	0.023
		Bin_9	Burkholderiales Pseudomonadales Xanthomonadales	0.61	0.04	0.025
	Burkholderiales sp. Cv44	Bin_1	Burkholderiales	0.75	0.15	0.119
		Bin_2	Burkholderiales	0.74	0.16	0.000
		Bin_3	Rhizobiales	0.00	0.00	0.045
		Bin_4	Rhizobiales	0.00	0.00	0.000
		Bin_5	Opitutales	0.57	0.11	0.024
		Bin_7	Burkholderiales	0.69	0.08	0.086
		Bin_8	Rhizobiales	0.70	0.02	0.046
		Bin_9	Burkholderiales Pseudomonadales Xanthomonadales	0.58	0.05	0.000
	Burkholderiales sp. Cv52	Bin_1	Burkholderiales	0.76	0.12	0.119
		Bin_2	Burkholderiales	1.00	0.50	96.482
		Bin_3	Rhizobiales	0.00	0.00	0.000
		Bin_4	Rhizobiales	0.00	0.00	0.000
		Bin_5	Opitutales	0.65	0.16	0.171
		Bin_7	Burkholderiales	0.69	0.11	0.065
		Bin_8	Rhizobiales	0.66	0.02	0.023
		Bin_9	Burkholderiales Pseudomonadales Xanthomonadales	0.70	0.03	0.050
	<i>Cephalotococcus capnophilus</i> Cv41	Bin_1	Burkholderiales	0.00	0.00	0.030
Bin_2		Burkholderiales	0.00	0.00	0.000	
Bin_3		Rhizobiales	0.00	0.00	0.000	
Bin_4		Rhizobiales	0.00	0.00	0.000	
Bin_5		Opitutales	0.96	0.63	80.562	
Bin_7		Burkholderiales	0.00	0.00	0.065	
Bin_8	Rhizobiales	0.00	0.00	0.023		

		Bin_9	Burkholderiales Pseudomonadales Xanthomonadales	0.00	0.00	0.000
	Rhizobiales sp. JR021-5	Bin_1	Burkholderiales	0.72	0.01	0.000
		Bin_2	Burkholderiales	0.70	0.01	0.000
		Bin_3	Rhizobiales	0.00	0.00	0.000
		Bin_4	Rhizobiales	0.00	0.00	0.000
		Bin_5	Opituales	0.00	0.00	0.000
		Bin_7	Burkholderiales	0.00	0.00	0.000
		Bin_8	Rhizobiales	0.00	0.00	0.000
			Bin_9	Burkholderiales Pseudomonadales Xanthomonadales	0.00	0.00
	<i>Ventosimonas gracilis</i> Cv58	Bin_1	Burkholderiales	0.68	0.07	0.059
		Bin_2	Burkholderiales	0.68	0.09	0.045
		Bin_3	Rhizobiales	0.00	0.00	0.045
		Bin_4	Rhizobiales	0.00	0.00	0.000
		Bin_5	Opituales	0.42	0.10	0.049
		Bin_7	Burkholderiales	0.69	0.08	0.108
		Bin_8	Rhizobiales	0.69	0.03	0.023
			Bin_9	Burkholderiales Pseudomonadales Xanthomonadales	0.00	0.00

Supplementary Table 7. Genomic similarity measurement between BGCs from isolated genomes and metagenomic bins.

Host	Genome	BGC	Metagenome	Order_identified	gANI	AF	%_bacterial_contigs_mapped
C_rohweri	primus	Arylpolylene	Bin_4_2	Rhizobiales	0.00	0.00	0.000
			Bin_5_2	Rhizobiales	0.00	0.00	0.000
			Bin_5_5	Pseudomonadales	0.71	0.45	0.000
			Bin_8	Pseudomonadales	0.00	0.00	0.000
	Ventosimonas_26_1	Arylpolylene	Bin_4_2	Rhizobiales	0.00	0.00	0.000
			Bin_5_2	Rhizobiales	0.00	0.00	0.000
			Bin_5_5	Pseudomonadales	1.00	0.98	100.000
			Bin_8	Pseudomonadales	0.00	0.00	28.000
	Ventosimonas_27_2	Arylpolylene	Bin_4_2	Rhizobiales	0.00	0.00	0.000
			Bin_5_2	Rhizobiales	0.00	0.00	0.000
			Bin_5_5	Pseudomonadales	1.00	0.98	100.000
			Bin_8	Pseudomonadales	0.00	0.00	28.000
	Xanthomonadales_60_5	Arylpolylene	Bin_4_2	Rhizobiales	0.00	0.00	0.000
			Bin_5_2	Rhizobiales	0.00	0.00	0.000
			Bin_5_5	Pseudomonadales	0.00	0.00	0.000
			Bin_8	Pseudomonadales	0.00	0.00	0.000
	Burkholderiales_20_2	NRP	Bin_5_3	Burkholderiales	0.00	0.00	0.000
			Bin_12_1	Burkholderiales	0.00	0.00	0.000
			Bin_12_2	Neisseriales	0.00	0.00	0.000
			Bin_14_1	Burkholderiales	1.00	0.98	100.000
Bin_14_2			Burkholderiales	0.00	0.00	0.000	
Bin_20			Burkholderiales	0.00	0.00	0.000	
Burkholderiales_20_4	NRP	Bin_5_3	Burkholderiales	0.00	0.00	0.000	
		Bin_12_1	Burkholderiales	0.00	0.00	0.000	
		Bin_12_2	Neisseriales	0.00	0.00	0.000	
		Bin_14_1	Burkholderiales	0.00	0.00	0.000	

			Bin_14_2	Burkholderiales	0.00	0.00	0.000
			Bin_20	Burkholderiales	0.00	0.00	0.000
	Burkholderiales_20_5	NRP	Bin_5_3	Burkholderiales	0.00	0.00	0.000
			Bin_12_1	Burkholderiales	0.00	0.00	0.000
			Bin_12_2	Neisseriales	0.00	0.00	0.000
			Bin_14_1	Burkholderiales	0.00	0.00	0.000
			Bin_14_2	Burkholderiales	0.00	0.00	0.000
			Bin_20	Burkholderiales	0.00	0.00	0.000
	Burkholderiales_25_3	NRP	Bin_5_3	Burkholderiales	0.00	0.00	0.000
			Bin_12_1	Burkholderiales	0.00	0.00	0.000
			Bin_12_2	Neisseriales	0.00	0.00	0.000
			Bin_14_1	Burkholderiales	0.00	0.00	0.000
			Bin_14_2	Burkholderiales	0.99	0.98	100.000
			Bin_20	Burkholderiales	0.00	0.00	0.000
	Burkholderiales_25_4	NRP	Bin_5_3	Burkholderiales	0.00	0.00	0.000
			Bin_12_1	Burkholderiales	0.00	0.00	0.000
			Bin_12_2	Neisseriales	0.00	0.00	0.000
			Bin_14_1	Burkholderiales	0.00	0.00	0.000
			Bin_14_2	Burkholderiales	0.00	0.00	0.000
			Bin_20	Burkholderiales	0.00	0.00	0.000
	Ventosimonas_26_2	NRP	Bin_5_3	Burkholderiales	0.00	0.00	0.000
			Bin_12_1	Burkholderiales	0.00	0.00	0.000
			Bin_12_2	Pseudomonadales	0.96	0.86	92.658
			Bin_14_1	Burkholderiales	0.00	0.00	0.000
			Bin_14_2	Burkholderiales	0.00	0.00	0.000

			Bin_20	Burkholderiales	0.00	0.00	0.000
	Ventosimonas_27_1	NRP	Bin_5_3	Burkholderiales	0.00	0.00	0.000
			Bin_12_1	Burkholderiales	0.00	0.00	0.000
			Bin_12_2	Pseudomonadales	0.96	0.82	92.536
			Bin_14_1	Burkholderiales	0.00	0.00	0.000
			Bin_14_2	Burkholderiales	0.00	0.00	0.000
			Bin_20	Burkholderiales	0.00	0.00	0.000
	Ventosimonas_320_1	NRP	Bin_5_3	Burkholderiales	0.00	0.00	0.000
			Bin_12_1	Burkholderiales	0.00	0.00	0.000
			Bin_12_2	Neisseriales	0.88	0.48	0.000
			Bin_14_1	Burkholderiales	0.00	0.00	0.000
			Bin_14_2	Burkholderiales	0.00	0.00	0.000
			Bin_20	Burkholderiales	0.00	0.00	0.000
	Ventosimonas_320_2	NRP	Bin_5_3	Burkholderiales	0.00	0.00	0.000
			Bin_12_1	Burkholderiales	0.00	0.00	0.000
			Bin_12_2	Neisseriales	0.85	0.36	0.000
			Bin_14_1	Burkholderiales	0.00	0.00	0.000
			Bin_14_2	Burkholderiales	0.00	0.00	0.000
			Bin_20	Burkholderiales	0.00	0.00	0.000
	Xanthomonadales_60_1	NRP	Bin_5_3	Burkholderiales	0.00	0.00	0.000
			Bin_12_1	Burkholderiales	0.00	0.00	0.000
			Bin_12_2	Neisseriales	0.00	0.00	0.000
			Bin_14_1	Burkholderiales	0.00	0.00	0.000
			Bin_14_2	Burkholderiales	0.00	0.00	0.000
			Bin_20	Burkholderiales	0.00	0.00	0.000

	Xanthomonadales_60_3	NRP	Bin_5_3	Burkholderiales	0.00	0.00	0.000
			Bin_12_1	Burkholderiales	0.00	0.00	0.000
			Bin_12_2	Neisseriales	0.00	0.00	0.000
			Bin_14_1	Burkholderiales	0.00	0.00	0.000
			Bin_14_2	Burkholderiales	0.00	0.00	0.000
			Bin_20	Burkholderiales	0.00	0.00	0.000
	Xanthomonadales_60_4	Siderophore	Bin_5_4	Burkholderiales	0.00	0.00	0.000
	Xanthomonadales_60_2	Ladderane	Bin_9_2	Xanthomonadales	0.96	0.52	100.000
			Bin_9_3	Enterobacterales	0.00	0.00	0.000
	Burkholderiales_20_1	T1PK	Bin_10_3	Burkholderiales	0.00	0.00	0.000
			Bin_15_1	Burkholderiales	1.00	0.98	100.000
			Bin_15_4	Burkholderiales	0.80	0.39	0.000
	Burkholderiales_25_2	T1PK	Bin_10_3	Burkholderiales	0.00	0.00	0.000
			Bin_15_1	Burkholderiales	0.00	0.00	0.000
			Bin_15_4	Burkholderiales	0.99	0.98	90.909
	Burkholderiales_20_3	Betalactone	Bin_12_3	Pseudomonadales	0.00	0.00	0.000
			Bin_13	Burkholderiales	0.00	0.00	0.000
			Bin_15_2	Pseudomonadales	0.00	0.00	0.000
			Bin_15_3	Burkholderiales	1.00	0.98	100.000
	Burkholderiales_25_1	Betalactone	Bin_12_3	Pseudomonadales	0.00	0.00	0.000
			Bin_13	Burkholderiales	0.00	0.00	0.000
			Bin_15_2	Pseudomonadales	0.00	0.00	0.000
			Bin_15_3	Burkholderiales	0.99	0.98	97.222
C_varians_005	Burkholderiales_44_3	Betalactone	Bin_1	Campylobacterales	0.00	0.00	0.000
	Burkholderiales_36_2	Arylpolylene	Bin_3_1	Rhizobiales	0.00	0.00	0.000
			Bin_4_1	Pseudomonadales	0.00	0.00	0.000

			Bin_6_4	Burkholderiales	0.00	0.00	0.000
			Bin_10_1	Xanthomonadales	0.00	0.00	0.000
			Bin_10_2	Pseudomonadales	0.00	0.00	0.000
	Ventosimonas_58	Arylpol-yene	Bin_3_1	Rhizobiales	0.00	0.00	0.000
			Bin_4_1	Pseudomonadales	0.97	0.80	93.125
			Bin_6_4	Burkholderiales	0.00	0.00	0.000
			Bin_10_1	Xanthomonadales	0.00	0.00	0.000
			Bin_10_2	Pseudomonadales	0.00	0.00	0.000
	Burkholderiales_44_2	NRP	Bin_4_4	Burkholderiales	0.00	0.00	0.000
			Bin_5	Burkholderiales	0.00	0.00	0.000
			Bin_6_2	Burkholderiales	0.00	0.00	0.000
			Bin_9_1	Burkholderiales	0.00	0.00	0.000
	Burkholderiales_44_1	T1PK	Bin_7	Burkholderiales	1.00	0.98	100.000
			Bin_8	Burkholderiales	0.00	0.00	0.000
	Burkholderiales_52_3	T1PK	Bin_7	Burkholderiales	0.80	0.49	0.000
			Bin_8	Burkholderiales	0.00	0.00	0.000
	Burkholderiales_36_1	Terpene	Bin_9_2	Burkholderiales	0.00	0.00	0.000
	Burkholderiales_52_1	Terpene	Bin_9_2	Burkholderiales	0.00	0.00	0.000
C_varians_010	Burkholderiales_36_1	Terpene	Bin_7_1	Burkholderiales	0.00	0.00	0.000
			Bin_7_2	Burkholderiales	0.00	0.00	0.000
	Burkholderiales_52_1	Terpene	Bin_7_1	Burkholderiales	0.00	0.00	0.000
			Bin_7_2	Burkholderiales	0.00	0.00	0.000

Supplementary Table 8. Reference BGC and query protein selected in the genomic core analysis.

Database	Reference BGC	Reference BGC length (bp)	Reference BGC length (gene number)	Query protein domain	Query protein coordinates in the reference BGC
Arylpolyene	<i>C. spinosus</i> Bin_5_1	41,203	44	APE_KS2	38,905-40,107
NRP	<i>C. rohweri</i> Xanthomonadales Cag60_3	53,609	45	AMP-binding	64,471-65,991
T1PK	<i>C. rohweri</i> Burkholderiales Cag25_2	47,911	35	PKS_KS	197,260-205,170
Siderophore	<i>C. pellans</i> Bin_2_7	9,526	11	lucA_lucC	2,926-4,788

Bibliographie

1. Feldhaar, H. *et al.* Nutritional upgrading for omnivorous carpenter ants by the endosymbiont *Blochmannia*. *BMC Biol.* **5**, 1–11 (2007).
2. Duron, O. *et al.* Tick-Bacteria Mutualism Depends on B Vitamin Synthesis Pathways. *Curr. Biol.* **28**, 1896–1902.e5 (2018).
3. Schultz, T. R. & Brady, S. G. Major evolutionary transitions in ant agriculture. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5435–5440 (2008).
4. Aanen, D. K. *et al.* High symbiont relatedness stabilizes mutualistic cooperation in fungusgrowing termites. *Science (80-)*. **326**, 1103–1106 (2009).
5. Hulcr, J. & Stelinski, L. L. The Ambrosia Symbiosis: From Evolutionary Ecology to Practical Management. *Annu. Rev. Entomol.* **62**, 285–303 (2017).
6. Raymann, K. & Moran, N. A. The role of the gut microbiome in health and disease of adult honey bee workers. *Curr. Opin. Insect Sci.* **26**, 97–104 (2018).
7. Flórez, L. V., Biedermann, P. H. W., Engl, T. & Kaltenpoth, M. Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. *Nat. Prod. Reports* **32**, 904–936 (2015).
8. Duron, O. *et al.* The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biol.* **6**, 1–12 (2008).
9. Hughes, D. P. *et al.* Behavioral mechanisms and morphological symptoms of zombie ants dying from fungal infection. *BMC Ecol.* **11**, (2011).
10. Vega, F. E., Meyling, N. V., Luangsa-Ard, J. J. & Blackwell, M. *Fungal entomopathogens. Insect Pathology* (Elsevier, 2012). doi:10.1016/B978-0-12-384984-7.00006-3.
11. Castillo, J. C., Reynolds, S. E. & Eleftherianos, I. Insect immune responses to nematode parasites. *Trends Parasitol.* **27**, 537–547 (2011).
12. Stephen, K. & Ipek Kurtböke, D. Screening of oomycete fungi for their potential role in reducing the biting midge (Diptera: Ceratopogonidae) larval populations in Hervey bay, Queensland, Australia. *Int. J. Environ. Res. Public Health* **8**, 15601574 (2011).
13. Antonelli, A. *et al.* Amazonia is the primary source of Neotropical biodiversity. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6034–6039 (2018).
14. Griffiths, H. M. *et al.* Ants are the major agents of resource removal from tropical rainforests. *J. Anim. Ecol.* **87**, 293–300 (2018).
15. Noll, F. B., Nascimento, F. S. & Vasconcelos, H. Special Section: Social Insects in the Neotropics. *Neotrop. Entomol.* **47**, 729–730 (2018).
16. Cook, S. C. & Davidson, D. W. Nutritional and functional biology of exudate-feeding ants. *Entomol. Exp. Appl.* **118**, 1–10 (2006).
17. Anderson, K. E. *et al.* Highly similar microbial communities are shared among related and trophically similar ant species. *Mol. Ecol.* **21**, 2282–2296 (2012).
18. Hu, Y. *et al.* Herbivorous turtle ants obtain essential nutrients from a highly conserved

- nitrogen-recycling gut microbiome. *Nat. Commun.* **9**, 1–14 (2018).
19. Stroud, J. T. *et al.* Is a community still a community? Reviewing definitions of key terms in community ecology. *Ecol. Evol.* **5**, 4757–4765 (2015).
 20. Angelini, C., Altieri, A. H., Silliman, B. R. & Bertness, M. D. Interactions among foundation species and their consequences for community organization, biodiversity, and conservation. *Bioscience* **61**, 782–789 (2011).
 21. Mouquet, N., Gravel, D., Massol, F. & Calcagno, V. Extending the concept of keystone species to communities and ecosystems. *Ecol. Lett.* **16**, 1–8 (2013).
 22. Sanders, N. J., Gotelli, N. J., Heller, N. E. & Gordon, D. M. Community disassembly by an invasive species. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2474–2477 (2003).
 23. Margalef, R. Information theory in biology. *Gen. Syst. Yearb.* **3**, 36–71 (1958).
 24. Menhinick, E. F. A Comparison of Some Species-Individuals Diversity Indices Applied to Samples of Field Insects. *Ecology* **45**, 859–861 (1964).
 25. Fedor, P. & Zvaríková, M. Biodiversity Indices. in (ed. Fath, B. B. T.-E. of E. (Second E.)) 337–346 (Elsevier, 2019). doi:<https://doi.org/10.1016/B978-0-12-409548-9.10558-5>.
 26. Pielou, E. C. The measurement of diversity in different types of biological collections. *J. Theor. Biol.* **13**, 131–144 (1966).
 27. McIntosh, R. P. An Index of Diversity and the Relation of Certain Concepts to Diversity. *Ecology* **48**, 392–404 (1967).
 28. Kraft, N. J. B. *et al.* Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* **29**, 592–599 (2015).
 29. Soliveres, S. *et al.* Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. *Nature* **536**, 456–459 (2016).
 30. Houadria, M., Salas-Lopez, A., Orivel, J., Blüthgen, N. & Menzel, F. Dietary and Temporal Niche Differentiation in Tropical Ants-Can They Explain Local Ant Coexistence? *Biotropica* **47**, 208–217 (2015).
 31. Wei, M., Lloyd, H. & Zhang, Y. Neighbour–stranger discrimination by Yellow-bellied Tit *Parus venustus*: evidence for the “dear-enemy” effect. *J. Ornithol.* **152**, 431–438 (2011).
 32. Pavey, S. A., Collin, H., Nosil, P. & Rogers, S. M. The role of gene expression in ecological speciation. *Ann. N. Y. Acad. Sci.* **1206**, 110–129 (2010).
 33. Vellend, M. Conceptual synthesis in community ecology. *Q. Rev. Biol.* **85**, 183–206 (2010).
 34. Baguette, M., Benton, T. G. & Bullock, J. M. *Dispersal ecology and evolution*. (Oxford University Press, 2012).
 35. Fodelianakis, S., Valenzuela-Cuevas, A., Barozzi, A. & Daffonchio, D. Direct quantification of ecological drift at the population level in synthetic bacterial communities. *ISME J.* **15**, 55–66 (2021).
 36. Peña-Claros, M. Changes in Forest Structure and Species Composition during Secondary Forest Succession in the Bolivian Amazon. *Biotropica* **35**, 450–461 (2003).

37. Wiederkehr, F. *et al.* Urbanisation affects ecosystem functioning more than structure in tropical streams. *Biol. Conserv.* **249**, 108634 (2020).
38. Finlay, B. J., Esteban, G. F., Olmo, J. L. & Tyler, P. A. Global distribution of free-living microbial species. *Ecography (Cop.)*. **22**, 138–144 (1999).
39. Ghoul, M. & Mitri, S. The Ecology and Evolution of Microbial Competition. *Trends Microbiol.* **24**, 833–845 (2016).
40. de Bary, A. *Die erscheinung der symbiose: Vortrag gehalten auf der versammlung deutscher naturforscher und aerzte zu cassel.* (1879).
41. Martin, B. D. & Schwab, E. Current Usage of Symbiosis and Associated Terminology. *Int. J. Biol.* **5**, 32–45 (2012).
42. Leung, T. L. F. & Poulin, R. Parasitism, commensalism, mutualism, exploring the many shades of symbioses. *Vie Milieu* **58**, 107–115 (2008).
43. Zapalski, M. K. Is absence of proof a proof of absence? Comments on commensalism. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **302**, 484–488 (2011).
44. Husa, E. A. & Goodrich-Blair, H. It Takes a Village: Ecological and Fitness Impacts of Multipartite Mutualism. *Annu. Rev. Microbiol.* **67**, 161–178 (2013).
45. Brownlie, J. C. & Johnson, K. N. Symbiont-mediated protection in insect hosts. *Trends Microbiol.* **17**, 348–354 (2009).
46. Salem, H., Florez, L., Gerardo, N. & Kaltenpoth, M. An out-of-body experience: The extracellular dimension for the transmission of mutualistic bacteria in insects. *Proc. R. Soc. B Biol. Sci.* **282**, (2015).
47. Burgdorfer, W. & Brinton, L. P. Mechanisms of transovarial infection of spotted fever rickettsiae in ticks. *Ann. N. Y. Acad. Sci.* **266**, 61–72 (1975).
48. Onchuru, T. O., Javier Martinez, A., Ingham, C. S. & Kaltenpoth, M. Transmission of mutualistic bacteria in social and gregarious insects. *Curr. Opin. Insect Sci.* **28**, 50–58 (2018).
49. Hajek, A. E., Steinkraus, D. C. & Castrillo, L. A. Sleeping beauties: Horizontal transmission via resting spores of species in the entomophthoromycotina. *Insects* **9**, 1–23 (2018).
50. Perez, M. & Juniper, S. K. Is the trophosome of *Ridgeia piscesae* monoclonal? *Symbiosis* **74**, 55–65 (2018).
51. Engelmoer, D. J. P., Behm, J. E. & Toby Kiers, E. Intense competition between arbuscular mycorrhizal mutualists in an in vitro root microbiome negatively affects total fungal abundance. *Mol. Ecol.* **23**, 1584–1593 (2014).
52. Mouton, L. *et al.* Virulence, multiple infections and regulation of symbiotic population in the *Wolbachia*-*Asobara tabida* symbiosis. *Genetics* **168**, 181–189 (2004).
53. Currie, C. R., Scottt, J. A., Summerbell, R. C. & Malloch, D. Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature* **398**, 701–704 (1999).
54. Scott, J. J. *et al.* Bacterial Protection of Beetle-Fungus Mutualism. *Science (80-)*. **322**, 63 (2008).
55. Currie, C. R. & Stuart, A. E. Weeding and grooming of pathogens in agriculture by ants.

- Proc. R. Soc. B Biol. Sci.* **268**, 1033–1039 (2001).
56. Freter, R., Brickner, H., Botney, M., Cleven, D. & Aranki, A. Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. *Infect. Immun.* **39**, 676–685 (1983).
 57. Freter, R., Brickner, H., Fekete, J., Vickerman, M. M. & Carey, K. E. Survival and implantation of *Escherichia coli* in the intestinal tract. *Infect. Immun.* **39**, 686–703 (1983).
 58. Chaucheyras-Durand, F. & Durand, H. Probiotics in animal nutrition and health. *Benef. Microbes* **1**, 3–9 (2010).
 59. Evans, H. C. Co-evolution of entomogenous fungi and their insect hosts. *Co-evolution Fungi with Plants Anim.* (1988).
 60. Li, X., Guo, W. & Ding, J. Mycangial fungus benefits the development of a leaf-rolling weevil, *Euops chinesis*. *J. Insect Physiol.* **58**, 867–873 (2012).
 61. Choat, J. H. & Clements, K. D. Vertebrate herbivores in marine and terrestrial environments: a nutritional ecology perspective. *Annu. Rev. Ecol. Syst.* **29**, 375–403 (1998).
 62. Chung, S. H. *et al.* Herbivore exploits orally secreted bacteria to suppress plant defenses. *PNAS* **110**, 15728–15733 (2013).
 63. Stewart, G. S. & Smith, C. P. Urea nitrogen salvage mechanisms and their relevance to ruminants, non-ruminants and man. *Nutr. Res. Rev.* **18**, 49–62 (2005).
 64. Bermingham, E. N., Maclean, P., Thomas, D. G., Cave, N. J. & Young, W. Key bacterial families (Clostridiaceae, Erysipelotrichaceae and Bacteroidaceae) are related to the digestion of protein and energy in dogs. *PeerJ* **2017**, (2017).
 65. Sapountzis, P. *et al.* Acromyrmex leaf-cutting ants have simple gut microbiota with nitrogen-fixing potential. *Appl. Environ. Microbiol.* **81**, 5527–5537 (2015).
 66. Redman, R., Sheehan, K., Stout, R., Rodriguez, R. & Hensen, J. Thermotolerance generated by plant/fungus symbiosis. *Science (80-.)*. **298**, 1581 (2002).
 67. Ferguson, L. V. *et al.* Seasonal shifts in the insect gut microbiome are concurrent with changes in cold tolerance and immunity. *Funct. Ecol.* **32**, 2357–2368 (2018).
 68. Lemoine, M. M., Engl, T. & Kaltenpoth, M. Microbial symbionts expanding or constraining abiotic niche space in insects. *Curr. Opin. Insect Sci.* **39**, 14–20 (2020).
 69. Adams, A. E., Kazenel, M. R. & Rudgers, J. A. Does a foliar endophyte improve plant fitness under flooding? *Plant Ecol.* **218**, 711–723 (2017).
 70. Heijtz, R. D. *et al.* Normal gut microbiota modulates brain development and behavior. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 3047–3052 (2011).
 71. Douglas, A. E. Review Lessons from Studying Insect Symbioses. *CHOM* **10**, 359–367 (2011).
 72. Barke, J. *et al.* A mixed community of actinomycetes produce multiple antibiotics for the fungus farming ant *Acromyrmex octospinosus*. *BMC Biol.* **8**, 1–10 (2010).
 73. Chevrette, M. G. *et al.* The antimicrobial potential of *Streptomyces* from insect

- microbiomes. *Nat. Commun.* **10**, 1–11 (2019).
74. Martín-Vivaldi, M. *et al.* Special structures of hoopoe eggshells enhance the adhesion of symbiont-carrying uropygial secretion that increase hatching success. *J. Anim. Ecol.* **83**, 1289–1301 (2014).
 75. Engl, T. *et al.* Evolutionary stability of antibiotic protection in a defensive symbiosis. *Proc. Natl. Acad. Sci.* **115**, E2020–E2029 (2018).
 76. Jensen, P. R. Natural products and the gene cluster revolution. *Trends Microbiol.* **24**, 968–977 (2016).
 77. Miller, M. B. & Bassler, B. L. Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **55**, 165–199 (2001).
 78. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412–421 (2014).
 79. Kramer, J., Özkaya, Ö. & Kümmerli, R. Bacterial siderophores in community and host interactions. *Nat. Rev. Microbiol.* **18**, 152–163 (2020).
 80. Baltz, R. H. Renaissance in antibacterial discovery from actinomycetes. *Curr. Opin. Pharmacol.* **8**, 557–563 (2008).
 81. Velkov, T., Roberts, K. D., Nation, R. L., Thompson, P. E. & Li, J. Pharmacology of polymyxins: new insights into an ‘old’ class of antibiotics. *Future Microbiol.* **8**, 114–137 (2013).
 82. Esipov, S. E., Adanin, V. M., Baskunov, B. P., Kiprianova, E. A. & Garagulia, A. D. New antibiotically active fluoroglucide from *Pseudomonas aurantiaca*. *Antibiotiki* **20**, 1077–1081 (1975).
 83. Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* **13**, 42–51 (2015).
 84. Jarvis, W. R. & Martone, W. J. Predominant pathogens in hospital infections. *J. Antimicrob. Chemother.* **29**, 19–24 (1992).
 85. Hover, B. M. *et al.* Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat. Microbiol.* **3**, 415–422 (2018).
 86. Moreau, C. S. A practical guide to DNA extraction , PCR , and gene-based DNA sequencing in insects. *Halteres* **5**, 32–42 (2014).
 87. Rubin, B. E. R. *et al.* DNA extraction protocols cause differences in 16S rRNA amplicon sequencing efficiency but not in community profile composition or structure. *Microbiologyopen* **3**, 910–921 (2014).
 88. Birer, C., Tysklind, N., Zinger, L. & Duplais, C. Comparative analysis of DNA extraction methods to study the body surface microbiota of insects: A case study with ant cuticular bacteria. *Mol. Ecol. Resour.* **17**, e34–e45 (2017).
 89. Sanger, F., Nicklen, S. & Coulson, A. . DNA sequencing with chain-terminating. *Proc Natl Acad Sci USA* **74**, 5463–5467 (1977).
 90. OpenStax *et al.* *Microbiology: OpenStax.* (2018).

91. Anderson, S. Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res.* **9**, 3015–3027 (1981).
92. Park, S. T. & Kim, J. Trends in next-generation sequencing and a new era for whole genome sequencing. *Int. Neurol. J.* **20**, 76–83 (2016).
93. Shendure, J. *et al.* Molecular biology: Accurate multiplex polony sequencing of an evolved bacterial genome. *Science (80-.)*. **309**, 1728–1732 (2005).
94. King, C. & Scott-Horton, T. Pyrosequencing: a simple method for accurate genotyping. *J. Vis. Exp.* **11**, 1–1 (2008).
95. Craig, D. W. *et al.* Identification of genetic variants using barcoded multiplexed sequencing. *Nat. Methods* **5**, 887–893 (2008).
96. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 4576–4579 (1990).
97. Field, K. G. *et al.* Molecular phylogeny of the animal kingdom. *Science (80-.)*. **239**, 748–753 (1988).
98. Hebert, P. D., Ratnasingham, S. & DeWaard, J. R. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings. Biol. Sci.* **270**, 96–99 (2003).
99. Peay, K. G., Kennedy, P. G. & Bruns, T. D. Fungal Community Ecology : *October* **58**, 799–810 (2008).
100. Kress, W. J. & Erickson, D. L. DNA barcodes: Genes, genomics, and bioinformatics. *PNAS* **105**, 2761–2762 (2008).
101. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).
102. Creer, S. *et al.* Ultrasequencing of the meiofaunal biosphere: Practice, pitfalls and promises. *Mol. Ecol.* **19**, 4–20 (2010).
103. Stoeck, T., Hayward, B., Taylor, G. T., Varela, R. & Epstein, S. S. A multiple PCR-primer approach to access the microeukaryotic diversity in environmental samples. *Protist* **157**, 31–43 (2006).
104. Lentendu, G. *et al.* Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: A multiple barcoding approach. *Mol. Ecol.* **23**, 3341–3355 (2014).
105. Metsky, H. C. *et al.* Zika virus evolution and spread in the Americas. *Nature* **546**, 411–415 (2017).
106. Lee, S., Abecasis, G. R., Boehnke, M. & Lin, X. Rare-variant association analysis: Study designs and statistical tests. *Am. J. Hum. Genet.* **95**, 5–23 (2014).
107. Hu, Y. *et al.* Herbivorous turtle ants obtain essential nutrients from a conserved nitrogen-recycling gut microbiome. *Nat. Commun.* **9**, 1–14 (2018).
108. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2011).

109. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
110. Hamady, M. & Knight, R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res.* **19**, 1141–1152 (2009).
111. Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J. & Knight, R. Error-correcting barcoded primers allow hundreds of samples to be pyrosequenced in multiplex. *Nat. Methods* **5**, 235–237 (2008).
112. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
113. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
114. Amir, A. *et al.* Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *Am. Soc. Microbiol.* **2**, 1–7 (2017).
115. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **2016**, 1–22 (2016).
116. Bokulich, N. A. *et al.* Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* **10**, 57–59 (2013).
117. Pruesse, E. *et al.* SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
118. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
119. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, 501–504 (2005).
120. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 1–9 (2009).
121. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
122. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
123. R Development Core Team. R: A language and environment for statistical computing. (2008).
124. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 1–14 (2018).
125. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
126. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina

- sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
127. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
 128. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
 129. Mikheenko, A., Saveliev, V. & Gurevich, A. MetaQUAST: Evaluation of metagenome assemblies. *Bioinformatics* **32**, 1088–1090 (2016).
 130. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
 131. Aziz, R. K. *et al.* The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 1–15 (2008).
 132. Markowitz, V. M. *et al.* IMG ER: A system for microbial genome annotation expert review and curation. *Bioinformatics* **25**, 2271–2278 (2009).
 133. Besemer, J. & Borodovsky, M. GeneMark: Web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res.* **33**, 451–454 (2005).
 134. Noguchi, H., Taniguchi, T. & Itoh, T. Meta gene annotator: Detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* **15**, 387–396 (2008).
 135. Lowe, T. M. & Eddy, S. R. TRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1996).
 136. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, (2004).
 137. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **32**, 138–141 (2004).
 138. Glass, E. M., Wilkening, J., Wilke, A., Antonopoulos, D. & Meyer, F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb. Protoc.* **2010**, (2010).
 139. Chen, I. M. A. *et al.* IMG/M: Integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* **45**, D507–D516 (2017).
 140. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
 141. Kang, D. D. *et al.* MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **2019**, 1–13 (2019).
 142. Eren, A. M. *et al.* Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, 1–29 (2015).
 143. Wang, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. MetaCluster-TA: Taxonomic annotation for metagenomic data based on assembly-assisted binning. *BMC Genomics* **15**, 1–9 (2014).
 144. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:

- Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
145. Medema, M. H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339–W346 (2011).
 146. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, 29–37 (2011).
 147. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403–410 (1990).
 148. Shannon, C. E. & Weaver, W. *The mathematical theory of communication.* (1949).
 149. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
 150. Kruskal, W. H. & Wallis, W. A. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* **47**, 583–621 (1952).
 151. Jaccard, P. Nouvelles recherches sur la distribution florale. *Bull. Soc. Vand. Sci. Nat.* **44**, 223–270 (1908).
 152. Sorenson, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. *K. Danske Vidensk. Selsk.* **5**, 1–34 (1948).
 153. Lozupone, C. & Knight, R. UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
 154. Lozupone, C. A., Hamady, M., Kelley, S. T. & Knight, R. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol.* **73**, 1576–1585 (2007).
 155. Anderson, M. J. M. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).
 156. CLARKE, K. R. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18**, 117–143 (1993).
 157. Dixon, P. A package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
 158. Davidson, D. W., Cook, S. C., Snelling, R. R. & Chua, T. H. Explaining the abundance of ants in lowland tropical rainforest canopies. *Science (80-.).* **300**, 969–972 (2003).
 159. Martins, C., Souza, R. F. & Bueno, O. C. Presence and distribution of the endosymbiont *Wolbachia* among *Solenopsis* spp. (Hymenoptera: Formicidae) from Brazil and its evolutionary history. *J. Invertebr. Pathol.* **109**, 287–296 (2012).
 160. De Oliveira, T. B. *et al.* Bacterial communities in the midgut of Ponerine ants (Hymenoptera: Formicidae: Ponerinae). *Sociobiology* **63**, 637–644 (2016).
 161. Moreau, C. S. & Rubin, B. E. R. Diversity and persistence of the gut microbiome of the giant neotropical bullet ant. *Integr. Comp. Biol.* **57**, 682–689 (2017).
 162. Ramalho, M. O., Bueno, O. C. & Moreau, C. S. Microbial composition of spiny ants (Hymenoptera: Formicidae: Polyrhachis) across their geographic range. *BMC Evol. Biol.*

- 17**, 1–15 (2017).
163. Ramalho, M. O., Bueno, O. C. & Moreau, C. S. Species-specific signatures of the microbiome from *Camponotus* and *Colobopsis* ants across developmental stages. *PLoS One* **12**, 1–22 (2017).
 164. Rubin, B. E. R., Kautz, S., Wray, B. D. & Moreau, C. S. Dietary specialization in mutualistic acacia-ants affects relative abundance but not identity of host-associated bacteria. *Mol. Ecol.* **28**, 900–916 (2019).
 165. Ramalho, M. O. *et al.* Development but not diet alters microbial communities in the Neotropical arboreal trap jaw ant *Daceton armigerum*: an exploratory study. *Sci. Rep.* **10**, 1–12 (2020).
 166. Hoffmeister, T. S., Vet, L. E. M., Biere, A., Holsinger, K. & Filser, J. Ecological and evolutionary consequences of biological invasion and habitat fragmentation. *Ecosystems* **8**, 657–667 (2005).
 167. Ricklefs, R. E. Community Diversity: Relative Roles of Local and Regional Processes. *Science (80-.)*. **235**, 167–171 (1987).
 168. Maire, V. *et al.* Habitat filtering and niche differentiation jointly explain species relative abundance within grassland communities along fertility and disturbance gradients. *New Phytol.* **196**, 497–509 (2012).
 169. Arthur, W. *The niche in competition and evolution.* (1987).
 170. Hickman, J. C. Energy allocation and niche differentiation in four co-existing annual species of *Polygonum* in western North America. *J. Ecol.* 317–326 (1977).
 171. Konopka, A. What is microbial community ecology. *ISME J.* **3**, 1223–1230 (2009).
 172. Allison, S. D. & Martiny, J. B. H. Resistance, resilience, and redundancy in microbial communities. *PNAS* **105**, 11512–11519 (2008).
 173. Hollister, E. B. *et al.* Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. *ISME J.* **4**, 829–838 (2010).
 174. Raes, J., Letunic, I., Yamada, T., Jensen, L. J. & Bork, P. Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Mol. Syst. Biol.* **7**, 1–9 (2011).
 175. Liu, S. *et al.* pH levels drive bacterial community structure in the Qiantang River as determined by 454 pyrosequencing. *Front. Microbiol.* **6**, 1–7 (2015).
 176. Sunagawa, S. *et al.* Structure and Function of the Global Ocean Microbiome. *Science (80-.)*. **348**, (2015).
 177. Stres, B. *et al.* Influence of temperature and soil water content on bacterial, archaeal and denitrifying microbial communities in drained fen grassland soil microcosms. *FEMS Microbiol. Ecol.* **66**, 110–122 (2008).
 178. Nicol, G. W., Leininger, S., Schleper, C. & Prosser, J. I. The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. *Environ. Microbiol.* **10**, 2966–2978 (2008).
 179. Kuffner, M. *et al.* Effects of season and experimental warming on the bacterial

- community in a temperate mountain forest soil assessed by 16S rRNA gene pyrosequencing. *FEMS Microbiol. Ecol.* **82**, 551–562 (2012).
180. Shen, C. *et al.* Soil pH drives the spatial distribution of bacterial communities along elevation on Changbai Mountain. *Soil Biol. Biochem.* **57**, 204–211 (2013).
 181. Hosokawa, T., Kikuchi, Y., Nikoh, N., Shimada, M. & Fukatsu, T. Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PLoS Biol.* **4**, 1841–1851 (2006).
 182. Jiggins, F. M. & Hurst, G. D. D. Rapid insect evolution by symbiont transfer. *Science* (80-). **332**, 185–186 (2011).
 183. Warnecke, F. *et al.* Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**, 560–565 (2007).
 184. Akman Gündüz, E. & Douglas, A. E. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proc. R. Soc. B Biol. Sci.* **276**, 987–991 (2012).
 185. Raymann, K. & Moran, N. A. The role of the gut microbiome in health and disease of adult honey bee workers. *Curr. Opin. Insect Sci.* **26**, 97–104 (2018).
 186. Block, W., Erzincliglu, Y. Z. & Worland, M. R. Cold resistance in all life stages of two blowfly species (Diptera, Calliphoridae). *Med. Vet. Entomol.* **4**, 213–219 (1990).
 187. Hückesfeld, S., Niederegger, S., Schlegel, P., Heinzl, H.-G. & Spiess, R. Feel the heat : The effect of temperature on development, behavior and central pattern generation in 3rd instar *Calliphora vicina* larvae. *J. Insect Physiol.* **57**, 136–146 (2010).
 188. Peccia, J., Werth, H. M., Shelly, M. & Mark, H. Effects of relative humidity on the ultraviolet induced inactivation of airborne bacteria. *Aerosol Sci. Technol.* **35**, 728–740 (2001).
 189. Engl, T. *et al.* Ancient symbiosis confers desiccation resistance to stored grain pest beetles. *Mol. Ecol.* **27**, 2095–2108 (2018).
 190. Leonardo, T. E. & Mondor, E. B. Symbiont modifies host life-history traits that affect gene flow. *Proc. R. Soc. B Biol. Sci.* **273**, 1079–1084 (2006).
 191. Goodacre, S. L. *et al.* Microbial modification of host long-distance dispersal capacity. *BMC Biol.* **7**, 1–8 (2009).
 192. Funaro, C. F. *et al.* Army ants harbor a host-specific clade of Entomoplasmatales bacteria. *Appl. Environ. Microbiol.* **77**, 346–350 (2011).
 193. Łukasik, P. *et al.* The structured diversity of specialized gut symbionts of the New World army ants. *Mol. Ecol.* **26**, 3808–3825 (2017).
 194. Sanders, J. G. *et al.* Stability and phylogenetic correlation in gut microbiota: Lessons from ants and apes. *Mol. Ecol.* **23**, 1268–1283 (2014).
 195. Wernegreen, J. J., Kauppinen, S. N., Brady, S. G. & Ward, P. S. One nutritional symbiosis begat another: Phylogenetic evidence that the ant tribe Camponotini acquired *Blochmannia* by tending sap-feeding insects. *BMC Evol. Biol.* **9**, (2009).
 196. Corby-Harris, V. *et al.* *Parasaccharibacter apium*, gen. Nov., sp. Nov., Improves Honey Bee (Hymenoptera: Apidae) resistance to *Nosema*. *J. Econ. Entomol.* **109**, 537–543

- (2016).
197. Mitra, E., Stathopoulos, S., Siden-Kiamos, I., Christophides, G. K. & Louis, C. Asaia accelerates larval development of *Anopheles gambiae*. *Pathog. Glob. Health* **107**, 305–311 (2013).
 198. Dussutour, A. & Simpson, S. J. Ant workers die young and colonies collapse when fed a high-protein diet. *Proc. R. Soc. B Biol. Sci.* **279**, 2402–2408 (2012).
 199. Neuvonen, M. M. *et al.* The genome of Rhizobiales bacteria in predatory ants reveals urease gene functions but no genes for nitrogen fixation. *Sci. Rep.* **6**, 1–11 (2016).
 200. Wenseleers, T. & Billen, J. No evidence for Wolbachia-induced parthenogenesis in the social Hymenoptera. *J. Evol. Biol.* **13**, 277–280 (2000).
 201. Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X. Y. & Fukatsu, T. Wolbachia as a bacteriocyte-associated nutritional mutualist. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 769–774 (2010).
 202. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013).
 203. Bokulich, N. *et al.* Optimizing taxonomic classification of marker gene amplicon sequences. *PeerJ Prepr.* (2018) doi:10.7287/peerj.preprints.3208v2.
 204. Team, R. C. R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria* (2019).
 205. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
 206. Hammer, Ø., Harper, D. A. T. & Ryan, P. D. Past: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **4**, (2001).
 207. Kassambara, A. ggpubr: ‘ggplot2’ Based Publication Ready Plots. *R Packag. version 0.3.0.* (2020).
 208. Pohlert, T. The Pairwise Multiple Comparison of Mean Ranks Sums. *R Packag.* (2014).
 209. Gómez-Rubio, V. ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). *J. Stat. Softw.* **77**, 3–5 (2017).
 210. Tang, Y., Horikoshi, M. & Li, W. Ggfortify: Unified interface to visualize statistical results of popular r packages. *R J.* **8**, 478–489 (2016).
 211. Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
 212. Green, J. L., Bohannan, B. J. M. & Whitaker, R. J. Microbial biogeography: From taxonomy to traits. *Science (80-.)*. **320**, 1039–1043 (2008).
 213. Shaffer, B. & Lighthart, B. Survey of Culturable Airborne Bacteria at Four Diverse Locations in Oregon: Urban, Rural, Forest, and Coastal. *Microb. Ecol.* **34**, 167–177 (1997).
 214. Song, S. J. *et al.* Is there convergence of gut microbes in blood-feeding vertebrates? *Philos. Trans. R. Soc. B Biol. Sci.* **374**, (2019).
 215. Muegge, B. D. *et al.* Diet drives convergence in gut microbiome functions across

- mammalian phylogeny and within humans. *Science (80-.)*. **332**, 970–974 (2011).
216. Sanders, J. G. *et al.* Dramatic differences in gut bacterial densities correlate with diet and habitat in rainforest ants. *Integr. Comp. Biol.* **57**, 705–722 (2017).
 217. Nikoh, N. *et al.* Evolutionary origin of insect-Wolbachia nutritional mutualism. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10257–10262 (2014).
 218. Bisch, G. *et al.* Genome evolution of Bartonellaceae symbionts of ants at the opposite ends of the trophic scale. *Genome Biol. Evol.* **10**, 1687–1704 (2018).
 219. Russell, J. A. *et al.* Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *Proc. Natl. Acad. Sci.* **106**, 21236–21241 (2009).
 220. Łukasik, P. *et al.* The structured diversity of specialized gut symbionts of the New World army ants. *Mol. Ecol.* **26**, 3808–3825 (2017).
 221. Ley, R. *et al.* Evolution of mammals and their gut microbes. *Science (80-.)*. **320**, 1647–1651 (2008).
 222. Nelson, T. M., Rogers, T. L. & Brown, M. V. The gut bacterial community of mammals from marine and terrestrial habitats. *PLoS One* **8**, 1–8 (2013).
 223. Papadimitriou, K., Pot, B. & Tsakalidou, E. How microbes adapt to a diversity of food niches. *Curr. Opin. Food Sci.* **2**, 29–35 (2015).
 224. Macfarlane, G. T., Steed, H. & Macfarlane, S. Bacterial metabolism and health-related effects of galacto-oligosaccharides and other prebiotics. *J. Appl. Microbiol.* **104**, 305–344 (2008).
 225. Ramirez-Farias, C. *et al.* Effect of inulin on the human gut microbiota: stimulation of *Bifidobacterium adolescentis* and *Faecalibacterium prausnitzii*. *Br. J. Nutr.* **101**, 541–550 (2009).
 226. Ivarsson, E., Roos, S., Liu, H. Y. & Lindberg, J. E. Fermentable non-starch polysaccharides increases the abundance of *Bacteroides-Prevotella-Porphyromonas* in ileal microbial community of growing pigs. *Animal* **8**, 1777–1787 (2014).
 227. Chung, W. S. F. *et al.* Modulation of the human gut microbiota by dietary fibres occurs at the species level. *BMC Biol.* **14**, 1–13 (2016).
 228. Greiner, T. & Bäckhed, F. Effects of the gut microbiota on obesity and glucose homeostasis. *Trends Endocrinol. Metab.* **22**, 117–123 (2011).
 229. Hang, I. *et al.* Effect of high contents of dietary animal-derived protein or carbohydrates on canine faecal microbiota. *BMC Vet. Res.* **8**, (2012).
 230. Magnusson, K. R. *et al.* Relationships between diet-related changes in the gut microbiome and cognitive flexibility. *Neuroscience* **300**, 128–140 (2015).
 231. Moreau, C. S. Symbioses among ants and microbes. *Curr. Opin. Insect Sci.* **39**, 1–5 (2020).
 232. Madigosky, S. R. Tropical microclimatic considerations. in *Forest canopies* (eds. Lowman, M. D. & Rinker, H. B.) 24–48 (Elsevier Academic Press, 2004).
 233. Kumagai, T. *et al.* Vertical Profiles of Environmental Factors within Tropical Rainforest, Lambir Hills National Park, Sarawak, Malaysia. *J. For. Res.* **6**, 257–264 (2001).

234. Basset, Y., Hammond, P. M., Barrios, H., Holloway, J. D. & Miller, S. E. Vertical stratification of arthropod assemblages. in *Arthropods of Tropical Forests: Spatio-Temporal Dynamics and Resource use in the Canopy* (eds. Basset, Y., Novotny, V., Miller, S. E. & Kitching, R. L.) 17–27 (Cambridge University Press., 2003).
235. Gruntenko, N. E. *et al.* Various Wolbachia genotypes differently influence host *Drosophila* dopamine metabolism and survival under heat stress conditions. *BMC Evol. Biol.* **17**, (2017).
236. Burke, G., Fiehn, O. & Moran, N. Effects of facultative symbionts and heat stress on the metabolome of pea aphids. *ISME J.* **4**, 242–252 (2010).
237. Rio, R. V. M., Wu, Y. N., Filardo, G. & Aksoy, S. Dynamics of multiple symbiont density regulation during host development: Tsetse fly and its microbial flora. *Proc. R. Soc. B Biol. Sci.* **273**, 805–814 (2006).
238. Gibbs, A. G. & Rajpurohit, S. Cuticular lipids and water balance. in *Insect hydrocarbons biology, biochemistry, and chemical ecology* (eds. Blomquist, G. J. & Bagnères, A. G.) 100–120 (2010).
239. Anbutsu, H. *et al.* Small genome symbiont underlies cuticle hardness in beetles. *Proc. Natl. Acad. Sci.* **114**, E8382–E8391 (2017).
240. Vigneron, A. *et al.* Insects recycle endosymbionts when the benefit is over. *Curr. Biol.* **24**, 2267–2273 (2014).
241. Lapinski, W. & Tschapka, M. Desiccation resistance reflects patterns of microhabitat choice in a Central American assemblage of wandering spiders. *J. Exp. Biol.* **217**, 2789–2795 (2014).
242. HOOD, W. G. & TSCHINKEL, W. R. Desiccation resistance in arboreal and terrestrial ants. *Physiol. Entomol.* **15**, 23–35 (1990).
243. Stacey, G. The Rhizobium-legume nitrogen-fixing symbiosis. in *Biology of the nitrogen cycle* 147–163 (Elsevier, 2007).
244. Chen, Y. S., Yanagida, F. & Shinohara, I. Isolation and identification of lactic acid bacteria from soil using an enrichment procedure. *Lett. Appl. Microbiol.* **40**, 195–200 (2005).
245. Yanagida, F., Chen, Y. S. & Shinohara, T. Searching for bacteriocin-producing lactic acid bacteria in soil. *J. Gen. Appl. Microbiol.* **52**, 21–28 (2006).
246. Wolschin, F., Hölldobler, B., Gross, R. & Zientz, E. Replication of the endosymbiotic bacterium *Blochmannia floridanus* is correlated with the developmental and reproductive stages of its ant host. **70**, 4096–4102 (2004).
247. Zientz, E., Beyaert, I., Gross, R. & Feldhaar, H. Relevance of the endosymbiosis of *Blochmannia floridanus* and carpenter ants at different stages of the life cycle of the host. **72**, 6027–6033 (2006).
248. Gómez-Valero, L. *et al.* Patterns and rates of nucleotide substitution, insertion and deletion in the endosymbiont of ants *Blochmannia floridanus*. *Mol. Ecol.* **17**, 4382–4392 (2008).
249. Degnan, P. H., Lazarus, A. B., Brock, C. D. & Wernegreen, J. J. Host-symbiont stability and fast evolutionary rates in an ant-bacterium association: Cospeciation of

- Camponotus species and their endosymbionts, Candidatus blochmannia. *Syst. Biol.* **53**, 95–110 (2004).
250. Brown, B. P. & Wernegreen, J. J. Deep divergence and rapid evolutionary rates in gut-associated Acetobacteraceae of ants. *BMC Microbiol.* **16**, 1–17 (2016).
 251. Stouthamer, R., Breeuwer, J. A. J. & Hurst, G. D. D. Wolbachia pipientis: Microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* **53**, 71–102 (1999).
 252. Russell, J. A. *et al.* Specialization and geographic isolation among Wolbachia symbionts from ants and Lycaenid butterflies. *Evolution (N. Y.)* **63**, 624–640 (2009).
 253. Kautz, S., Rubin, B. E. R. & Moreau, C. S. Bacterial infections across the ants: Frequency and prevalence of Wolbachia, Spiroplasma, and Asaia. *Psyche (London)* **2013**, (2013).
 254. Eilmus, S. & Heil, M. Bacterial associates of arboreal ants and their putative functions in an obligate ant-plant mutualism. *Appl. Environ. Microbiol.* **75**, 4324–4332 (2009).
 255. Moreau, C. S. & Rubin, B. E. R. Diversity and persistence of the gut microbiome of the giant neotropical bullet ant. *Integr. Comp. Biol.* **57**, 682–689 (2017).
 256. Larson, H. K. *et al.* Distribution and dietary regulation of an associated facultative Rhizobiales-related bacterium in the omnivorous giant tropical ant, Paraponera clavata. *Naturwissenschaften* **101**, 397–406 (2014).
 257. Pyle, R., Bentzien, M. & Opler, P. Insect conservation. *Annu. Rev. Entomol.* **26**, 233–258 (1981).
 258. McIntyre, N. Ecology of urban arthropods: a review and a call to action. *Ann. Entomol. Soc. Am.* **93**, 271–276 (2009).
 259. Belt, K. T., Hohn, C., Gbakima, A. & Higgins, J. A. Identification of culturable stream water bacteria from urban, agricultural, and forested watersheds using 16S rRNA gene sequencing. *J. Water Health* **5**, 395–406 (2007).
 260. Reeson, A. F., Jankovic, T., Kasper, M. L., Rogers, S. & Austin, A. D. Application of 16S rDNA-DGGE to examine the microbial ecology associated with a social wasp *Vespa germanica*. *Insect Mol. Biol.* **12**, 85–91 (2003).
 261. Overzier, E. *et al.* Anaplasma phagocytophilum in questing ixodes ricinus ticks: Comparison of prevalences and partial 16S rRNA gene variants in urban, pasture, and natural habitats. *Appl. Environ. Microbiol.* **79**, 1730–1734 (2013).
 262. Poinar, G. Nematode parasites and associates of ants: Past and present. *Psyche (London)* **2012**, (2012).
 263. Santos, A. V., Dillon, R. J., Dillon, V. M., Reynolds, S. E. & Samuels, R. I. Occurrence of the antibiotic producing bacterium Burkholderia sp. in colonies of the leaf-cutting ant *Atta sexdens rubropilosa*. *FEMS Microbiol. Lett.* **239**, 319–323 (2004).
 264. Miranda, V. A., Navarro, P. D., Davidowitz, G., Bronstein, J. & Stock, S. P. Effect of insect host age and diet on the fitness of the entomopathogenic nematode-bacteria mutualism. *Symbiosis* **61**, 145–153 (2013).
 265. Bates, S. T. *et al.* Global biogeography of highly diverse protistan communities in soil. *ISME J.* **7**, 652–659 (2013).

266. Lawley, B., Ripley, S., Bridge, P. & Convey, P. Molecular analysis of geographic patterns of eukaryotic diversity in antarctic soils. *Appl. Environ. Microbiol.* **70**, 5963–5972 (2004).
267. Fell, J. W., Scorzetti, G., Connell, L. & Craig, S. Biodiversity of micro-eukaryotes in Antarctic Dry Valley soils with <5% soil moisture. *Soil Biol. Biochem.* **38**, 3107–3119 (2006).
268. Tian, F. *et al.* Bacterial, archaeal and eukaryotic diversity in Arctic sediment as revealed by 16S rRNA and 18S rRNA gene clone libraries analysis. *Polar Biol.* **32**, 93–103 (2009).
269. Zhao, H. *et al.* Drivers and assemblies of soil eukaryotic microbes among different soil habitat types in a semi-arid mountain in China. *PeerJ* **2018**, (2018).
270. Richards, T. A., Vepritskiy, A. A., Gouliamova, D. E. & Nierzwicki-Bauer, S. A. The molecular diversity of freshwater picoeukaryotes from an oligotrophic lake reveals diverse, distinctive and globally dispersed lineages. *Environ. Microbiol.* **7**, 1413–1425 (2005).
271. Lopez-Garcia, P., Rodriguez-Valera, F., Pedros-Allo, C. & Moreira, D. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**, 603–607 (2001).
272. Karling, J. S. CHYTRIDIOSIS OF SCALE INSECTS. *Am. J. Bot.* **35**, 246–254 (1948).
273. Evans, H. C. Mycopathogens of insects of epigeal and aerial habitats. *Insect-Fungus Interact.* 205–238 (1989).
274. Martin, W. W. Two Additional Species of Catenaria (Chytridiomycetes, Blastocladiales) Parasitic in Midge Eggs. *Mycologia* **70**, 461–467 (1978).
275. Matsuura, K., Tanaka, C. & Nishida, T. Symbiosis of a termite and a sclerotium-forming fungus: Sclerotia mimic termite eggs. *Ecol. Res.* **15**, 405–414 (2000).
276. Evans, H. C. & Samson, R. A. Sporodiniella umbellata, an entomogenous fungus of the Mucorales from cocoa farms in Ecuador. *Can. J. Bot.* **55**, 2981–2984 (1977).
277. Horn, B. W. & Lichtwardt, R. W. Studies on the Nutritional Relationship of Larval Aedes Aegypti (Diptera: Culicidae) with Smittium Culisetae (Trichomycetes). *Mycologia* **73**, 724–740 (1981).
278. Roberts, D. W. & Humber, R. A. Entomogenous Fungi. in *Biology of Conidial Fungi* 201–236 (Elsevier, 1981). doi:10.1016/B978-0-12-179502-3.50014-5.
279. Humber, R. A. Entomophthoromycota: A new phylum and reclassification for entomophthoroid fungi. *Mycotaxon* **120**, 477–492 (2012).
280. Armer, C. A., Rao, S. & Berry, R. E. Armer, C. A., Rao, S., & Berry, R. E. (2004). Insect cellular and chemical limitations to pathogen development: the Colorado potato beetle, the nematode Heterorhabditis marelatus, and its symbiotic bacteria. *J. Invertebr. Pathol.* **87**, 114–122 (2004).
281. Taylor, D. B., Szalanski, A. L., Adams, B. J. & Peterson, R. D. Susceptibility of House Fly (Diptera: Muscidae) Larvae to Entomopathogenic Nematodes (Rhabditida: Heterorhabditidae, Steinernematidae). *Environ. Entomol.* **27**, 1514–1519 (1998).
282. Li, X.-Y., Cowles, R. S., Cowles, E. A., Gaugler, R. & Cox-Foster, D. L. Li, X. Y., Cowles, R. S., Cowles, E. A., Gaugler, R., & Cox-Foster, D. L. (2007). Relationship between the

- successful infection by entomopathogenic nematodes and the host immune response. *Int. J. Parasitol.* **37**, 365–374 (2007).
283. van Sambeek, J. & Wiesner, A. van Sambeek, J., & Wiesner, A. (1999). Successful parasitization of locusts by entomopathogenic nematodes is correlated with inhibition of insect phagocytes. *J. Invertebr. Pathol.* **73**, 154–161 (1999).
 284. Zhao, X., Ferdig, M. T., Li, J. & Christensen, B. M. Biochemical pathway of melanotic encapsulation of *Brugia malayi* in the mosquito, *Armigeres subalbatus*. *Dev. Comp. Immunol.* **19**, 205–215 (1995).
 285. Aliota, M. T., Chen, C.-C., Dagoro, H., Fuchs, J. F. & Christensen, B. M. Filarial worms reduce *Plasmodium* infectivity in mosquitoes. *PLoS Negl Trop Dis* **5**, e963 (2011).
 286. Vegni Talluri, M. & Cancrini, G. A N ULTRASTRUCTURAL STUDY ON THE EARLY CELLULAR RESPONSE TO *DIROFILARIA IMMITIS* (NEMATODA) IN THE MALPIGHIA N TUBULES OF *Aedes aegypti* (refractory strains). *Parasite* **1**, 343–348 (1994).
 287. Cupp, M. S., Chen, Y. & Cupp, E. W. Cellular Hemolymph Response of *Simulium vittatum* (Diptera: Simuliidae) to Intrathoracic Injection of *Onchocerca lienalis* (Filarioidea: Onchocercidae) Microfilariae. *J. Med. Entomol.* **34**, 56–63 (1997).
 288. Bartholomay, L. C. *et al.* NIH Public Access. **330**, 88–90 (2011).
 289. Seymour, R. & Briggs, J. D. Occurrence and control of *Aphanomyces* (Saprolegniales: fungi) infections in laboratory colonies of larval Anopheles. *J. Am. Mosq. Control Assoc.* **1**, 100–102 (1985).
 290. Biedermann, P. H. W. & Vega, F. E. Ecology and Evolution of Insect – Fungus Mutualisms. *Annu. Rev. Entomol.* **65**, 22.1–22.25 (2020).
 291. Kobune, S., Kajimura, H., Masuya, H. & Kubono, T. Symbiotic fungal flora in leaf galls induced by *Illiciomyia yukawai* (Diptera: Cecidomyiidae) and in its mycangia. *Microb. Ecol.* **63**, 619–627 (2012).
 292. Stefanini, I. Yeast-insect associations: It takes guts. *Yeast* **35**, 315–330 (2018).
 293. Goodrich-Blair, H. & Clarke, D. J. Mutualism and pathogenesis in *Xenorhabdus* and *Photorhabdus*: Two roads to the same destination. *Mol. Microbiol.* **64**, 260–268 (2007).
 294. Shaffer, M., Thurimella, K. & Lozupone, C. A. SCNIC: Sparse correlation network investigation for compositional data. *bioRxiv* (2020) doi:10.1101/2020.11.13.380733.
 295. Friedman, J. & Alm, E. J. Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput. Biol.* **8**, 1–11 (2012).
 296. Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
 297. Meadow, J. F. & Zabinski, C. A. Spatial heterogeneity of eukaryotic microbial communities in an unstudied geothermal diatomaceous biological soil crust: Yellowstone National Park, WY, USA. *FEMS Microbiol. Ecol.* **82**, 182–191 (2012).
 298. Kaya, H. K. & Thurston, G. S. Soil microorganisms affecting entomopathogenic nematodes. in *Nematodes and the biological control of insect pests* 97–104 (1993).
 299. Lara, E., Berney, C., Harms, H. & Chatzinotas, A. Cultivation-independent analysis

- reveals a shift in ciliate 18S rRNA gene diversity in a polycyclic aromatic hydrocarbon-polluted soil. *FEMS Microbiol. Ecol.* **62**, 365–373 (2007).
300. Bridge, P. D., Roberts, P. J., Spooner, B. M. & Panchal, G. On the unreliability of published DNA sequences. *New Phytol.* **160**, 43–48 (2003).
 301. Marcial Gomes, N. C. *et al.* Dynamics of Fungal Communities in Bulk and Maize Rhizosphere Soil in the Tropics. *Appl. Environ. Microbiol.* **69**, 5737 (2003).
 302. He, J., Xu, Z. & Hughes, J. Analyses of soil fungal communities in adjacent natural forest and hoop pine plantation ecosystems of subtropical Australia using molecular approaches based on 18S rRNA genes. *FEMS Microbiol. Lett.* **247**, 91–100 (2005).
 303. Voglmayr, H. *et al.* The diversity of ant-associated black yeasts: Insights into a newly discovered world of symbiotic interactions. *Fungal Biol.* **115**, 1077–1091 (2011).
 304. Poinar, G. & Yanoviak, S. P. *Myrmeconema neotropicum* n. g., n. sp., a new tetradonematid nematode parasitising South American populations of *Cephalotes atratus* (Hymenoptera: Formicidae), with the discovery of an apparent parasite-induced host morph. *Syst. Parasitol.* **69**, 145–153 (2008).
 305. Yanoviak, S. P., Kaspari, M., Dudley, R. & Poinar, G. Parasite-induced fruit mimicry in a tropical canopy ant. *Am. Nat.* **171**, 536–544 (2008).
 306. Ekstam, B., Johansson, B., Dinnézt, P. & Ellström, P. Predicting risk habitats for the transmission of the small liver fluke, *Dicrocoelium dendriticum* to grazing ruminants. *Geospat. Health* **6**, 125–131 (2011).
 307. Manga-González, M. Y., González-Lanza, C., Cabanas, E. & Campo, R. Contributions to and review of dicrocoeliosis, with special reference to the intermediate hosts of *Dicrocoelium dendriticum*. *Parasitology* **123**, (2001).
 308. Olsen, O. W. *Animal parasites: their life cycles and ecology.* (1986).
 309. Benjamin, R. K. *et al.* Insect- and other arthropod-associated fungi. in *Biodiversity of Fungi: Inventory and Monitoring Methods* 395–433 (Elsevier, Academic Press, 2004).
 310. Humber, R. A. Evolution of entomopathogenicity in fungi. *J. Invertebr. Pathol.* **98**, 262–266 (2008).
 311. Vega, F. E. & Dowd, P. F. The role of yeasts as insect endosymbionts. *Insect–fungal Assoc. Ecol. Evol.* 211–243 (2005).
 312. Poinar, G. O. Diplogasterid nematodes (diplogasteridae: Rhabditida) and their relationship to insect disease. *J. Invertebr. Pathol.* **13**, 447–454 (1969).
 313. Koppenhöfer, A. M., Kaya, H. K. & Taormino, S. P. Infectivity of entomopathogenic nematodes (Rhabditida: Steinernematidae) at different soil depths and moistures. *Journal of Invertebrate Pathology* vol. 65 193–199 (1995).
 314. Hashmi, G. & Gaugler, R. Genetic Diversity in Insect-Parasitic Nematodes (Rhabditida: Heterorhabditidae). *J. Invertebr. Pathol.* **72**, 185–189 (1998).
 315. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn’s disease. *Cell Host Microbe* **15**, 382–392 (2014).
 316. Wang, H. *et al.* *Mycoplasma pneumoniae* and *Streptococcus pneumoniae* caused

- different microbial structure and correlation network in lung microbiota. *J. Thorac. Dis.* **8**, 1316–1322 (2016).
317. Sokol, H. *et al.* Specificities of the intestinal microbiota in patients with inflammatory bowel disease and *Clostridium difficile* infection. *Gut Microbes* **9**, 55–60 (2018).
 318. Faust, K. *et al.* Cross-biome comparison of microbial association networks. *Front. Microbiol.* **6**, 1–13 (2015).
 319. Weng, Y. J. *et al.* Correlation of diet, microbiota and metabolite networks in inflammatory bowel disease. *J. Dig. Dis.* **20**, 447–459 (2019).
 320. Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351 (2012).
 321. Binetruy, F. *et al.* Microbial community structure reveals instability of nutritional symbiosis during evolutionary radiation of *Amblyomma* ticks. *Mol. Ecol.* **29**, 1016–1029 (2020).
 322. Hansen, A. K. & Moran, N. A. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2849–2854 (2011).
 323. Boucias, D. G. *et al.* The hindgut lumen prokaryotic microbiota of the termite *Reticulitermes flavipes* and its responses to dietary lignocellulose composition. *Mol. Ecol.* **22**, 1836–1853 (2013).
 324. Zheng, H. *et al.* Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *PNAS* **116**, 25909–25916 (2019).
 325. Scarborough, C. L., Ferrari, J. & Godfray, H. C. J. Aphid protected from pathogen by endosymbiont. *Science (80-.)*. **310**, 1781–1781 (2005).
 326. Hedges, L. M., Brownlie, J. C., O’Neill, S. L. & Johnson, K. N. Wolbachia and virus protection in insects. *Science (80-.)*. **322**, 702–702 (2008).
 327. Flórez, L. V. *et al.* An antifungal polyketide associated with horizontally acquired genes supports symbiont-mediated defense in *Lagria villosa* beetles. *Nat. Commun.* **9**, 1–10 (2018).
 328. Matarrita-Carranza, B. *et al.* Evidence for widespread associations between neotropical hymenopteran insects and Actinobacteria. *Front. Microbiol.* **8**, 1–17 (2017).
 329. Ziemert, N., Alanjary, M. & Weber, T. The evolution of genome mining in microbes – a review. *Nat. Prod. Rep.* **33**, 988–1005 (2016).
 330. Adamek, M., Spohn, M., Stegmann, E. & Ziemert, N. Mining Bacterial Genomes for Secondary Metabolite Gene Clusters. in *Antibiotics: Methods and Protocols* (ed. Sass, P.) vol. 1520 23–47 (Springer Science+Business Media, 2017).
 331. Crits-Christoph, A., Diamond, S., Butterfield, C. N., Thomas, B. C. & Banfield, J. F. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* **558**, 440–444 (2018).
 332. Paulus, C. *et al.* New natural products identified by combined genomics-metabolomics profiling of marine *Streptomyces* sp. MP131-18. *Sci. Rep.* **7**, 1–11 (2017).

333. Beemelmans, C., Guo, H., Rischer, M. & Poulsen, M. Natural products from microbes associated with insects. *Beilstein J. Org. Chem.* **12**, 314–327 (2016).
334. Kroiss, J. *et al.* Symbiotic streptomycetes provide antibiotic combination prophylaxis for wasp offspring. *Nat. Chem. Biol.* **6**, 261–263 (2010).
335. Cuadrat, R. R. C., Ionescu, D., Dávila, A. M. R. & Grossart, H. P. Recovering genomics clusters of secondary metabolites from lakes using genome-resolved metagenomics. *Front. Microbiol.* **9**, 1–13 (2018).
336. Waterworth, S. C. *et al.* Horizontal gene transfer to a defensive symbiont with a reduced genome amongst a multipartite beetle microbiome. *MBio* **11**, 1–23 (2020).
337. Altamia, M. A. *et al.* Secondary metabolism in the gill microbiota of shipworms (teredinidae) as revealed by comparison of metagenomes and nearly complete symbiont genomes. *mSystems* **5**, 1–18 (2020).
338. Donia, M. S., Fricke, W. F., Ravel, J. & Schmidt, E. W. Variation in tropical reef symbiont metagenomes defined by secondary metabolism. *PLoS One* **6**, 1–11 (2011).
339. Wilson, M. C. & Piel, J. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem. Biol.* **20**, 636–647 (2013).
340. Ashen, J. B. & Goff, L. J. Molecular and ecological evidence for species specificity and coevolution in a group of marine algal-bacterial symbioses. *Appl. Environ. Microbiol.* **66**, 3024–3030 (2000).
341. Edlund, A., Loesgen, S., Fenical, W. & Jensen, P. R. Geographic distribution of secondary metabolite genes in the marine actinomycete *Salinispora arenicola*. *Appl. Environ. Microbiol.* **77**, 5916–5925 (2011).
342. Duplais, C. *et al.* Gut bacteria are essential for normal cuticle development in herbivorous turtle ants. *Nat. Commun.* **12**, 1–6 (2021).
343. Chen, I. M. A. *et al.* IMG/M v.5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **47**, D666–D677 (2019).
344. Price, S. L. *et al.* Renewed diversification is associated with new ecological opportunity in the Neotropical turtle ants. *J. Evol. Biol.* **27**, 242–258 (2014).
345. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
346. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).
347. Ziemert, N. *et al.* The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* **7**, 1–9 (2012).
348. Konstantinidis, K. T. & Tiedje, J. M. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2567–2572 (2005).
349. Varghese, N. J. *et al.* Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* **43**, 6761–6771 (2015).

350. Novichkov, V., Kaznadzey, A., Alexandrova, N. & Kaznadzey, D. NSimScan: DNA comparison tool with increased speed, sensitivity and accuracy. *Bioinformatics* **32**, 2380–2381 (2016).
351. Wang, Q. *et al.* Genomic identification of rhizobia-related strains and threshold of ANI and core-genome for family, genus and species. *Int. J. Environ. Agric. Res. ISSN* **2**, 76–86 (2016).
352. Bushnell, B., Rood, J. & Singer, E. BBMerge – Accurate paired shotgun read merging via overlap. *PLoS One* **12**, 1–15 (2017).
353. Yeong, B. M. *BiG-SCAPE: exploring biosynthetic diversity through gene cluster similarity networks*. https://git.wageningenur.nl/yeong001/BGC_networks (2016).
354. Navarro-Munoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
355. Schorn, M. A. *et al.* Sequencing rare marine actinomycete genomes reveals high density of unique natural product biosynthetic gene clusters. *Microbiol. (United Kingdom)* **162**, 2075–2086 (2016).
356. Adamek, M. *et al.* Comparative genomics reveals phylogenetic distribution patterns of secondary metabolites in *Amycolatopsis* species. *BMC Genomics* **19**, 1–15 (2018).
357. Morris, J. H. *et al.* ClusterMaker: A multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics* **12**, 1–14 (2011).
358. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
359. Price, M. N., Dehal, P. S. & Arkin, A. P. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
360. Schöner, T. A. *et al.* Aryl Polyenes, a Highly Abundant Class of Bacterial Aryl Polyenes, a Highly Abundant Class of Bacterial Natural Products, Are Functionally Related to Antioxidative Carotenoids. *ChemBioChem* **17**, 247–253 (2016).
361. Zhang, W., Li, Z., Miao, X. & Zhang, F. The screening of antimicrobial bacteria with diverse novel nonribosomal peptide synthetase (NRPS) genes from South China sea sponges. *Mar. Biotechnol.* **11**, 346–355 (2009).
362. Shevchuk, O. *et al.* Polyketide synthase (PKS) reduces fusion of *Legionella pneumophila*-containing vacuoles with lysosomes and contributes to bacterial competitiveness during infection. *Int. J. Med. Microbiol.* **304**, 1169–1181 (2014).
363. Chaturvedi, K. S., Hung, C. S., Crowley, J. R., Stapleton, A. E. & Henderson, J. P. The siderophore yersiniabactin binds copper to protect pathogens during infection. *Nat. Chem. Biol.* **8**, 731–736 (2012).
364. Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 1–15 (2007).
365. Jenke-Kodama, H. & Dittmann, E. Evolution of metabolic diversity: Insights from microbial polyketide synthases. *Phytochemistry* **70**, 1858–1866 (2009).
366. Rokas, A., Mead, M. E., Steenwyk, J. L., Raja, H. A. & Oberlies, N. H. Biosynthetic gene

- clusters and the evolution of fungal chemodiversity. *Nat. Prod. Rep.* **37**, 868–878 (2020).
367. Storey, M. A. *et al.* Metagenomic Exploration of the Marine Sponge *Mycale hentscheli* Uncovers Multiple Polyketide-Producing Bacterial Symbionts. *MBio* **11**, 1–16 (2020).
368. Thompson, J. N. Coevolution: The Geographic Mosaic of Coevolutionary Arms Races. *Curr. Biol.* **15**, 992–994 (2005).
369. Nuismer, S. L. Parasite Local Adaptation in a Geographic Mosaic. *Evolution (N. Y.)*. **60**, 24–30 (2006).
370. Caldera, E. J., Chevrette, M. G., McDonald, B. R. & Currie, C. R. Local Adaptation of Bacterial Symbionts within a Geographic Mosaic of Antibiotic Coevolution. *Appl. Environ. Microbiol.* **85**, 1–13 (2019).
371. Itoh, H. *et al.* Host–symbiont specificity determined by microbe–microbe competition in an insect gut. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22673–22682 (2019).
372. Gil, R. *et al.* The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9388–9393 (2003).
373. Powell, J. E., Leonard, S. P., Kwong, W. K., Engel, P. & Moran, N. A. Genome-wide screen identifies host colonization determinants in a bacterial gut symbiont. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 13887–13892 (2016).
374. Hölldobler, B. & Wilson, E. O. *The ants.* (1990).
375. Lanan, M. C., Rodrigues, P. A. P., Agellon, A., Jansma, P. & Wheeler, D. E. A bacterial filter protects and structures the gut microbiome of an insect. *ISME J.* **10**, 1866–1876 (2016).
376. McCutcheon, J. P. & Moran, N. A. Functional convergence in reduced genomes of bacterial symbionts spanning 200 my of evolution. *Genome Biol. Evol.* **2**, 708–718 (2010).
377. López-Sánchez, M. J. *et al.* Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. *PLoS Genet.* **5**, 1–11 (2009).
378. Flynn, P. J., D’Amelio, C. L., Sanders, J. G., Russell, J. A. & Moreau, C. S. Localization of bacterial communities within gut compartments across *Cephalotes* turtle ants. *In prep.*
379. Charlop-Powers, Z. *et al.* Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. *Proc. Natl. Acad. Sci.* **113**, 14811–14816 (2016).

Titre de la thèse : Dynamiques des associations entre les fourmis amazoniennes et leurs communautés microbiennes

Résumé : Les associations entre hôte et communautés microbiennes sont nombreuses et dépendantes de plusieurs facteurs, tant sur le plan écologique qu'évolutif. Dans le cadre de cette thèse, nous avons étudié les facteurs structurant les associations entre fourmis amazoniennes et leurs communautés microbiennes. Les chapitres 1 et 2 décrivent respectivement la diversité des communautés bactériennes et eucaryotes associées à 38 espèces de fourmis amazoniennes. Ces différentes espèces possèdent diverses écologies et ont été collectées dans deux habitats distincts. Nous avons ainsi pu comparer les communautés microbiennes associées aux fourmis amazoniennes en fonction de différents paramètres de l'hôte, tels que son habitat, son mode de nidification, son régime alimentaire, sa taxonomie et son histoire évolutive. Pour cela nous avons procédé à une analyse métabarcoding en ciblant les gènes 16S rRNA et 18S rRNA respectivement dans les chapitres 1 et 2. Nos résultats révèlent que les communautés microbiennes associées aux fourmis amazoniennes sont structurées par différents facteurs tels que le régime alimentaire, le mode de nidification ainsi que la taxonomie cependant cette structuration ne semble pas impactée par l'habitat. Le chapitre 3 présente une analyse de la diversité des clusters de gènes de biosynthèse (CGB) des bactéries symbiotiques des fourmis du genre *Cephalotes*. Ces CGB permettent aux bactéries de produire des métabolites secondaires dont certains jouent un rôle dans la structuration et la maintenance des communautés bactériennes, qu'elles soient associées à un hôte ou non. Nos résultats révèlent que les communautés bactériennes dominantes de 17 espèces de fourmis *Cephalotes* possèdent de nombreux clusters de gènes de biosynthèse dont certains présente une haute similarité entre des CGB issu d'un même symbiote à travers différentes espèces hôtes de fourmis, et entre les CGB provenant de différents ordres bactériens au sein d'une même espèce hôte. Ces résultats préliminaires ouvrent la voie pour mieux comprendre le rôle fonctionnel de ces métabolites dans les interactions bactériennes et les interactions bactérie-hôte.

Mots clefs : insecte, hôte, bactérie, eucaryote, symbiote, métabarcoding, métagénomique, cluster de gène de biosynthèse, métabolite secondaire, qiime2.

Thesis title: Association dynamics between Amazonian ants and their microbial communities

Abstract: Host-associated microbial communities are numerous et depend on several ecological and evolutive factors. In this thesis, we studied the factors structuring the associations between Amazonian ants and their microbial communities. The chapters 1 and 2 describe respectively the diversity of bacterial and eukaryote communities associated with 38 Amazonian ant species. These different species have diverse ecologies and have been collected in two distinct habitats. As such, we were able to compare the microbial communities associated with Amazonian ants according to different host parameters, like its habitat, nesting mode, diet, taxonomy and evolutive history. To this aim, we did a metabarcoding analysis targeting the 16S rRNA gene and 18S rRNA gene in chapter 1 and 2 respectively. Our results reveal that microbial communities associated with Amazonian ants are structured by different factors, like host diet, nesting mode and taxonomy, however habitat does not seem to be a structuring factor. Chapter 3 present an analysis of the biosynthetic gene clusters (BGC) diversity of symbiotic bacteria from the ant genus *Cephalotes*. These BGC allow bacteria to produce secondary metabolites, some of which play a part in the structuration and maintenance of bacterial communities, in natura or associated with a host. Our results reveal that the main bacterial communities associated with 17 different species of *Cephalotes* ants possess numerous biosynthetic gene clusters, with some of them presenting high similarity between BGC from the same symbiont across different ant host species, and between BGC from the same bacterial order within the same host species. These preliminary results pave the way to better understand the functional role of these metabolites for bacteria-bacteria interactions and host-bacteria interactions.

Keywords: insect, host, bacteria, eukaryote, symbiont, metabarcoding, metagenomics, biosynthetic gene cluster, secondary metabolite, qiime2.