Dietary adaptations in the genomes and microbiomes of convergent myrmecophagous mammals

Sophie Teullet

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Génétique et Génomique

École doctorale GAIA

Unité de recherche UMR 5554 ISEM

Dietary adaptations in the genomes and microbiomes of convergent myrmecophagous mammals

Présentée par Sophie Teullet

Le 27 novembre 2023

Sous la direction de Frédéric Delsuc

Devant le jury composé de

Kevin KOHL, Professeur assistant, Université de Pittsburgh, Pittsburgh, USA	Rapporteur
Simonetta GRIBALDO, Professeure, Institut Pasteur, Paris, France	Rapportrice
Marie SEMON, Professeure, Ecole Normale Supérieure, Lyon, France	Présidente du jury
A MURAT EREN, Professeur, Université d'Oldenburg, Oldenburg, Germany	Examinateur
Mélanie DEBIAIS-THIBAUD, Maitresse de conférences, Université de Montpellier, Montpellier, France	Examinatrice
Frédéric DELSUC, Directeur de recherche, CNRS, Montpellier, France	Directeur de thèse

UNIVERSITÉ DE MONTPELLIER

“This is some gourmet anteater shit.”

Gut fiction, 2023

PREFACE AND ACKNOWLEDGEMENTS

It's not about the destination, it's about the journey

You are about to read the summary of three years of research on the adaptation to ant and termite consumption in mammals; a work I did during my PhD. But doing a PhD is much more than doing scientific research. That is why I wanted, in an informal way, to talk about what my journey as a PhD student inspired me, what made me grow personally, and not just scientifically, and most importantly about things I did beside my research work.

Doing a PhD is a unique experience, both professionally and personally, as you get the chance to work on your own project for three years. You can focus on this project, you gain autonomy, you meet people, you learn new methods, you exchange on many diverse and fascinating subjects with passionate people, you travel, you do various things from fieldwork to lab work and analyses… But, you also try things, it does not work, you try to understand why, you are frustrated, you try again, several times, it finally works and you are ecstatic, thrilled, you also stress for presentations before being relieved and celebrating: doing a PhD can be an emotional roller-coaster! It has its ups but also its downs… It is a weird period during which you are not a student anymore, you start working in a research lab but you are still learning. It is a transitory period during which you change a lot and that is not always easy.

From the beginning, we are all aware that doing a PhD will be challenging because doing research is not easy. You get stressed, frustrated, you work a lot, you are confronted to the requirements of the academic research system with all its flaws, your work gets criticized and you must detach yourself from these critics… and when you start your PhD you are still discovering all of that and learning everything so it can be hard. But, you will manage it, people will be there for you, people will help you, you will have a team working with you, you will be fine. Cases for which it can be hard to manage the difficulties of the PhD are when you add other layers of difficulties. These additional difficulties can be multiple, complex and, above all, unique to everyone. We all have to face personal problems that can impact our work. We have to deal with the disadvantages of the research system that can push us to produce more while cumulating post docs in order to try to have a chance to get a permanent position in the

end. We do not receive the same supervision and relationships between mentors and mentees can be complex, which can also impact our work as we are still learning how to do it. On that note, I want to mention the survey done by A. Murat Eren about mentorship in academia that I found truly enlightening (https://merenlab.org/2021/06/01/mentorship-survey/). This survey gave voice to Early Career Researchers (ECRs) to talk about their experiences as mentees, highlighting both good and bad experiences enabling both mentors and mentees to reflect on their own relationships and practices. To continue listing some of the difficulties than can add up while doing a PhD, let us not forget the covid situation and numerous lockdowns that impacted our research in various degrees (working from home, less interactions, harder to get help, field and lab work delayed or canceled, etc), but which overall impacted many PhD students worldwide. When you add up all these difficulties, you can find yourself in a very tough situation where it is hard to stay motivated and not discouraged. If you are not surrounded by people to help you go through these hard times, it can be very deleterious for yourself. It is now known, and more and more people are talking about it, that PhD students constitute a population more at risk of feeling lonely, developing addictions, having suicidal thoughts, burn outs, and depressions (*e.g.*, Levecque *et al*, 2017; Combes, 2022). But how can we change that?

During my PhD, I witnessed a lot of situations, in different labs, in France and abroad, where PhD students were not feeling well at all, either throughout their whole PhD or during some periods. It revolted me because in many cases solutions could be found to help these students or at least to make their PhD experience less tough. I am talking about PhD students but let us not forget about other non-permanent people (*e.g.*, post docs, technicians) who are often in similar situations. One easy solution is first to TALK. Talk about your problems, get stuffs off your chest, find people willing to listen to you, support you, because it will help you clear your mind, find solutions, and above all feel better and not lonely. It is also about asking more often your colleagues and friends how they really are feeling. Thankfully, in my lab and from other labs too, I always found there was a wonderful, caring, and supportive PhD student community. Having this community around you is important because being there for each other and supporting each other when you know what the other is going through is really helpful. But when facing certain problems, you might also need talk to the head of your lab, or someone in your lab who you trust, they should help you find solutions, take actions, and

guide you to other resources if needed; they have more power. Besides, at a larger scale, other organisms can help you, such as associations like "Femmes et Sciences" (www.femmesetsciences.fr) or the "Confédération des Jeunes Chercheurs" (cjc.jeunes-chercheurs.org). If you are struggling during your PhD, no matter the reason: do not stay alone.

As one of the two PhD student representatives of my lab, I wanted to alert people in our lab about how PhD students were truly feeling. Indeed, we noticed several of our colleagues were facing many problems, and were not well at all, feeling demotivated, sad, anxious, under pressure, exhausted, etc. With the support of our lab direction, we decided to conduct a survey among PhD students. We presented these results in front of the PhD students with whom we exchanged a lot trying to find short- and long-term solutions to improve their well-being at work. We also presented these results to supervisors to alert them about the situation of the PhD students at ISEM and hopefully raise awareness on the doctoral student situation in our lab. Students feeling bad had to deal with problems in their thesis environment, including dealing with their supervision. Besides, many students said the reality of research weights a lot on their motivation. There were problems easy to solve, for instance those about students needing more material (computers, desk chairs, etc). Other problems were not straightforward to solve because the research system require time to change. But other problems related to the thesis environment could be solved again by more communication: defining how many times you should meet with your supervisor(s) (the ideal frequency highlighted by this survey was once a week for example), how your meetings should be organized, how you should communicate when working remotely, how much autonomy you need, etc. And above all, ask "how are you doing? But really, how are you?" to prevent problems before getting overwhelmed. Among the solutions PhD students proposed, what came up most often was: to encourage more activities and discussions between students to prevent them from feeling lonely and increase social interactions, but also to encourage more supervisors to take management/communication trainings, to limit the number of students per supervisor (at the moment our doctoral school imposes a limit of three PhD students full time, or six in co-supervision per supervisor but Master students should be counted as well), to encourage co-supervisions (but not add too many supervisors).

Discussions generated by this survey were very enriching and inspiring, and I hope there will be a follow-up on this survey because, in most cases, solutions can be found to help students to truly enjoy their PhD experience when they are facing difficulties and not feeling well.

We should not just focus on the well-being of PhD students but also on the well-being of every person in our labs; no matter their gender, the language they speak, their disabilities, etc. Everyone should feel included at work. Thoughts and discussions about Diversity, Equity, and Inclusion (DEI) are more often taking place in labs, conferences, seminars. A working group was recently created at ISEM: the IDEAL group ("Inclusion, Diversité, Equité, Action et Lutte contre les discriminations"). I decided to be part of this group to learn more about these questions. We started discussions about the actions we wanted to see our group do, started writing a DEI statement for our lab, discussed about the language to use during seminars, but also the creation of a resting/breastfeeding room. I hope this group will grow in the future and more people will join it to help us improve the well-being of everyone at ISEM.

Many members of this group are, like me, also members of the "Femmes et Sciences" association; one source of additional inequity being the one between men and women. This French association aims at promoting research to girls, helping women in STEM, and promoting research done by women. The mentoring program for female PhD students and post docs is one of the association's famous actions (Batut *et al*, 2021). I had the chance to do it in my first year of PhD and it really helped me gain self-confidence and navigate the difficulties of starting a PhD during covid times. Members of "Femmes et Sciences" also gather, often during "Power Lunches" or "Power Apéros", to exchange on diverse topics related to difficulties women can encounter during their carrier. These discussions are always inspiring and give hope as you see all these women full of energy and desire to fight for change. Closely related to "Femmes et Sciences" I also want to mention the association "Mother in Science" (www.mothersinscience.com) which focuses on improving the conditions of mothers in academia. One of the main actions of "Femmes et Sciences" is to spread word on the inequities between men and women in research, to promote science to girls and the work of women scientists, often left behind. To do that, members of the association often give conferences or are invited by medias but mostly go to middle and high schools. I had the

chance to participate in these school interventions and it was an incredible experience which made me learn a lot and gave me faith in the new generation.

During my interventions in schools with "Femmes et Sciences" I also had the opportunity to present and popularize my research. I really enjoyed it. Our research work is not accessible to the general public but it is crucial that people have access to science; and if there are two things that show how important this is, it is the climate change crisis and covid crisis. Instead of scaring people we should teach them. Scientific popularization is important. I had the chance to do it during my PhD and I will try to continue, as best as I can in the future. I had the opportunity to share my research with diverse persons, from students to people in a bar, in different formats and it was always enriching. I went to middle and high schools with "Femmes et Sciences" and the "Déclics" initiative (www.cerclefser.org/fr/declics) during which researchers are going to high schools to present their work. I also went in a bar thanks to the "PhD pub" of Montpellier (www.phdpub.fr) which gives the opportunity to PhD students to present their projects in a bar, every month. Many other initiatives exist and I hope I will have the opportunity to participate in them. Here are some of them: Pint of Science (pintofscience.fr) also to present your research with a beer in your hand, the "Comptoir des sciences" (www.cerclefser.org/fr/comptoir-des-sciences) which are videoconferences between researchers and high school students, the "DECODER" journal (journal-decoder.fr) in which you can popularize one of your published article with middle and high school students playing the role of reviewers. The mailing list "La Scitoyenne" lists these actions and many more (www.cerclefser.org/fr/la-scitoyenne). There are also national events organized every year in many French cities, during which numerous activities are held to meet scientists, learn and experiment with your friends and family, for instance during the "Fête de la science" (www.fetedelascience.fr) and "Nuit des chercheurs" (nuitdeschercheurs-france.eu).

Being attentive to the well-being of students in research labs, fighting for more inclusion and less inequities in academia, promoting women in STEM, and popularizing scientific research to the general public are all causes that are close to my heart. I started being involved during my PhD, I still have a lot to learn but I am eager to do it. I hope to convince more people along the way, and to contribute at my small way to making academia a more welcoming, caring, and inclusive environment.

References

Batut, J., Kvaskoff, M., Morris, M.C., 2021. When mentoring matters: a French mentoring program for women in science. Nat Biotechnol 39, 776–779.

Combes, A.B., 2022. Comment l'université broie les jeunes chercheurs. Précarité, harcèlement, loi du silence., Autrement. ed. Paris.

Levecque, K., Anseel, F., De Beuckelaer, A., Van der Heyden, J., Gisle, L., 2017. Work organization and mental health problems in PhD students. Research Policy 46, 868–879.

"Good company in a journey makes the way seem shorter" (I. Walton)

Doing a PhD has its ups and downs and is not always easy. Having people surrounding you, helping you, supporting you, loving you, sometimes cooking for you, or buying you chocolate in these moments is precious. I want to thank the persons who accompanied me during these last three years (and even before) and who always supported me, encouraged me, cheered me up and just were there for me.

Scientific research is not something you can do alone, you always need people to discuss with you about new ideas, to help you conduct field and lab work, to debug far too complicated scripts, to share papers with you... research is a team work. I also want to thank the persons who helped me throughout these three years and helped me grow professionally.

First, I want to thank the members of the jury for accepting to review this thesis and to be present at the defense.

Merci aux membres de mon comité de thèse, Emilie, Clémentine, Carole, Olivier et Tom. Merci d'avoir accepté de faire partie de ce comité, vos précieux conseils m'ont beaucoup aidée.

I also want to thank all the people I had the opportunity to collaborate with, notably people that I had the chance to meet on the field and helped us during the two fieldtrips I did during my PhD (Benoit, Roxane, Edith at Pasteur; Wendy, Dylan, Nico in South Africa). Thanks a lot Chris for all the very interesting discussions we had, I hope we will meet in person someday! Merci Guillaume pour votre accueil à Pasteur avec Simonetta et ton aide précieuse tout au long de notre collaboration.

Merci à Sylvie, ma mentore à Femmes et Sciences. Merci pour vos conseils, votre bienveillance et votre écoute, vous m'avez aidée à grandir pendant ces trois années de thèse. Plus généralement, merci à Femmes et sciences (et dans F&S s'il y a bien une femme à remercier c'est May !), merci pour les power lunchs aux discussions très enrichissantes qui m'ont ouvert l'esprit sur plein de sujets, merci pour le mentorat, merci de m'avoir donné l'opportunité d'aller à la rencontre des jeunes dans les lycées et collèges pour partager mon expérience, quelque chose qui me tenait à cœur, que j'ai adoré et que j'espère pouvoir continuer à faire tout au long de ma carrière tant j'ai appris lors de ces rencontres.

Je souhaite remercier l'ISEM, qui nous offre un cadre exceptionnel pour mener à bien notre recherche. Je remercie tout particulièrement l'équipe de direction de l'ISEM, Nicolas, Carole, PierrO, Sébastien, Charlie notamment pour votre écoute et aide lorsqu'il a été question du bien être des doctorant.e.s au labo. Une attention spéciale pour PierrO : merci infiniment pour ton écoute, tes conseils et ton aide, ta dévotion envers les docs, c'est incroyable ! Tu as toujours les bons mots pour nous rebooster quand on est un peu down, toujours une petite grimace pour nous faire rire dans les couloirs, tu trouves toujours le temps pour parler de nos problèmes et trouver des solutions… MERCI pour tout ! Une pensée également pour les admins du labo qui nous sont d'une aide précieuse, qui n'ont pas toujours les tâches les plus sympathiques et sans qui nous aurions bien du mal à mener notre recherche. Merci Florence pour votre aide et efficacité notamment dans l'organisation de toutes mes missions de thèse.

Bien entendu je ne peux parler de l'ISEM sans évoquer l'équipe PEM ! Merci à toute l'équipe pour votre accueil depuis mon stage de master (pour moi la toulousaine fraichement arrivée à Montpellier !). Merci pour les discussions enrichissantes les midis au RA ou lors des

réunions d'équipe, pour vos conseils, la bonne humeur et les apéros ! Plus généralement merci à toutes les personnes du rez-de-chaussée du 22. Merci pour les petites discussions en salle café, les apéros encore, le fameux repas de noël, les moments de partage en mission en Guyane ou en Italie autour d'un ti punch ou d'un spritz ! Toutes ces petites choses du quotidien qui rendent le cadre de travail plus agréable !

Il y a des isémiens et isémiennes que je souhaite remercier plus particulièrement. Tout d'abord merci à toute la team ConvergeAnt ! Merci à toi Fred de m'avoir donné l'opportunité de travailler sur ce projet dès mon M2, merci pour ta confiance tout au long de ces trois années pas toujours faciles et marquées par les confinements, merci de m'avoir laissée la liberté de m'approprier ce vaste sujet dans les directions qui me passionnaient le plus. Et puis merci de m'avoir appris qu'il est important de toujours avoir du biltong sur soi et de prendre le temps, après une journée de terrain, de se poser autour d'un verre de ti punch en Guyane ou d'Amarula en Afrique du Sud ! Merci Rémi, pour ton aide tout au long de mon stage et de ma thèse, pour tes conseils et merci pour les moments partagés hors labo autour d'une bière à Zoobrew ou d'un week end des docs ! Merci Amandine, même si on ne s'est pas vu beaucoup à l'ISEM merci pour ton aide pour les manips. Pendant ma thèse j'ai également pu encadrer un étudiant, Victor, que je tiens également à remercier. Un grand merci à Marie-Ka… pour beaucoup de choses ! J'ai énormément appris en manipant à tes côtés et j'ai adoré nos conversations toujours enrichissantes. Mais plus que les aspects professionnels merci pour tes encouragements, ta bienveillance, tes conseils, ton écoute et ton soutien à des moments où j'en avais particulièrement besoin, et merci pour tous les échanges au coin de ton bureau, le midi devant le 22 ou autour d'une bière place des beaux arts, merci pour ces moments de partage à discuter voyages, vie, de manière presque philosophique parfois, à s'échanger des bonnes adresses à Montpellier, à me faire découvrir la cuisine indienne, bref merci Marie-Ka ! Merci Mathilde. Je me souviens du tout début de nos thèses, commencées le même jour, où tu m'envoyais un mail avec pour objet « choix du bureau ». Il s'en est passé des choses depuis ce mail et merci d'avoir toujours été là pour moi, merci pour ton écoute et ton soutien incroyable dans des moments pas toujours faciles, merci pour les têtes à têtes du midi et puis merci pour tous les moments partagés hors ISEM, les soirées jeux, les aprems à la rivière, les brocantes, les ventes de plantes et les soirées cuisine. J'espère qu'on continuera à partager autant de bons moments après nos thèses !

L'ISEM c'est aussi les docs et plus généralement des non permanents, docs, post docs, masters, de passage quelques mois ou présents depuis le début de ma thèse. Vous formez une team incroyable qui se serre les coudes et s'entraide et cette solidarité fait du bien. Alors merci à tou.te.s ! Merci pour les apéros, les week ends des docs, les JDD, les pauses café, les soirées Charlie's/Barberousse, la bonne humeur, l'entraide ! Merci Eliette et Mathilde les voisines de bureau, merci pour les biscuits et cookies (on a toujours besoin de sucre !), Nathan pour la découverte de jeux et le thé glacé plus que nécessaire l'été, Nico pour m'avoir coachée au billard (et merci Kim, merci à vous deux les copains), Amira, Noémie, Alexis, Bérénice, Alba pour ton partage de la culture italienne, Céline, Narla pour ce magnifique maquillage pailleté à la fête des fanfares, Rémi, Quentin, Maxime, Gwen et Marie pour les bons restos à Ferrara, Louise parce qu'on formait un sacré duo d'organisation d'évènements isémiens, Alice pour cette magnifique session photos à la plage, Léa toujours partante pour rigoler (tous les quatre avec Arthur et Alex dans votre bureau on n'était pas toujours très productifs mais on rigolait bien), Manue, Elodie, Arthur, Laura, Marie pour les danses endiablées au Barberousse et les sorties ciné où le film n'était pas toujours choisi pour le scénario et parfois pour le casting, Lila thank you for your good mood, I did not think I would met someone from Kingston here!, Jean-Loup, Heitor and Marcos thank you for all the fun in Montpellier (and Sète discovering the famous mussels/fries), Félix, Killian, Adrien, Romain merci pour les boites de sardines, Lucas, Iris thanks for your kindness and this great moment in Sète, Zach et Matthieu pour avoir été des super co-bureaux, Yohan et Marjo pour les apéros docs et les spritzs à Ferrara, Angèle pour cette fameuse soirée quizz au Tarbrew et parce qu'on a formé une sacrée team au CU… et si j'en ai oublié.e.s je m'excuse ! MERCI ! THANK YOU!

Parmi ces non permanents il y a une petite team que je tiens plus particulièrement à remercier : la team cancoillotte !! Merci Manue, Elodie, Tutur et Alex, merci pour tous ces fous rires, pour les soirées jeux, pour les bières à Zoobrew (special thanks to Oko), pour les soirées Barberousse où on finissait à tour de rôle dans des états plus ou moins stables, pour la meilleure cancoillotte que j'ai jamais mangée (bon le Cantal reste au-dessus), pour les soirées iroquoises, pour le pédalo sur la Vltava (et le Hugo in love), pour les pauses cafés et midis au labo racontage de vie et vidage de sac, merci pour la bonne humeur, pour votre soutien énorme… merci les copains !

Finir sa thèse ne marque pas que la fin de trois années mais la fin de huit années d'études sup et je n'oublie pas les personnes rencontrées bien avant ma thèse, qui m'ont accompagnée tout au long de mes études, encouragée et soutenue dans les moments difficiles et c'est aussi l'occasion de les remercier ici.

Merci aux ami.e.s rencontré.e.s pendant mes études à Toulouse et Montréal, que d'aventures vécues dans ces deux villes !

Merci au master EE ! Merci Clara, Laurine, Julie, Paul, Laurie, Océane, Tristan, David, Samantha, merci d'avoir formé cette team incroyable à Toulouse (et bien après aussi, malgré la distance, malgré deux confinements), merci pour la solidarité, la fameuse soirée fonclette, les visios parfois transformées en séance psy pour vider son sac, merci pour tout ! Et merci pour votre soutien à des moments où j'en avais bien besoin. Merci d'avoir été là.

Et rencontré.e.s avant le master il y a bien sûr Potos Fac ! Comment ne pas vous remercier ? Et comment vous remercier tellement vous m'avez tant apporté pendant 5 ans à Toulouse ! On a vécu tellement de beaux moments, tellement partagé, tellement ri, on a voyagé aussi, on s'est soutenu dans les moments difficiles, on a fêté des anniversaires, des départs au Canada, des fins d'exams, on a travaillé aussi un peu (quand même !), merci pour tout et merci après toutes ces années d'être encore là, malgré les emplois du temps chargés, la distance et parfois le décalage horaire, MERCI. Merci Louise, Océane, Hugo, Mylène, Baptiste, Cyrine, Camille, Fanny, Tristan, Julie, Juliette, Marie ! Merci pour les soirées au Snaper Rock, pour les vacances en Normandie, pour les fous rires en cours, pour les soirées dans des apparts beaucoup trop petits, Mylène et Louise cette soirée au QG à Toulon, Cyrine pour avoir supporté la chaleur de Montpellier, Baptiste pour les bonnes adresses restos, Julie pour les découvertes incroyables à la Réunion, et j'en oublie plein ! Mais juste MERCI !

Ma petite Marie, tu sais déjà tout, tu sais déjà que je ne pourrai jamais assez te remercier pour tout ce que tu as fait pour moi, pour ton soutien inconditionnel depuis huit ans maintenant, dans les bons comme dans les mauvais moments, merci pour tout. Je pourrai citer 1 000 souvenirs que j'ai avec toi tant on a vécu de choses toutes les deux depuis ce premier jour de L1 où, assise au premier rang d'un amphi du U2, tu te retournais et on commençait à papoter. Je citerai simplement quelques bons souvenirs de ces trois dernières années : merci de m'avoir fait découvrir ta petite vie en Roumanie, merci pour ces vacances

des princesses d'Hendaye incroyables, merci pour les surprises à Montpellier, merci pour les visios qui faisaient un bien fou, merci pour les mots réconfortants, MERCI !

Merci aux amies rencontrées lors de mon échange à Montréal, vous avez marqué mes années d'études et je suis très reconnaissante, que malgré la distance, nous arrivions à partager encore autant de beaux moments, merci Léa pour les bonnes adresses à Prague, pour le voyage organisé en last minute à Malte, pour cette glace que nous étions obligées de manger à la Valette, pour les danses folles au Wiston Churchill à Montréal avec Célia, pour les fous rires... Merci Alice pour les voyages, pour m'avoir fait découvrir la culture néerlandaise, pour les visios vidage de sac pendant nos thèse, pour ta bonne humeur… et puis merci Sophie, Mariya, Célia, Julie, Timea, Estelle, merci pour tous les bons moments à Montréal et ailleurs dans le monde !

Mélanie, on se suit depuis bien longtemps maintenant, la sixième si je ne me trompe pas ! Merci pour tout, pour les fous rires en cours de philo, pour les cafés au Milk, pour les heures au téléphone à déchiffrer les DMs de maths, pour les aprems à la piscine… merci pour tout ! Si on imaginait quand on était encore à la Ponétie, en train de râler sur notre emploi du temps et nos profs, qu'on en serait là aujourd'hui !

Merci infiniment Carmen pour les illustrations juste magnifiques de cette thèse. Merci d'avoir accepté de participer à ce projet un peu particulier. Et des souvenirs me reviennent, avec Mélanie, en voyage scolaire en Espagne, merci aussi pour ces moments-là.

Si les dessins de cette thèse vous plaisent, allez admirer le travail incroyable de Carmen, aka R. Oncedor (instagram : @roncedor @mistybulle).

Je souhaite aussi remercier le rayon chocolat de carrouf, les chocolatiers et toutes les personnes qui m'ont acheté du chocolat !

Et puis il y a ceux et celles qui me suivent depuis encore bien plus longtemps que le collège, la famille. Merci à toute ma famille pour son soutien infaillible et ses encouragements. Merci papa et maman. On n'a pas toujours traversé des moments faciles mais on les a traversés ensemble. Merci pour votre soutien, votre aide, vos encouragements, vos conseils,

pour m'avoir accompagnée dans chacune des étapes de la vie qui m'ont amené ici, merci de m'avoir rassurée parfois et poussée dans d'autres moments, merci d'avoir toujours cru en moi. Merci d'être les parents incroyables que vous êtes. Merci d'être toujours là pour moi. Je vous aime. Et promis c'est fini les études maintenant ! Merci Delphine, pour ton sourire, ta bonne humeur, ton rire, mais aussi pour m'avoir ouvert les yeux, m'avoir fait grandir et mûrir, je suis extrêmement fière d'être ta grande sœur. Je t'aime. Et merci mémé, mamie, tonton, tata, Elise, Adeline, Lucie et toute la famille. Une pensée émue va à mes deux grands-pères, j'aurais tant aimé partager cette fin de thèse avec vous. Bien sûr j'ai une pensée toute particulière pour toi Bernadette, la meilleure des belles mamans. De la petite Sophie qui venait te voir le samedi matin à la clinique vétérinaire, rêvant encore à l'époque de le devenir, à la grande Sophie qui décroche une thèse en biologie évolutive, tu as toujours été là pour moi, merci. Tu me manques.

Et le meilleur pour la fin : Alex ! Comment te remercier ? Je pourrai commencer par le commencement en fait, dire que notre histoire à débuter par une affaire de post it couplée à une bonne dynamique après une explication du tarot douteuse et aujourd'hui cette histoire s'est remplie d'aventures, de hauts, de bas, de bacalhau, de rencontres, de glaces à la banquise, de souvenirs, de montagne, de voyages, de boosters pokémon, de sorties photos, de joies, de copains, de restos, de concerts, de thèses, de cinés, de soutien, de patates, de fous rires, de monstera, de craquages et surtout d'amour. Je ne serai pas en train d'écrire cette thèse si tu n'avais pas été là pour moi, alors merci. Quelle année 2023, finir nos deux thèses à quelques mois d'écart, une aventure aussi ! Une aventure que nous avons vécue main dans la main. Merci pour tout. Je t'aime.

A tou.te.s, un grand MERCI ! THANK YOU!

TABLE OF CONTENTS

Table of contents

GENERAL INTRODUCTION

1. Convergent evolution and its implications in evolutionary biology

1.1. Assessing evolution's repeatability and predictability

Evolutionary biology aims at understanding the origin, maintenance, and evolution of living organisms' diversity. One fascinating evolutionary process is the repeated independent evolution of similar phenotypes in different lineages. Explaining such a phenomenon has become a major research topic because of its importance in improving our knowledge of evolution repeatability and predictability, raising questions about the evolutionary forces involved. In its thought experiment of replaying the tape of life, Gould (1989) argued that it would be impossible that evolution leads to the same outcome because of historical contingency, the fact that random events like mutations or genetic drift (*i.e*., random sorting of alleles) influence the different paths evolution can take, making it unpredictable. This hypothesis is often referred to as the Radical Contingency Thesis (RCT; Powell, 2012). The role of contingency in explaining biodiversity is further supported by the fact that evolution acts on available material, which can differ between species subject to similar selective pressures, a concept called evolutionary tinkering as first defined by Jacob (1977). Contrary to Gould's contingency theory, Conway-Morris (1998) argued that evolution could be, up to a certain degree, predictable because of natural selection which consists of the differential survival and reproduction of individuals in a population, given a time and a place. This opposed hypothesis is often referred to as the Robust Repeatability Thesis (RRT; Powell, 2012). To understand evolution predictability and repeatability, major questions therefore reside in deciphering the relative contribution of constraints (*i.e*., historical contingency) and determinism (*i.e*., natural selection) in shaping phenotypes evolution with the later making it more predictable (Blount *et al*, 2018).

To answer such questions, many laboratory experiments have been conducted to follow populations evolving under similar environmental conditions (Orgogozo, 2015; Losos, 2017; Blount *et al*, 2018). One of these famous experiments is the *Escherichia coli* long-term evolution experiment (LTEE) of Richard Lenski and colleagues, which started in 1988 and is still ongoing. In this experiment, genetic changes of 12 initial identical populations are monitored

throughout their evolution under similar culture conditions. This experiment notably revealed genetic changes occurring in all or some populations, including apparitions of novel traits, and that similar changes could evolve following different evolutionary paths (*e.g*., Lenski *et al*, 1991; Blount *et al*, 2008, 2012; Lenski, 2017). These results suggest that convergence and contingency both play a role in shaping the evolution of these bacterial populations. Laboratory experiments can therefore tell us about the initial conditions (*e.g*., genetic, environmental) leading to repeated evolution, and knowledge of these can help us predict evolution (Orgogozo, 2015). Yet, these experiments are often conducted on specific model microorganisms, can be limited in time, and do not reflect natural conditions. Natural experiments constitute an alternative but they are often long-term experiments that are not always easily done because the initial conditions are harder to control and ancestral populations are more heterogeneous (Blount *et al*, 2018). Thus, comparative studies on populations or species that have evolved independently toward similar phenotypes in response to similar selective pressures are key to decipher the mechanisms influencing evolution repeatability (Powell and Mariscal, 2015; Losos, 2017; Blount *et al*, 2018). Among such studies, one can cite the evolutionary radiation of *Anolis* lizards on the Greater Antilles where similar ecomorphs evolved independently on each island despite historical contingency (Losos *et al*, 1998). Other famous comparative studies are the long-term studies on stickleback populations of lakes and streams that have revealed morphological and genetic similarities evolving independently in these populations in response to the different selective pressures imposed by the two habitats (Kaeuffer *et al*, 2012).

Overall, these experimental and comparative studies have shown that evolution can be, to some degree, repeatable despite the influence of stochasticity and that different paths can lead to similar phenotypes (Orgogozo, 2015; Powell and Mariscal, 2015; Losos, 2017; Blount *et al*, 2018). These studies therefore highlight the influence of natural selection acting in response to similar selective pressures on the available material (*i.e*., genotypic diversity) which is shaped by the evolutionary history of the evolving species (*i.e*., historical contingency). Finally, such analyses emphasize the importance of studying cases of repeated and independent evolution of similar phenotypes (*i.e*., convergent or parallel evolution) to answer questions and test hypotheses about the repeatability of evolution. In the following part of this introduction a brief definition of convergence and parallel evolution is given (see

part 1.2), as well as an overview of some of the methods used to study convergence at the molecular level, as it has been the main focus of this PhD project (see part 1.3).

1.2. Convergent and parallel evolution

Independent repeated evolution of the same trait in different lineages could be the result of convergence (also called convergent evolution) or parallelism (also called parallel evolution). There have been several debates regarding the definition of these two phenomena. Parallelism often designates the independent evolution of similar phenotypes in different lineages sharing a common ancestry and being closely related (*e.g.*, between populations) whereas convergence refers to the independent evolution of similar phenotypes in distantly related lineages that do not share a direct common ancestor, for instance between distantly related species (Arendt and Reznick, 2008; Elmer and Meyer, 2011; Waters and McCulloch, 2021). Some definitions are also based on the developmental and molecular mechanisms underlying the repeated evolution of similar phenotypes. When changes occur in the same genetic or developmental pathways it often refers to parallelism by opposition to convergence where similar changes are not needed (Arendt and Reznick, 2008; Christin *et al*, 2010; Manceau *et al*, 2010; Elmer and Meyer, 2011; Losos, 2011; Stayton, 2015a, b). Several factors can influence the distribution of mutations and explain why similar or different genetic changes might occur. Among them are the pleiotropic (*i.e.*, influence of one genetic locus on several phenotypes) and epistatic (*i.e.*, interactions between genetic loci) effects of certain mutations, the number of genetic or developmental pathways involved in a phenotype, the size of the mutation target or the position of the gene in regulatory networks but also the effect of phenotypic plasticity or the strength of selection (Gompel and Prud'Homme, 2009; Stern and Orgogozo, 2009; Losos, 2011; Storz, 2016). These complex and multiple factors reduce the probability of similar molecular changes occurring in multiple lineages, making difficult the clear distinction between cases of parallelism or convergence. Besides, numerous cases where convergent phenotypes of closely related taxa have evolved through different mechanisms and in distantly related taxa through the same have been reported (Arendt and Reznick, 2008; Waters and McCulloch, 2021).

Several authors therefore argued that no clear distinction should be made between convergence and parallelism (Arendt and Reznick, 2008; Pearce, 2012; Powell, 2012; Waters and McCulloch, 2021). Arendt and Reznick (2008) proposed that only the term convergence should be used to designate any case of independent evolution of similar phenotypes. Recently, focusing on genomic evidence for repeated evolution, some authors have even proposed the term "repeated sorting" instead of parallelism to designate the evolution of repeated traits from shared standing genetic variation and to avoid confusion with the parallelism/convergence dichotomy often referring to the degree of relatedness between the compared taxa (Waters and McCulloch, 2021). Besides, convergence and parallelism are not mutually exclusive and can co-occur, as shown in several study systems, depending notably on the phylogenetic relationships between taxa and the degree of genetic exchanges (*e.g.*, introgression, ancient polymorphism) between them (Waters and Mc Culloch, 2021). Therefore, several authors now consider that parallelism and convergence should be seen as a continuum (Arendt and Reznick, 2008; Pearce, 2012; Waters and McCulloch, 2021). This continuum can be defined based on the level at which changes leading to similar phenotypes occur (Fig 1) and with parallelism being highly constrained as opposed to convergence. This continuum can also be viewed as a relationship with the time since shared ancestry (*i.e.*, the proportion of parallelism decreasing with the time since shared ancestry) (Fig 1).

Studying the underlying mechanisms involved in the evolution of convergent myrmecophagous phenotypes has been the focus of my PhD. Throughout this thesis, convergence will be defined as the repeated independent evolution of similar phenotypes in multiple lineages in response to similar selective pressures and therefore not inherited from a common ancestor (Arendt and Reznick, 2008; Stayton, 2015a, b). Convergence can be the result of adaptation under similar selective pressures (*i.e.*, natural selection) but also of constraints (*i.e.*, historical contingency) (Losos, 2011). The relative contribution of these two processes is discussed throughout the thesis. Numerous examples of convergent and parallel evolution have been reported in various organisms (*e.g.*, McGhee, 2011; Losos, 2017; Blount *et al*, 2018). My PhD project, more specifically, aims at studying the role of genomic and metagenomic adaptations in convergent evolution through the example of the convergent adaptation to myrmecophagy in mammals. This system model is presented in part 2 before detailing the specific goals of the thesis (see part 3).

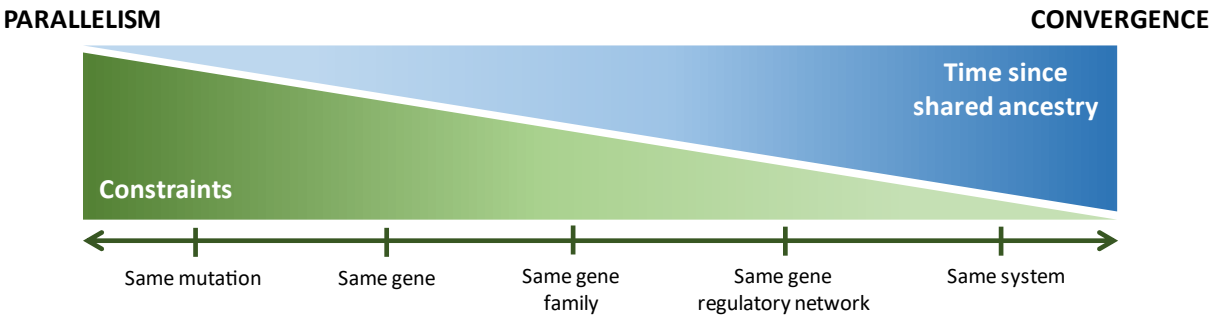


Figure 1. The parallelism-convergence continuum. Changes can occur at different levels and they can be placed along this continuum, from changes corresponding to the same mutation to changes occurring in the same system. These changes vary in the degree of imposed constraints. This continuum can also be seen as a result of the time since shared ancestry. Light colors indicate few constraints (green) or time since shared ancestry (blue) whereas dark colors indicate higher constraints or divergence. This figure is inspired from David Baum's talk at the SMBE 2023 meeting: "Parallelism-convergence continuum and its implication for homology assessment".

1.3. Methods to study convergent evolution

Studying convergence relies on comparisons between taxa. Cases where we can compare ancestral (thanks to fossils or laboratory experiments) and actual states are ideal but rare (Mahler *et al*, 2017). Besides, convergence is not always the result of natural selection and adaptation but can be caused by constraints, or even occur randomly (Losos, 2011; Stayton, 2015b). To study convergence, quantify it, and disentangle the numerous factors influencing it, statistical methods have been used (Stayton, 2015a, b; Mahler *et al*, 2017). With progress in phylogenetic reconstructions, phylogenetic comparative methods (*i.e.*, statistical comparative methods taking into account phylogenetic relationships) have been developed and improved our understanding of convergence (Garland *et al*, 2005; Mahler and Ingram, 2014; Mahler *et al*, 2017). Yet, those comparative methods present some limits as they depend on the evolutionary processes assumed in the underlying model (Mahler *et al*, 2017). Today, convergence can be studied at different levels (*e.g.*, phenotypes, genotypes, microbiota) using diverse types of data (*e.g.*, morphometrics, genomics, transcriptomics, metagenomics). Combining several methods allows us to adopt integrative approaches to fully understand the

multiple and complex causes of evolutionary convergences. Understanding the adaptive genomic mechanisms involved in the evolution of convergent phenotypes notably reside in deciphering whether the same genetic and/or developmental pathways are involved and whether the observed changes are adaptive. As it has been the subject of my PhD project, I will focus here on how molecular data can be used, by giving examples of some approaches, to learn more about the mechanisms involved in convergent evolution. Methods used during this PhD project will be further developed in the following chapters of this thesis manuscript.

- **Assessing the strength and direction of selection**

Assessing whether molecular changes are adaptive is possible by determining the strength and direction of selection using DNA sequences (Yang and Bielwaski, 2000; Fay and Wu, 2001; Ellegren, 2008; Jeffares *et al*, 2015). A measure widely used to assess whether protein-coding genes are under selection is the ratio ω of the non-synonymous substitutions rate dN (*i.e*., nucleotide changes that modify the protein sequence) over the synonymous substitutions rate dS (*i.e*., nucleotide changes that do not alter the protein sequence) (Yang and Bielwaski, 2000; Ellegren, 2008; Jeffares *et al*, 2015). A ratio lower than 1 reflects negative (or purifying) selection, meaning there is selection against non-synonymous substitutions to conserve the protein sequence. A ratio greater than 1 indicates positive (or adaptive or diversifying) selection, and a ratio equal to 1 suggests neutral evolution (Yang and Bielwaski, 2000; Ellegren 2008; Jeffares *et al*, 2015). Therefore, it is possible to assess whether similar phenotypes in different taxa are subject to similar selective pressures and hence might represent cases of adaptive convergence. For example, this measure has been used to study the repeated losses of enamel in mammals and highlighted relaxed selection acting on the associated genes in enamelless and toothless lineages (Meredith *et al*, 2009; Springer *et al*, 2019).

- **Studying convergence at the sequence level**

To identify genes involved in convergent phenotypes, convergent shifts in substitutions or gene evolutionary rates can be studied at the sequence level. Adaptive convergent changes in specific amino acids can occur in different taxa living under the same conditions and combined

with selection analyses (see above) could give insights into the adaptive significance of these changes. For example, convergent amino acid substitutions have been identified in numerous genes in several independent lineages of electric fishes (Wang and Yang, 2021). Several of these genes are under positive selection suggesting adaptive molecular convergences (Wang and Yang, 2021). In echolocating mammals (*i.e*., bats, toothed whales), convergent amino acids substitutions have been identified in hearing genes showing signal of positive selection (Shen *et al*, 2012), and genes coding for proteins involved in fast-twitch muscle contraction (Lee *et al*, 2018). Foote *et al* (2015) also identified convergent changes in genes under positive selection in several aquatic mammal species, potentially involved in marine adaptation. However, they also identified such changes in sister taxa suggesting that those shifts might not be due to marine adaptation and that adaptive convergent amino acid changes can be rare (Foote *et al*, 2015). Indeed, as the number of different possible substitutions is constrained by the number of different nucleotides (n = 4) and amino acids (n = 20) found at a specific site, convergent amino acid substitutions might occur by chance and this is likely the cause of many of the observed convergences detected in genome-wide studies (Rey *et al*, 2019; Zhou and Zhang, 2020). Additionally, other processes such as introgression, mutation biases, or incomplete lineage sorting complicate the identification of true convergences (Rey *et al*, 2019; Zhou and Zhang, 2020). The main challenge when studying convergent substitutions thus resides in differentiating true adaptive convergent substitutions (*i.e*., foreground substitutions) from those occurring because of other factors (*i.e*., background substitutions) and several methods have been developed in this purpose (Rey *et al*, 2019; Duchemin *et al*, 2023; Fukushima and Pollock, 2023). Another approach consists in studying convergent shifts in genes evolutionary rates along branches of a phylogeny (Chikina *et al*, 2016; Kowalczyk *et al*, 2019; Partha *et al*, 2019). These rates can be affected by environmental changes in two ways: (i) an acceleration due to adaptive changes caused by the same selective pressures or a relaxation of the constraints on specific functions that are not useful anymore, or (ii) a decrease caused by an increase of the constraints on a specific gene whose function become more important (Chikina *et al*, 2016). For example, several genes whose evolutionary rates have been increased or decreased, have been identified and linked to marine (Chikina *et al*, 2016) and subterranean (Partha *et al*, 2017) life adaptations in mammals.

- **Reconstructing gene family evolution**

Losses or gains of functions can be identified and the evolution of candidate genes families studied by identifying gene duplications and losses events associated with environmental or lifestyle changes. For instance, gene losses associated with taste, light perception, and claw keratin have also been identified in snakes and could be linked with the regression of the associated phenotypic traits (Emerling *et al*, 2017). Gene losses have been identified among several species of frugivorous (Wang *et al*, 2020) as well as vampire (Blumer *et al*, 2022) bats, and potentially linked to their specific diet. Also related with dietary specialization, the evolution of chitinases is marked by gene losses in bird (Chen and Zhao, 2019) and mammalian (Emerling *et al*, 2018) species not having chitin-rich diets (*e.g*., herbivores, carnivores). Additionally, in mammals, the evolution of olfactory and taste receptor gene families is characterized by duplications and losses specific to certain clades and seems to be linked with ecological characteristics (*e.g*., lifestyle, diet, vision) leading to a huge variation in the size of these gene repertoires between species (*e.g*., Jiang *et al*, 2012; Feng *et al*, 2014; Hughes *et al*, 2018; Shan *et al*, 2018). Studying the evolution of gene families to understand convergent adaptations is further detailed in Chapter I of this thesis through the study of two gene families (chitinase and taste receptor genes) in myrmecophagous mammals.

- **Comparing gene expression**

Phenotypes evolve as a result of the expression of genotypes. Therefore, to improve our understanding of the roles of candidate genes in the evolution of convergent phenotypes, gene expression profiles can be compared between taxa thanks to transcriptomic data (*i.e*., the set of all RNAs expressed in a sample). Tissular location and levels of global gene expression, as well as candidate gene expression, can be compared and help explain phenotypic differences exposed to different selective pressures. For instance, such analyses have revealed convergent patterns in gene expression profiles of venom gland transcriptomes among 20 venomous snake species (Zancolli *et al*, 2022), or shared gene expression patterns involved in caste determination in eusocial insects (Berens *et al*, 2015). In addition to gene expression, studying proteomes or metabolomes (*i.e*., respectively the set of all proteins or metabolites found in a sample) can confirm patterns observed at the gene level and improve our knowledge of the

molecular bases of convergence. Genomic and proteomic comparative analyses have, for example, enabled the identification of molecular pathways involved in the adaptation to high altitudes in several human populations (Sharma *et al*, 2022). The use of transcriptomic data to understand the role of candidate genes in the evolution of convergent phenotypes will be further developed in Chapter I of this thesis.

- **Considering the role of symbiotic microorganisms**

Plants and animals live in close association with millions of microorganisms (*i.e*., regrouped under the term microbiota) which impact their health and more generally fitness. These microorganisms can have crucial roles in adapting to new environmental conditions. For instance, diversification of dietary habits in mammals is likely to have been influenced through changes in their gut microbiota leading to convergences in composition between lineages sharing similar diets (*e.g*., Ley *et al*, 2008; Muegge *et al*, 2011). Considering the role of symbiotic microbial communities can be complementary to genomic approaches described above. Together these approaches can help understand the role of the holobiont (*i.e*., the host and its associated microbes) in adapting under specific selective pressures. For instance, such approaches have helped to decipher the mechanisms underlying the adaptation toward highly specialized diets such as sanguivory in vampire bats (Mendoza *et al*, 2018) and myrmecophagy in the short-beaked echidna, giant anteater, and Malayan pangolin (Cheng *et al*, 2023). Comparisons of gut microbiota taxonomic and functional compositions to decipher its role in convergence are discussed in Chapter II of this thesis and the role of the holobiont in adapting to myrmecophagy is developed in the general discussion.

Overall, studying molecular adaptations underlying convergent phenotypes has improved our knowledge of this widespread phenomenon and of the multiple mechanisms involved. To fully understand evolutionary convergence, it is crucial to combine multiple approaches and study it at different levels. As it will be discussed in the last part of this introduction, during my PhD I used different methods and data to understand convergent evolution in myrmecophagous mammals. How combining these methods can help us to better understand

convergent adaptations is discussed throughout this thesis and especially in the general discussion part. For example, associating comparative genomics and transcriptomics can give insights into the functions of specific genes and thus reveals their putative role in adaptations (see Chapter I). Studying both genomic adaptations in the host and its associated microbial communities can shed light on the relative contributions of the two symbiotic partners in adapting to specific conditions (see Chapter II and the general discussion). Finally, associating molecular changes with phenotypic traits by correlating genomic and morphological data can help associate phenotypic and genotypic changes resulting from adaptations under similar selective pressures (see Chapter I and the general discussion).

2. Case study: myrmecophagous mammals, a striking example of convergent evolution

Several examples of morphological convergence have been described, notably as revealed by molecular studies, in placental mammals (Springer *et al*, 2004). Indeed, their rapid radiation around the Cretaceous-Tertiary boundary about 65 million years ago (Mya) (Springer *et al*, 2003; Álvarez-Carretero *et al*, 2022; Carlisle *et al*, 2023; Foley *et al*, 2023) was accompanied by an important ecological diversification leading the major super-orders to evolve convergently on three different continents: Afrotheria in Africa, Xenarthra in South America, and Boreoeutheria comprising Laurasiatheria and Euarchontoglires in the Northern hemisphere (Archibald and Deutschman, 2001; Springer *et al*, 2003; Meredith *et al*, 2011; Springer *et al*, 2017). Species from these different clades adapted to similar ecological niches left vacant by the disappearance of non-avian dinosaurs during the crisis, which led to numerous cases of morphological convergences (Delsuc *et al*, 2003; Springer *et al*, 2004; Luo, 2007; Wildman *et al*, 2007; Meredith *et al*, 2011). For instance, mole-like or hedgehog-like species have evolved in parallel in Laurasiatheria and Afrotheria (Springer *et al*, 2004). One striking example of convergence is the dietary specialization of consuming exclusively ants and termites (*i.e*., myrmecophagy) in species belonging to Afrotheria, Xenarthra, and Laurasiatheria (Delsuc *et al*, 2003; Springer *et al*, 2004).

Many mammalian species consume ants or termites but a species is considered strictly myrmecophagous when its diet is composed of more than 90% of ants and/or termites (Redford, 1987). Myrmecophagous placentals include species consuming a broad range of termite and/or ant species while others specialized on specific social insect species (Redford, 1986; Swart *et al*, 1999; Miranda *et al*, 2009; Pietersen *et al*, 2016; Allio, 2021; Sun *et al*, 2022). Myrmecophagous species can prey on ants and/or termites opportunistically and their diet can vary with prey availability and abundances as a result of seasonal variations of the habitat (Redford, 1986, 1987; Swart *et al*, 1999; Gallo *et al*, 2017; Weyer, 2018; Panaino *et al*, 2022). It is important to note that the degree of specialization toward ant and termite consumption varies between species. Pangolins and anteaters are highly specialized myrmecophagous species preying almost exclusively on ants and termites and present the most extreme phenotypes (*e.g*., Redford, 1987; Reiss, 2001). Yet, myrmecophagous species diet can include other insects or even plants, fruits, and seeds, occasionally for anteaters or more frequently for armadillos which present less pronounced morphological adaptations and are opportunistic myrmecophagous species (Redford, 1986; Brown, 2011; Vaz *et al*, 2012; personal observations during dissections of armadillo stomach contents). Additionally, several Coleoptera larvae were found while inspecting anteater stomach contents from French Guiana (personal observations). In four individuals of *Tamandua tetradactyla* (southern tamandua) between one and 12 larvae were found and ten were retrieved from one individual of *Myrmecophaga tridactyla* (giant anteater) (personal observations; Fig 2). Termitophilous beetles have been reported in fecal samples of several individuals of *T. tetradactyla* (Vaz *et al*, 2012; Sun *et al*, 2022). Analysis of the *COX1* gene of one larva retrieved in *M. tridactyla* revealed it belongs to the Dynastidae family and its sequence has the highest similarity with the one of *Heterogomphus telamon*. This beetle species is present in French Guiana and could be termitophilous and/or myrmecophilous. This result, which should be confirmed by analyzing mitochondrial DNA of other larvae, suggests these larvae are ingested opportunistically while foraging by the giant anteater.

Figure 2. Photo of the ten Coleoptera larvae identified in the stomach content of an adult *M. tridactyla*. One larva is shown on the right as an example.

Twenty-two placental species are considered myrmecophagous and are found in five distinct placental orders: Tubulidentata (aardvark), Pilosa (anteaters), Cingulata (armadillos), Pholidota (pangolins), and Carnivora (aardwolves) (Fig 3). Although they were grouped together in the Edentata clade based on their morphological resemblance, molecular data revealed that myrmecophagous species do not share a direct common ancestor (Delsuc *et al*, 2002; Meredith *et al*, 2011; Álvarez-Carretero *et al*, 2022; Fig 3). Myrmecophagy thus evolved independently at least five times in placentals. Additionally, the bat-eared fox (*Otocyon megalotis*; Carnivora, Canidae), an insectivorous species, is often considered a partially myrmecophagous placental. In mammals, myrmecophagy also evolves in monotremes and marsupials with respectively the short-beaked echidna (*Tachyglossus aculeatus*), which consumes invertebrates and includes large quantities of ants and termites in its diet (Rismiller and Grutzner, 2019), and the numbat (*Myrmecobius fasciatus*) which prey on termites (Cooper, 2011). The main focus of my PhD was to study the convergent adaptation to myrmecophagy in the five aforementioned placental orders. The numbat and echidna, as well as the bat-eared fox, are briefly mentioned in the following chapters of this thesis and included in some datasets for comparative purposes.

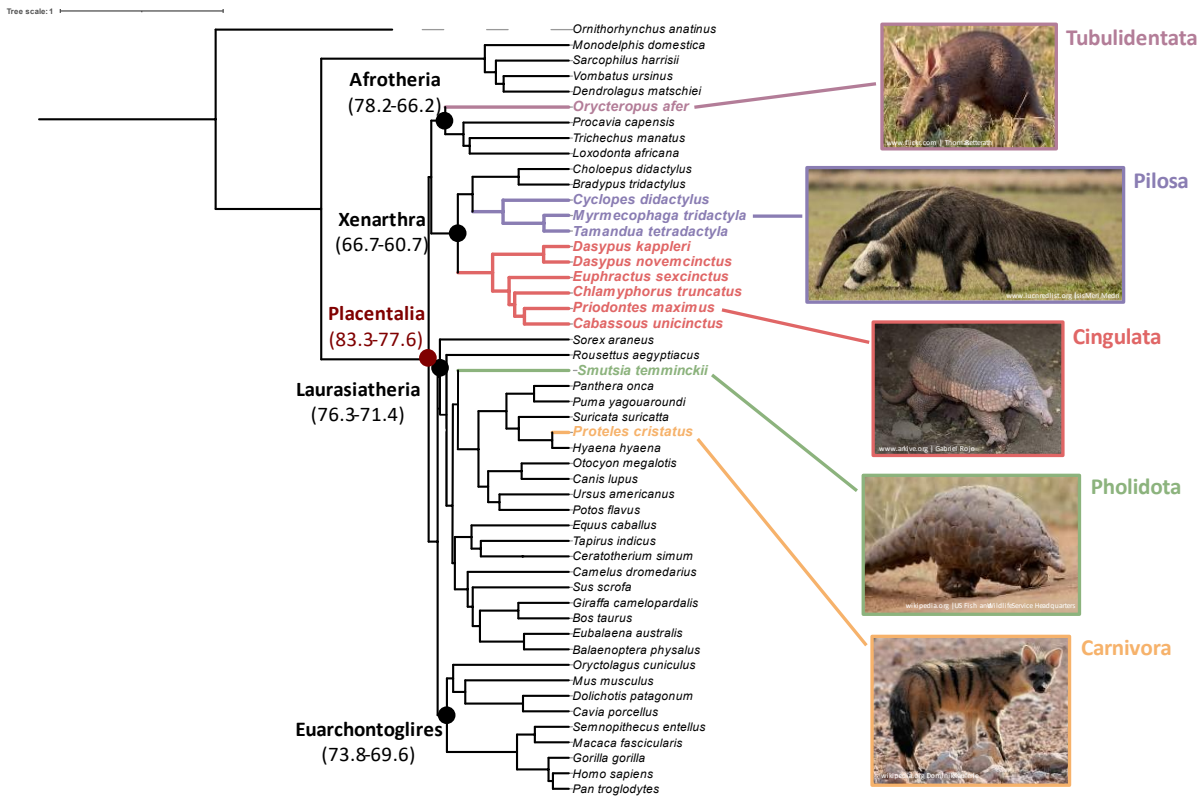


Figure 3. Phylogeny of myrmecophagous placentals. This phylogeny was obtained by pruning the species-level mammalian phylogeny reconstructed by Álvarez-Carretero *et al* (2022). Pholidota and Carnivora are not monophyletic in the phylogeny of Álvarez-Carretero *et al* although it is recognized that they share a common ancestor, the position of the Pholidota was modified manually to correct this. The placental node and the four main placental super-order nodes are indicated by circles. Their divergence times are in million years as estimated by Álvarez-Carretero *et al* (2022). Myrmecophagous species are colored according to their respective orders; whose names are indicated on the right of the figure.

Ant- and termite-eating mammals are characterized by remarkable convergently evolved morphological adaptations (Redford, 1987; Reiss, 2001). Their robust forearms and claws are used to dig into anthills and termite mounds (Griffiths, 1968). Most species also share an elongated skull with reduction or even complete loss of teeth, and modified jaw muscles (Reiss, 2001; Ferreira-Cardoso *et al*, 2019, 2020, 2022). Their hypertrophied salivary glands produce a lot of viscous saliva and their long, thin, and extensible tongue is used to catch and ingest prey without mastication thanks to its modified muscles, rather than for taste (Griffiths, 1968; Reiss, 2001; Casali *et al*, 2017). Absence of mastication is compensated by anatomical adaptations of their digestive tract usually possessing a strong muscular stomach (Patterson,

1975; Griffiths, 1968; Reiss, 2001). Because of their energetically poor diet and the ingestion of non-digestible material (*i.e*., soil), ant-eating species are also characterized by a low metabolic rate (McNab, 1984). However, not all species possess these marked anatomical characteristics. Pangolins and anteaters have the most extreme morphological adaptations (*e.g*., elongated skull and complete absence of teeth), whereas armadillos represent intermediate morphologies, and the aardwolf does not present marked adaptations typical of the myrmecophagous phenotype (Patterson, 1975; Reiss, 2001).

Differences in the degree of morphological specialization of these species could be explained by the species evolutionary history (*i.e*., phylogenetic constraints) and notably the timing of specialization toward myrmecophagy during their evolution. Ant- and termite-eating placentals include anciently diverged lineages such as pangolins, which diverged from Carnivora around ~40-26 Mya, or Xenarthrans diverging around ~66-60 Mya (with anteaters diverging ~30 Mya) (Álvarez-Carretero *et al*, 2022). Besides, ancient specialization toward ant and termite consumption in these lineages is supported by evidence from the fossil record. Indeed, *Eomanis* (~47 Mya; Storch, 1978), one of the earliest pangolin fossils, and *Palaeanodonta*, an extinct clade (60-30 Mya) potentially related to Pholidota, already presented morphological characteristics linked with myrmecophagy, such as the lack of teeth (Gaudin *et al*, 2009). Similar observations were made in anteater fossils (Gaudin and Braham, 1998). This evidence suggests that these species have been adapted to myrmecophagy for long times. Additionally, massive dental gene losses in anteaters (Emerling *et al*, 2023) confirm an ancient loss of teeth in anteaters and therefore ancient specialization toward this peculiar diet. On the contrary, aardwolves represent a more recently diverged myrmecophagous lineage with a divergence time from other hyaena species estimated around ~10 Mya (Eizirik *et al*, 2010). This suggests that aardwolves evolved toward myrmecophagy relatively more recently, which is consistent with the fact they do not present marked morphological myrmecophagous adaptations besides dental simplification (Reiss, 2001). Differences in time since myrmecophagous specialization could explain differences in the morphological and genomic adaptations observed in these species as well as the influence of phylogenetic constraints (*i.e*., historical contingency). Moreover, these divergences can explain the various adaptive paths taken by myrmecophagous species, which is something that will be discussed throughout the thesis.

3. The ConvergeAnt project and PhD's main goals

3.1. The ConvergeAnt project: an integrative approach to understand convergent adaptation to myrmecophagy in placental mammals

Myrmecophagous mammals represent a textbook example of convergence. Their independent convergent evolution toward the same highly specialized diet raises questions on the underlying mechanisms involved in their adaptation. **How myrmecophagous species convergently adapted to this diet? Were the same adaptive mechanisms involved between the different species?** The ConvergeAnt project (www.convergeant-project.com) proposes an integrative approach combining comparative morphology, genomics, and metagenomics to understand the underlying adaptive mechanisms involved in the evolution of convergent myrmecophagous phenotypes and how these three levels of study act together. The project first focused on the comparison of morphological adaptations linked to the myrmecophagous diet which were performed by Sergio Ferreira-Cardoso during his PhD and post-doc on the project. Comparative genomic analyses were first mainly conducted by Rémi Allio during his PhD and post-doc, and Christopher Emerling during his post-doc. A brief summary of the major findings revealed by the project at the time I started my PhD is given below.

Studies of the morphology of myrmecophagous species notably highlighted the different adaptive mechanisms involved between the different ant- and termite-eating species regarding their skull elongation, the masticatory muscles reduction of anteaters, and teeth loss in pangolins and anteaters (Ferreira-Cardoso *et al*, 2019, 2020, 2022). Comparative genomics gave further insights into molecular adaptations underlying teeth reduction in myrmecophagous species, notably through the study of genes involved in enamel and dentin development. Emerling *et al* (2023) found independent events of pseudogenization of 11 dental genes in Xenarthran species consistent with previous findings on losses of specific enamel and dental genes in toothless and enamelless mammals (*e.g.*, *ENAM,* Meredith *et al*, 2009; *ODAM,* Springer *et al*, 2019; *ACPT,* Mu *et al*, 2021). Additionally, comparative genomics focusing on chitinase genes (*i.e.*, enzymes able to degrade chitin, the main component of insects' exoskeleton) revealed that at least five chitinase paralogues were present in the last common ancestor of placentals and some were lost during the placental radiation especially

in non-insectivorous species resulting in a positive correlation between the percentage of invertebrates in the diet and the number of functional chitinase genes in the genome (Emerling *et al*, 2018). Myrmecophagous species carry different chitinase gene repertoires suggesting different molecular adaptations were involved in prey digestion, and were also probably constrained by their evolutionary history (Emerling *et al*, 2018). The genomic part of the ConvergeAnt project also aimed at generating high quality reference genomes of 11 ant-eating species and closely related species (see Allio, 2021; Allio *et al*, 2021). These genomes will help study in more detail candidate gene families of interest regarding adaptation to myrmecophagy, such as chitinase and taste receptor genes (see Chapter I). They will also be used to study global patterns of sequence convergence in coding and non-coding regions between species.

3.2. PhD research questions and main objectives

While the morphological part of the ConvergeAnt project was well advanced at the beginning of my PhD, questions remained regarding genomic adaptations underlying the myrmecophagous phenotype, and much remained to be done on the microbiome part as well as on understanding the links between the different parts of the project. The focus of my PhD project was thus to shed light on the **adaptations linked to the myrmecophagous diet both in the genomes and microbiomes of myrmecophagous species using comparative genomics and metagenomics to understand how these species perceive and digest their prey.**

The first chapter of this thesis presents the genomic part of the PhD project, which focuses on the study of two gene families, taste receptor and chitinase genes, to better understand their role in the convergent adaptation to myrmecophagy in placentals. The positive correlation between chitinase gene repertoires and diet found in Emerling *et al* (2018) also revealed discrepancies between species. More specifically, anteaters and pangolins have a diet composed of 100% of social insects but very different numbers of functional chitinase genes (Emerling *et al*, 2018) raising questions on how these gene repertoires are used to digest prey among the different myrmecophagous species. The main objective of the genomic part of my PhD was to understand the evolutionary history of chitinase genes in mammals and the role of the different paralogues in prey digestion using comparative genomics and

transcriptomics (see Chapter I part I.3). Preliminary analyses of placental taste receptor genes, conducted by Kathleen Garland during her Master degree under the supervision of Frédéric Delsuc and Christopher Emerling, revealed a reduction of taste receptor gene repertoires in ant-eating species suggesting reduced taste perception abilities (Garland, 2018). Part of my PhD project also aimed at understanding the evolutionary history of taste receptor genes in mammals and more specifically myrmecophagous species to study potential convergent patterns of gene loss/gain in relation to diet (see Chapter I part I.2).

The second chapter of this thesis focuses on the metagenomic part of the project. A previous study showed patterns of convergence in the gut microbiota taxonomic composition of myrmecophagous species compared to their non-myrmecophagous sister species (Delsuc *et al*, 2014). Yet, questions remained regarding the functions these gut symbionts ensure, notably in prey digestion and more specifically chitin digestion. Recent studies revealed the presence of chitinolytic bacteria carrying chitinase genes in the gut microbiota of one species of pangolin (*Manis javanica*) and of an anteater (*Myrmecophaga tridactyla*) (Ma *et al*, 2018; Cheng *et al*, 2023) as well as in mammalian species with a chitin-rich diet (*e.g*., Sanders *et al*, 2015) suggesting similar patterns might be observed in other myrmecophagous species. The main goal of my PhD was thus to study the potential role of the gut microbiota of myrmecophagous species in prey digestion by identifying chitinolytic bacteria and comparing their distribution among myrmecophagous species (see Chapter II part II.2).

Overall, by combining comparative genomics, transcriptomics, and metagenomics this PhD project aimed at providing insights into the adaptations underlying prey perception and digestion in myrmecophagous species. Moreover, studying the same gene family (*i.e*., chitinases) in both the host and its gut microbiota should shed light on the respective contributions of genomic adaptations of the host and its associated microbiome. Besides, this work should highlight the importance of combining complementary approaches to study such a system. Altogether, the integrative approach used during this PhD should help us understand the multiple and complex adaptive mechanisms underlying the convergent myrmecophagous phenotype in placental mammals, and whether the same or different mechanisms were involved. More generally, this work will shed light on the phenomenon of convergence as both natural selection in response to a similar diet, and constraints (*e.g*., historical contingency) might have played a role.

Personal contribution

During my PhD, I mainly worked on the metagenomic part of this project, notably on the manuscript presented in Chapter II part II.2 for which I did fieldwork, DNA extraction, sequencing, analyses, and writing. I also participated in the publication of the protocol presented in Chapter II part II.1.2. Another important part of my PhD project was to follow up on my Master project to work on the manuscript on chitinase comparative transcriptomics presented in Chapter I part I.3.2. Throughout my PhD, I collaborated with Christopher Emerling (Reedley College, Reedley, CA, USA) for projects on taste receptor evolution presented in Chapter I parts I.2.2 and I.2.3, with the later starting at the end of my PhD with the finalization of the dataset. I also collaborated with Guillaume Borrel (Institut Pasteur, Paris, France) with whom we started the project on the study of chitin digestion in the gut microbiota of placentals presented in Chapter II part II.3, for which the completion of the dataset and first analyses were done at the end of my PhD. Personal contribution on the different manuscripts is mentioned before each of the presented manuscript. Projects for which datasets were finalized and first preliminary analyses conducted at the end of the PhD will be pursued after the PhD.

Introduction references

Allio, R., 2021. Phylogenomics and comparative genomics in ant-eating mammals. Doctoral dissertation. University of Montpellier, Montpellier, France. English. NNT: 2021MONTG006. tel-03343948. https://theses.hal.science/tel-03343948

Allio, R., Tilak, M.-K., Scornavacca, C., Avenant, N.L., Kitchener, A.C., Corre, E., Nabholz, B., Delsuc, F., 2021. High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox. eLife 10, e63167.

Álvarez-Carretero, S., Tamuri, A.U., Battini, M., Nascimento, F.F., Carlisle, E., Asher, R.J., Yang, Z., Donoghue, P.C.J., dos Reis, M., 2022. A Species-Level Timeline of Mammal Evolution Integrating Phylogenomic Data. Nature 602, 263–267.

Archibald, J.D., Deutschman, D.H., 2001. Quantitative Analysis of the Timing of the Origin and Diversification of Extant Placental Orders. Journal of Mammalian Evolution 8, 107–124.

Arendt, J., Reznick, D., 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends in Ecology & Evolution 23, 26–32.

Baum, D., 2023. Parallelism-convergence continuum and its implication for homology assessment. Congress of the Society for Molecular Biology and Evolution (SMBE). Ferrara (Italy), 23rd-27th July 2023.

Berens, A.J., Hunt, J.H., Toth, A.L., 2015. Comparative Transcriptomics of Convergent Evolution: Different Genes but Conserved Pathways Underlie Caste Phenotypes across Lineages of Eusocial Insects. Molecular Biology and Evolution 32, 690–703.

Blount, Z.D., Borland, C.Z., Lenski, R.E., 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proc Natl Acad Sci U S A 105, 7899–7906.

Blount, Z.D., Barrick, J.E., Davidson, C.J., Lenski, R.E., 2012. Genomic Analysis of a Key Innovation in an Experimental *E. coli* Population. Nature 489, 513–518.

Blount, Z.D., Lenski, R.E., Losos, J.B., 2018. Contingency and determinism in evolution: Replaying life's tape. Science 362, eaam5979.

Blumer, M., Brown, T., Freitas, M.B., Destro, A.L., Oliveira, J.A., Morales, A.E., Schell, T., Greve, C., Pippel, M., Jebb, D., Hecker, N., Ahmed, A.-W., Kirilenko, B.M., Foote, M., Janke, A., Lim, B.K., Hiller, M., 2022. Gene losses in the common vampire bat illuminate molecular adaptations to blood feeding. Science Advances 8, eabm6494.

Brown, D.D., 2011. Fruit-Eating by an Obligate Insectivore: Palm Fruit Consumption in Wild Northern Tamanduas (*Tamandua mexicana*) in Panamá. eden 12, 63–65.

Carlisle, E., Janis, C.M., Pisani, D., Donoghue, P.C.J., Silvestro, D., 2023. A timescale for placental mammal diversification based on Bayesian modeling of the fossil record. Current Biology 33, 3073-3082.

Casali, D.M., Martins-Santos, E., Santos, A.L.Q., Miranda, F.R., Mahecha, G.A.B., Perini, F.A., 2017. Morphology of the tongue of Vermilingua (Xenarthra: Pilosa) and evolutionary considerations. Journal of Morphology 278, 1380–1399.

Chen, Y.-H., Zhao, H., 2019. Evolution of digestive enzymes and dietary diversification in birds. PeerJ 7, e6840.

Cheng, S.-C., Liu, C.-B., Yao, X.-Q., Hu, J.-Y., Yin, T.-T., Lim, B.K., Chen, W., Wang, G.-D., Zhang, C.-L., Irwin, D.M., Zhang, Z.-G., Zhang, Y.-P., Yu, L., 2023. Hologenomic insights into mammalian adaptations to myrmecophagy. National Science Review 10, nwac174.

Chikina, M., Robinson, J.D., Clark, N.L., 2016. Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. Mol Biol Evol 33, 2182–2192.

Christin, P.-A., Weinreich, D.M., Besnard, G., 2010. Causes and evolutionary significance of genetic convergence. Trends in Genetics 26, 400–405.

Cooper, C.E., 2011. *Myrmecobius fasciatus* (Dasyuromorphia: Myrmecobiidae). Mammalian Species 43, 129–140.

Conway-Morris, S., 1998. The Crucible of Creation: The Burgess Shale and the Rise of Animals. Oxford University Press.

Delsuc, F., Scally, M., Madsen, O., Stanhope, M.J., de Jong, W.W., Catzeflis, F.M., Springer, M.S., Douzery, E.J.P., 2002. Molecular Phylogeny of Living Xenarthrans and the Impact of Character and Taxon Sampling on the Placental Tree Rooting. Mol Biol Evol 19, 1656–1671.

Delsuc, F., Mauffrey, J.-F., Douzery, E., 2003. Une nouvelle classification des mammifères. Pour la science 303, 62–66.

Delsuc, F., Metcalf, J.L., Wegener Parfrey, L., Song, S.J., González, A., Knight, R., 2014. Convergence of gut microbiomes in myrmecophagous mammals. Molecular Ecology 23, 1301–1317.

Duchemin, L., Lanore, V., Veber, P., Boussau, B., 2023. Evaluation of Methods to Detect Shifts in Directional Selection at the Genome Scale. Mol Biol Evol 40, msac247.

Eizirik, E., Murphy, W.J., Koepfli, K.-P., Johnson, W.E., Dragoo, J.W., Wayne, R.K., O'Brien, S.J., 2010. Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. Molecular Phylogenetics and Evolution 56, 49–63.

Ellegren, H., 2008. Comparative genomics and the study of evolution by natural selection. Molecular Ecology 17, 4586–4596.

Elmer, K.R., Meyer, A., 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. Trends in Ecology & Evolution 26, 298–306.

Emerling, C.A., 2017. Genomic regression of claw keratin, taste receptor and light-associated genes provides insights into biology and evolutionary origins of snakes. Mol Phylogenet Evol 115, 40–49.

Emerling, C.A., Delsuc, F., Nachman, M.W., 2018. Chitinase genes (*CHIAs*) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. Science Advances 4, eaar6478.

Emerling, C.A., Gibb, G.C., Tilak, M.-K., Hughes, J.J., Kuch, M., Duggan, A.T., Poinar, H.N., Nachman, M.W., Delsuc, F., 2023. Genomic data suggest parallel dental vestigialization within the xenarthran radiation. Peer Community Journal 3, e75.

Fay, J.C., Wu, C.-I., 2001. The neutral theory in the genomic era. Current Opinion in Genetics & Development 11, 642–646.

Feng, P., Zheng, J., Rossiter, S.J., Wang, D., Zhao, H., 2014. Massive Losses of Taste Receptor Genes in Toothed and Baleen Whales. Genome Biology and Evolution 6, 1254–1265.

Ferreira-Cardoso, S., Delsuc, F., Hautier, L., 2019. Evolutionary Tinkering of the Mandibular Canal Linked to Convergent Regression of Teeth in Placental Mammals. Current Biology 29, 468-475.e3.

Ferreira-Cardoso, S., Fabre, P.-H., Thoisy, B. de, Delsuc, F., Hautier, L., 2020. Comparative masticatory myology in anteaters and its implications for interpreting morphological convergence in myrmecophagous placentals. PeerJ 8, e9690.

Ferreira-Cardoso, S., Claude, J., Goswami, A., Delsuc, F., Hautier, L., 2022. Flexible conservatism in the skull modularity of convergently evolved myrmecophagous placental mammals. BMC Ecology and Evolution 22, 87.

Foley, N.M., Mason, V.C., Harris, A.J., Bredemeyer, K.R., Damas, J., Lewin, H.A., Eizirik, E., Gatesy, J., Karlsson, E.K., Lindblad-Toh, K., Zoonomia Consortium, Springer, M.S., Murphy, W.J., 2023. A genomic timescale for placental mammal evolution. Science 380, eabl8189.

Foote, A.D., Liu, Y., Thomas, G.W.C., Vinař, T., Alföldi, J., Deng, J., Dugan, S., van Elk, C.E., Hunter, M.E., Joshi, V., Khan, Z., Kovar, C., Lee, S.L., Lindblad-Toh, K., Mancia, A., Nielsen, R., Qin, X., Qu, J., Raney, B.J., Vijay, N., Wolf, J.B.W., Hahn, M.W., Muzny, D.M., Worley, K.C., Gilbert, M.T.P., Gibbs, R.A., 2015. Convergent evolution of the genomes of marine mammals. Nature Genetics 47, 272–275.

Fukushima, K., Pollock, D.D., 2023. Detecting macroevolutionary genotype-phenotype associations using error-corrected rates of protein convergence. Nat Ecol Evol 7, 155–170.

Gallo, J., Abba, A., Elizalde, L., Di Nucci, D., Rios, T., Ezquiaga, M., 2017. First study on food habits of anteaters, *Myrmecophaga tridactyla* and *Tamandua tetradactyla*, at the southern limit of their distribution. Mammalia 81.

Garland, T., Bennett, A.F., Rezende, E.L., 2005. Phylogenetic approaches in comparative physiology. Journal of Experimental Biology 208, 3015–3035.

Garland, K., (2018). Evolution of bitter taste receptors in convergent myrmecophagous placentals. Master thesis. University of Montpellier.

Gaudin, T.J., Branham, D.G., 1998. The Phylogeny of the Myrmecophagidae (Mammalia, Xenarthra, Vermilingua) and the Relationship of *Eurotamandua* to the Vermilingua. Journal of Mammalian Evolution 5, 237–265.

Gaudin, T.J., Emry, R.J., Wible, J.R., 2009. The Phylogeny of Living and Extinct Pangolins (Mammalia, Pholidota) and Associated Taxa: A Morphology Based Analysis. J Mammal Evol 16, 235–305.

Gompel, N., Prud'homme, B., 2009. The causes of repeated genetic evolution. Dev. Biol. 332, 36–47.

Gould, S.J., 1989. Wonderful Life: The Burgess Shale and the Nature of History. W. W. Norton.

Griffiths, M., 1968. Echidnas. Pergamon Press, Oxford; New York.

Hughes, G.M., Boston, E.S.M., Finarelli, J.A., Murphy, W.J., Higgins, D.G., Teeling, E.C., 2018. The Birth and Death of Olfactory Receptor Gene Families in Mammalian Niche Adaptation. Mol Biol Evol 35, 1390–1406.

Jacob, F., 1977. Evolution and tinkering. Science 196, 1161–1166.

Jeffares, D.C., Tomiczek, B., Sojo, V., dos Reis, M., 2015. A Beginners Guide to Estimating the Non-synonymous to Synonymous Rate Ratio of all Protein-Coding Genes in a Genome, in: Peacock, C. (Ed.), Parasite Genomics Protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 65–90.

Jiang, P., Josue, J., Li, X., Glaser, D., Li, W., Brand, J.G., Margolskee, R.F., Reed, D.R., Beauchamp, G.K., 2012. Major taste loss in carnivorous mammals. PNAS 109, 4956–4961.

Kaeuffer, R., Peichel, C.L., Bolnick, D.I., Hendry, A.P., 2012. Parallel and non-parallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. Evolution 66, 402–418.

Kowalczyk, A., Meyer, W.K., Partha, R., Mao, W., Clark, N.L., Chikina, M., 2019. RERconverge: an R package for associating evolutionary rates with convergent traits. Bioinformatics 35, 4815–4817.

Lee, J.H., Lewis, K.M., Moural, T.W., Kirilenko, B., Borgonovo, B., Prange, G., Koessl, M., Huggenberger, S., Kang, C., Hiller, M., 2018. Molecular parallelism in fast-twitch muscle proteins in echolocating mammals. Science Advances 4, eaat9660.

Lenski, R.E., Rose, M.R., Simpson, S.C., Tadler, S.C., 1991. Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. The American Naturalist 138, 1315–1341.

Lenski, R.E., 2017. Convergence and Divergence in a Long-Term Experiment with Bacteria. The American Naturalist 190, S57–S68.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., Gordon, J.I., 2008. Evolution of Mammals and Their Gut Microbes. Science 320, 1647–1651.

Losos, J.B., Jackman, T.R., Larson, A., Queiroz, K. de, Rodríguez-Schettino, L., 1998. Contingency and Determinism in Replicated Adaptive Radiations of Island Lizards. Science 279, 2115–2118.

Losos, J.B., 2011. Convergence, adaptation, and constraint. Evolution 65, 1827–1840.

Losos, J.B., 2017. Improbable Destinies: Fate, Chance, and the Future of Evolution. Penguin.

Luo, Z.-X., 2007. Transformation and diversification in early mammal evolution. Nature 450, 1011–1019.

Ma, J.-E., Jiang, H.-Y., Li, L.-M., Zhang, X.-J., Li, G.-Y., Li, H.-M., Jin, X.-J., Chen, J.-P., 2018. The Fecal Metagenomics of Malayan Pangolins Identifies an Extensive Adaptation to Myrmecophagy. Front. Microbiol. 9, 2793.

Mahler, D.L., Ingram, T., 2014. Phylogenetic Comparative Methods for Studying Clade-Wide Convergence, in: Garamszegi, L.Z. (Ed.), Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Springer, Berlin, Heidelberg, pp. 425–450.

Mahler, D.L., Weber, M.G., Wagner, C.E., Ingram, T., 2017. Pattern and Process in the Comparative Study of Convergent Evolution. The American Naturalist 190, S13–S28.

Manceau, M., Domingues, V.S., Linnen, C.R., Rosenblum, E.B., Hoekstra, H.E., 2010. Convergence in pigmentation at multiple levels: mutations, genes, and function. Philos Trans R Soc Lond B Biol Sci 365, 2439–2450.

McGhee, G.R., 2011. Convergent Evolution: Limited Forms Most Beautiful. MIT Press.

McNab, B.K., 1984. Physiological convergence amongst ant-eating and termite-eating mammals. Journal of Zoology 203, 485–510.

Mendoza, M.L.Z., Xiong, Z., Escalera-Zamudio, M., Runge, A.K., Thézé, J., Streicker, D., Frank, H.K., Loza-Rubio, E., Liu, Shengmao, Ryder, O.A., Samaniego Castruita, J.A., Katzourakis, A., Pacheco, G., Taboada, B., Löber, U., Pybus, O.G., Li, Y., Rojas-Anaya, E., Bohmann, K., Carmona Baez, A., Arias, C.F., Liu, Shiping, Greenwood, A.D., Bertelsen, M.F., White, N.E., Bunce, M., Zhang, G., Sicheritz-Pontén, T., Gilbert, M.P.T., 2018. Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. Nat Ecol Evol 2, 659–668.

Meredith, R.W., Gatesy, J., Murphy, W.J., Ryder, O.A., Springer, M.S., 2009. Molecular Decay of the Tooth Gene Enamelin (*ENAM*) Mirrors the Loss of Enamel in the Fossil Record of Placental Mammals. PLOS Genetics 5, e1000634.

Meredith, R.W., Janečka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A., Eizirik, E., Simão, T.L.L., Stadler, T., Rabosky, D.L., Honeycutt, R.L., Flynn, J.J., Ingram, C.M., Steiner, C., Williams, T.L., Robinson, T.J., Burk-Herrick, A., Westerman, M., Ayoub, N.A., Springer, M.S., Murphy, W.J., 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. Science 334, 521–524.

Miranda, F., Veloso, R., Superina, M., Zara, F.J., 2009. Food Habits of Wild Silky Anteaters (*Cyclopes didactylus*) of São Luis do Maranhão, Brazil. eden 2009, 1–5.

Mu, Y., Huang, X., Liu, R., Gai, Y., Liang, N., Yin, D., Shan, L., Xu, S., Yang, G., 2021. *ACPT* gene is inactivated in mammalian lineages that lack enamel or teeth. PeerJ 9, e10219.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R., Gordon, J.I., 2011. Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans. Science 332, 970–974.

Orgogozo, V., 2015. Replaying the tape of life in the twenty-first century. Interface Focus 5, 20150057.

Panaino, W., Parrini, F., Phakoago, M.V., Smith, D., van Dyk, G., Fuller, A., 2022. Do seasonal dietary shifts by Temminck's pangolins compensate for winter resource scarcity in a semi-arid environment? Journal of Arid Environments 197, 104676.

Partha, R., Chauhan, B.K., Ferreira, Z., Robinson, J.D., Lathrop, K., Nischal, K.K., Chikina, M., Clark, N.L., 2017. Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. eLife 6, e25884.

Partha, R., Kowalczyk, A., Clark, N.L., Chikina, M., 2019. Robust Method for Detecting Convergent Shifts in Evolutionary Rates. Mol Biol Evol 36, 1817–1830.

Patterson, B., 1975. The Fossil Aardvarks (Mammalia: Tubulidentata). Bull. Mus. Comp. Zool. 147, 185–237.

Pearce, T., 2012. Convergence and Parallelism in Evolution: A Neo-Gouldian Account. The British Journal for the Philosophy of Science 63, 429–448.

Pietersen, D.W., Symes, C.T., Woodborne, S., McKechnie, A.E., Jansen, R., 2016. Diet and prey selectivity of the specialist myrmecophage, Temminck's ground pangolin. Journal of Zoology 298, 198–208.

Powell, R., 2012. Convergent evolution and the limits of natural selection. Euro Jnl Phil Sci 2, 355–373.

Powell, R., Mariscal, C., 2015. Convergent evolution as natural experiment: the tape of life reconsidered. Interface Focus 5, 20150040.

Redford, K.H., 1986. Dietary specialization and variation in two mammalian myrmecophages (variation in mammalian myrmecophagy). Revista Chilena de Historia Natural 59, 201–208.

Redford, K.H., 1987. Ants and Termites As Food, in: Genoways, H.H. (Ed.), Current Mammalogy. Springer US, Boston, MA, pp. 349–399.

Reiss, K.Z., 2001. Using Phylogenies to Study Convergence: The Case of the Ant-Eating Mammals. Integr Comp Biol 41, 507–525.

Rey, C., Lanore, V., Veber, P., Guéguen, L., Lartillot, N., Sémon, M., Boussau, B., 2019. Detecting adaptive convergent amino acid evolution. Philosophical Transactions of the Royal Society B: Biological Sciences 374, 20180234.

Rismiller, P.D., Grutzner, F., 2019. *Tachyglossus aculeatus* (Monotremata: Tachyglossidae). Mammalian Species 51, 75–91.

Sanders, J.G., Beichman, A.C., Roman, J., Scott, J.J., Emerson, D., McCarthy, J.J., Girguis, P.R., 2015. Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. Nature Communications 6, 8285.

Shan, L., Wu, Q., Wang, L., Zhang, L., Wei, F., 2018. Lineage-specific evolution of bitter taste receptor genes in the giant and red pandas implies dietary adaptation. Integrative Zoology 13, 152–159.

Sharma, V., Varshney, R., Sethy, N.K., 2022. Human adaptation to high altitude: a review of convergence between genomic and proteomic signatures. Hum Genomics 16, 21.

Shen, Y.-Y., Liang, L., Li, G.-S., Murphy, R.W., Zhang, Y.-P., 2012. Parallel Evolution of Auditory Genes for Echolocation in Bats and Toothed Whales. PLOS Genetics 8, e1002788.

Springer, M.S., Murphy, W.J., Eizirik, E., O'Brien, S.J., 2003. Placental mammal diversification and the Cretaceous–Tertiary boundary. PNAS 100, 1056–1061.

Springer, M.S., Stanhope, M.J., Madsen, O., de Jong, W.W., 2004. Molecules consolidate the placental mammal tree. Trends in Ecology & Evolution 19, 430–438.

Springer, M.S., Emerling, C.A., Meredith, R.W., Janečka, J.E., Eizirik, E., Murphy, W.J., 2017. Waking the undead: Implications of a soft explosive model for the timing of placental mammal diversification. Molecular Phylogenetics and Evolution 106, 86–102.

Springer, M.S., Emerling, C.A., Gatesy, J., Randall, J., Collin, M.A., Hecker, N., Hiller, M., Delsuc, F., 2019. Odontogenic ameloblast-associated (*ODAM*) is inactivated in toothless/enamelless placental mammals and toothed whales. BMC Evolutionary Biology 19, 31.

Stayton, C.T., 2015a. The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. Evolution 69, 2140–2153.

Stayton, C.T., 2015b. What does convergent evolution mean? The interpretation of convergence and its implications in the search for limits to evolution. Interface Focus 5, 20150039.

Stern, D.L., Orgogozo, V., 2009. Is Genetic Evolution Predictable? Science 323, 746–751.

Storch, G., 1978. *Eomanis waldi*, ein Schuppentier aus dem Mittel-Eozän der "Grube Messel" bei Darmstadt (Mammalia: Pholidota). Senckenberg lethaea 59, 503–529.

Storz, J.F., 2016. Causes of molecular convergence and parallelism in protein evolution. Nature Reviews Genetics 17, 239–250.

Sun, N.C.-M., Lin, C.-C., Liang, C.-C., Li, H.-F., 2022. Diet composition of an escaped captive-born southern tamandua (*Tamandua tetradactyla*) in a nonnative habitat in Asia. Ecology and Evolution 12, e9175. h

Swart, J.M., Richardson, P.R.K., Ferguson, J.W.H., 1999. Ecological factors affecting the feeding behaviour of pangolins (*Manis temminckii*). Journal of Zoology 247, 281–292.

Vaz, V.C., Santori, R.T., Jansen, A.M., Delciellos, A.C., D'Andrea, P.S., 2012. Notes on Food Habits of Armadillos (Cingulata, Dasypodidae) and Anteaters (Pilosa, Myrmecophagidae) at Serra Da Capivara National Park (Piauí State, Brazil). eden 13, 84–89.

Wang, K., Tian, S., Galindo-González, J., Dávalos, L.M., Zhang, Y., Zhao, H., 2020. Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. Mol Ecol 29, 4366–4381.

Wang, Y., Yang, L., 2021. Genomic Evidence for Convergent Molecular Adaptation in Electric Fishes. Genome Biol Evol 13, evab038.

Waters, J.M., McCulloch, G.A., 2021. Reinventing the wheel? Reassessing the roles of gene flow, sorting and convergence in repeated evolution. Molecular Ecology 30, 4162–4172.

Weyer, N.M., 2018. Physiological flexibility of free-living aardvarks (*Orycteropus afer*) in response to environmental fluctuations. Doctoral dissertation. University Witwatersrand, Witwatersrand, South Africa.

Wildman, D.E., Uddin, M., Opazo, J.C., Liu, G., Lefort, V., Guindon, S., Gascuel, O., Grossman, L.I., Romero, R., Goodman, M., 2007. Genomics, biogeography, and the diversification of placental mammals. PNAS 104, 14395–14400.

Yang, Z., Bielawski, J.P., 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. (Amst.) 15, 496–503.

Zancolli, G., Reijnders, M., Waterhouse, R.M., Robinson-Rechavi, M., 2022. Convergent evolution of venom gland transcriptomes across Metazoa. Proceedings of the National Academy of Sciences 119, e2111392119.

Zhou, Z., and Zhang, J., (2020). The Nature and Phylogenomic Impact of Sequence Convergence. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, Phylogenetics in the Genomic Era, chapter No. 4.6, pp. 4.6:1–4.6:17. No commercial publisher | Authors open access book. The book is freely available at https://hal.inria.fr/PGE.

CHAPTER I

GENOMIC ADAPTATIONS TO THE MYRMECOPHAGOUS DIET IN MAMMALS

I.1. Chapter introduction

Understanding genomic adaptations through the study of gene families

I.1.1. Evolution of multigenic families

Gene families and their evolution

Gene families (or multigenic families) are composed of a set of homologous genes (*i.e*., genes that share a common evolutionary origin; Koonin, 2005), which evolved from an ancestral gene through successive duplications (Nei and Rooney, 2005). Two types of homologous genes can be distinguished depending on the processes that occurred during their evolution from an ancestral gene: i) paralogous genes evolved from a common ancestral gene through successive duplications, and ii) orthologous genes evolved from a common ancestral gene through speciation during which each copy is found in each of the two new species (Koonin, 2005) (Fig I.1.A).

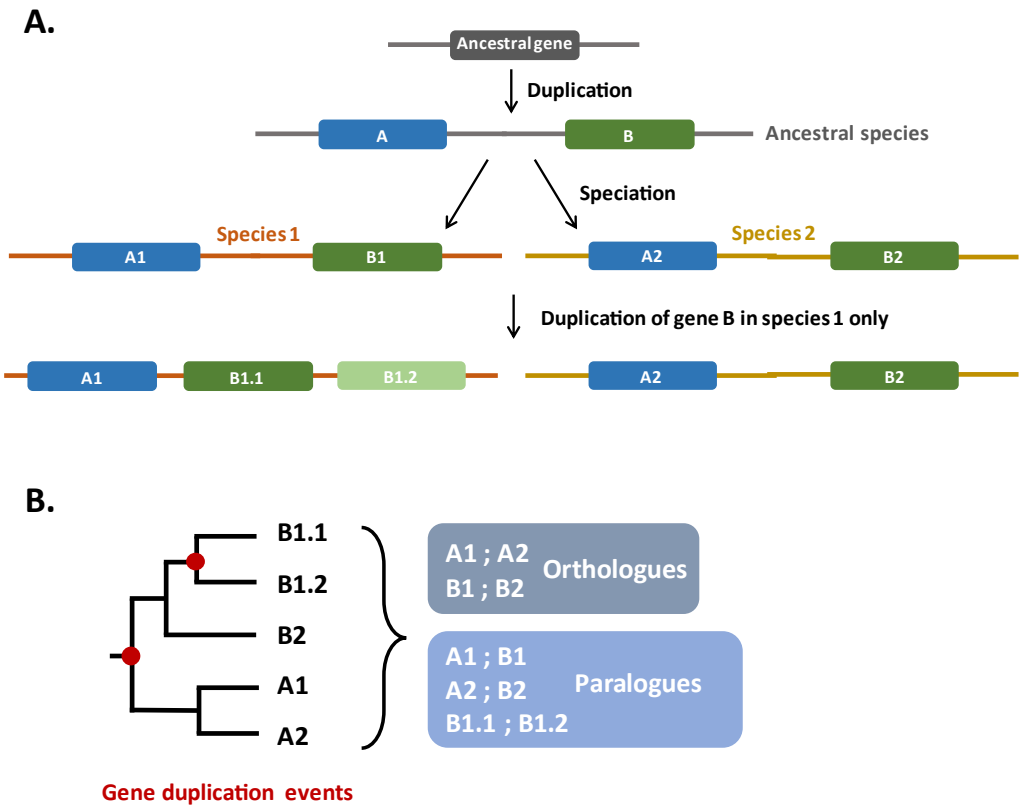


Figure I.1. Evolution of orthologous and paralogous genes within a multigenic family. A. Evolution of paralogues and orthologues from an ancestral gene. B. Phylogenetic relationships of paralogous and orthologous genes within a multigenic family. Adapted from Koonin (2005).

The different gene copies composing a multigenic family evolve according to different evolutionary processes usually classified in three historical main models of multigenic evolution which are not mutually exclusive and often occur altogether:

- **Divergent model**: it was first described by Ohno (1970). In this model, some copies under positive selection can i) acquire new functions (*i.e.*, neo-functionalization), and ii) mutations can accumulate in some copies which become non-functional (*i.e.*, pseudogenization). To complexify this model, the duplication-degeneration-complementation model has been proposed and adds sub-functionalization that is functions ensured by the ancestral genes can be split between the new copies. Related to this, gene copies can all keep the same function but some can acquire specific cellular or tissular expressions (*i.e.*, cellular or tissular adaptations) (Hughes, 1994; Force *et al*, 1999; Koonin, 2005; Nei and Rooney, 2005).

For example, such a model of gene family evolution played an important role in the evolution of color vision in vertebrates. Indeed, gene duplication followed by neo-functionalization of the opsin gene enabled trichromatic vision in primates (Hunt *et al*, 1998; Dulai *et al*, 1999; Carvalho *et al*, 2017). Sub-functionalization of gene copies with specific tissular expressions was observed in plants with genes of the chalcone synthase family (that participates in the biosynthesis of flavonoid pigments) for which the different copies are expressed in different organs depending on the developmental stage (Han *et al*, 2006). In humans, tissular specialization of gene copies has for instance been described for the glutamate dehydrogenase for which one copy is expressed in several tissues and the other in nerve tissues specifically (Burki and Kaessmann, 2004).

- **Concerted evolution**: this phenomenon occurs during meiosis and homogenizes the different paralogous gene copies through repeated inequal recombinations, also called gene conversion (Nei and Rooney, 2005). Genes involved in color vision were also subjected to gene conversion homogenizing paralogues, for instance in primates (Zhou and Li, 1996; Hunt *et al*, 1998; Carvalho *et al*, 2017). In teleost fishes, opsin genes seem to have undergone a complex evolutionary dynamic involving gene losses and duplications as well as gene conversion with potential cases of "resurrection" of pseudogenized copies via conversion with functional ones (Cortesi *et al*, 2015).

- **Birth-and-death model**: depending on the lineages, the number of duplications and pseudogenizations within a multigenic family can differ (Nei and Hughes, 1992, cited in Nei and Rooney, 2005). Most multigenic families (e.*g*., immune system genes, sensory genes, genes involved in the development or highly conserved housekeeping genes) could be evolving according to this process (Nei and Rooney, 2005).

It should be noted that several models of evolution after gene duplication have now been described to better account for the complexity of multigenic evolution and shed light on how it might affect an organism's fitness. To better understand how gene duplications can be fixed and the effects they might have on an organism, Innan and Kondrashov (2010) proposed a

classification of the different models of multigenic family evolution according to the fate of copies after gene duplication, from the original mutation giving birth to the new copy, through the accumulation of changes between copies to the fixation of these changes in a population, and the contribution of neutral, positive, and relaxed selection during these different stages. For example, having several gene copies can be beneficial as some copies can compensate for deleterious mutations arising in others (Innan and Kondrashov, 2010). Gene dosage effect, in which several copies of the same gene can increase the number of produced proteins, has also been described (Innan and Kondrashov, 2010), complexifying the classic divergent model, and might explain why some duplicates are maintained besides neo- and sub-functionalization (Conant *et al*, 2014).

Methods to study gene families' evolution

Different methods are used to study the evolution of multigenic families (Koonin, 2005). Reconstructing gene trees is often the first step to study gene families as it enables to identify paralogous and orthologous genes within the phylogeny (Fig I.1.B). Indeed, comparing sequence divergence between gene copies and analyzing gene tree topologies can help understand the processes that shaped multigenic families (Koonin 2005; Nei and Rooney, 2005). For example, in the case of divergent evolution where each duplicated gene is transmitted to the descendance of a species after speciation, the gene tree retraces the species tree (Fig I.2.A). When concerted evolution happens it homogenizes the different paralogues making their sequences more similar and erasing their evolutionary history (Fig I.2.B). Finally, during the birth-and-death process, specific gene duplications and losses occur resulting in species having different number of gene copies (Fig I.2.C).

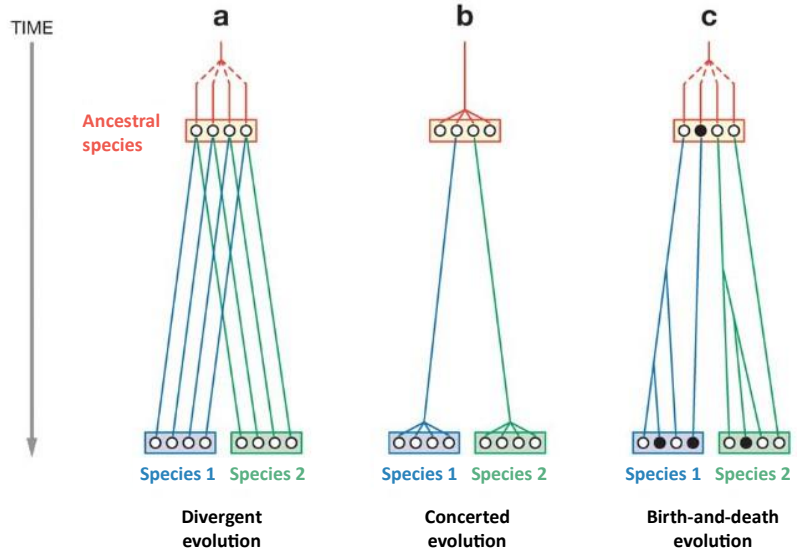


Figure I.2. Topologies of hypothetical gene trees under the three main models of evolution of multigenic families (divergent (a), concerted (b), and birth-and-death evolution (c)). Open circles indicate functional genes and close circles pseudogenes. From Nei and Rooney (2005).

To infer the number of duplications and pseudogenizations in different lineages, reconciliation methods can be used and are based on the comparison of gene trees against the species trees. Indeed, genes evolve within species and are thus subject to the speciation and extinction processes specific to species evolutionary histories, as well as biological processes impacting gene dynamics. These processes include gene duplications, pseudogenizations, gene conversions, Horizontal Gene Transfers (HGT; often occurring in prokaryotes), hybridization (the hybrid genome is a combination of two parental genomes and might have inherited different genes copies due to recombination), and Incomplete Lineage Sorting (ILS; occurring when speciation happens rapidly so the ancestral polymorphism cannot be sorted out between the two new species) (Boussau and Scornavacca, 2020; Schrempf and Szöllősi, 2020). Together these processes shape different and independent evolutionary histories of the species and their genes resulting in discrepancies between gene trees and the species tree (*i.e.*, phylogenetic conflict; Fig I.3) (Boussau and Scornavacca, 2020; Schrempf and Szöllősi, 2020). Reconciliation methods have been developed to model these different processes based on parsimony or probabilistic approaches to resolve phylogenetic conflicts (Doyon *et al*, 2011; Szöllősi *et al*, 2015; Boussau and Scornavacca, 2020; Schrempf and Szöllősi, 2020).

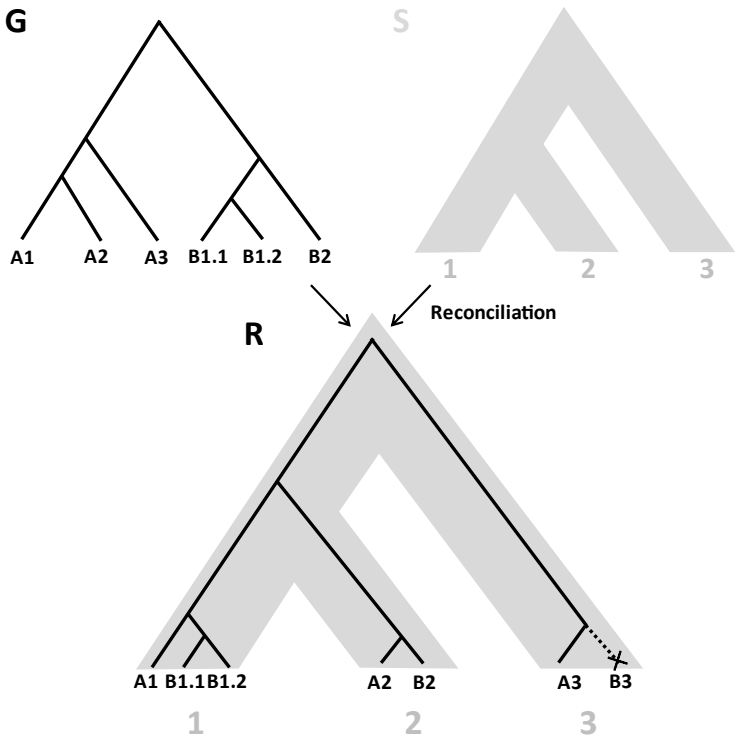


Figure I.3. Phylogenetic conflict between species and gene trees. G represents the gene tree (with genes A and B), S the species tree (with species 1, 2, and 3) and R the reconciled gene tree. Dashed line indicated gene loss of B3. Adapted from Boussau and Scornavacca (2020).

When using reconciliation methods, pseudogenization events are inferred but pseudogene sequences cannot be studied. To better understand the evolution of gene copies within a lineage or even within a species, genes can be identified in genome assemblies by mapping or blasting a reference gene of a closest relative. Once genes are identified their location within the genome (*i.e.*, synteny) can be compared between species (*i.e.*, comparative genomics) using orthologous genes not belonging to the gene family as reference points to place other genes (*e.g.*, Emerling *et al*, 2018; Leurs *et al*, 2021). This can for instance reveal inversions of genes that could occur during recombination, as well as tandem duplications in which paralogous genes are found one next to each other (Fig I.4).

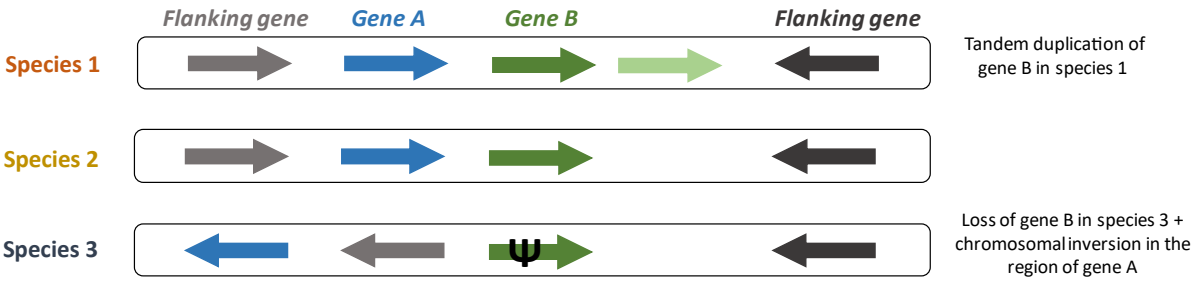


Figure I.4. Hypothetical synteny comparison of two paralogous genes across three different species. Arrows and their direction respectively represent the genes and their transcription direction. "Ψ" symbol indicates a pseudogene.

Additionally, aligning orthologous gene copies can allow the identification of inactivating mutations like splice-site mutations, frameshifts, and/or premature stop codons within the protein coding sequence. Identifying such inactivating mutations enables the characterization of pseudogenes and highlights putative molecular signatures linked to specific adaptations that might be shared across species or be specific to certain lineages. Some programs can be useful to detect frameshifts and stop codons as they consider the Open Reading Frame (ORF) when aligning multiple sequences. For example, MACSE has been designed to align datasets containing protein-coding sequences of genes including non-functional sequences (*i.e.*, pseudogenes) by allowing frameshifts to avoid disrupting the codon structure (Ranwez *et al*, 2011). It has notably enabled the identification of frameshifts not yet detected in public databases and the mapping of next-generation sequences containing introns and potentially sequencing errors against protein-coding references (Ranwez *et al*, 2011). MACSE is for instance used within the Pseudo$_{\text{Checker}}$ pipeline, which allows the study of a gene functionality by aligning target sequences to reference exon-annotated coding sequences and reporting inactivating mutations to compute a pseudogenization score (Alves, 2020).

To understand how gene copies, and the potential functions they ensure, evolved within a multigenic family, several omics approaches (*e.g.*, genomics, transcriptomics, proteomics) can be combined. For example, *in situ* hybridization, when possible, enables the identification of expression sites of specific gene copies, including during development, and makes possible inferences on their putative functions (*e.g.*, Leurs *et al*, 2021; Romero *et al*,

2021). When such experimentation cannot be conducted, transcriptomic data become a valuable resource to study the location and expression level of different gene copies. Indeed, expression differences might indicate potential sub-functionalization, tissular adaptations, and/or dosage effect acting on gene copies, and therefore uncover the functions of the different paralogues. Combined with reconciliation methods and syntenic comparisons, these approaches can greatly improve our understanding of gene families' evolution as it has been done for example to study genes involved in biomineralization of cartilaginous fishes (Leurs *et al*, 2021).

Finally, to fully characterize the evolution of paralogous gene functions, ancestral sequences can be reconstructed. This is usually done by first inferring ancestral sequences at each node of a phylogeny, for instance using maximum likelihood methods (Thornton, 2004; Dube *et al*, 2022), as done within RAxML-NG (Kozlov *et al*, 2019). Studying the sequence structure and domains (*e.g*., catalytic and binding sites) then enables predictions on the protein functions. These ancient genes can then be resurrected by synthetizing their sequences and cloning them into cells that will produce the ancestral proteins (Thornton, 2004; Dube *et al*, 2022). Such approach enables the realization of multiple *in vitro* functional assays under diverse conditions to fully describe protein functions. Besides, this approach can help characterizing the influence of certain substitutions on the protein function and thus to understand how they evolved (Thornton, 2004; Dube *et al*, 2022). Ancestral gene resurrection has been used, for instance, to study key mutations that influenced the evolution of the maximum absorption wavelength of opsins (Yokoyama *et al*, 2008), or the hemoglobin affinity to oxygen in high-altitude birds (Natarajan *et al*, 2016).

The use of these complementary methods (*i.e*., comparative genomic and transcriptomic, reconciliation, ancestral sequence reconstruction, synteny analyses) to understand the evolutionary dynamic of two gene families (taste receptor and chitinase genes) will be further developed and discussed in the following parts of this chapter.

I.1.2. Studying multigenic families to understand molecular adaptations

Evolution of gene families can be studied to understand molecular adaptations to specific selective pressures and more broadly can help understand evolutionary processes like convergence. Studying duplication and pseudogenization events that occurred during the evolution of a gene family can enable the identification of those associated with changes in selective pressures (*e.g.*, environmental changes, lifestyle changes) (Demuth and Hahn, 2009). For instance, in mammals, several studies focused on olfactory and gustatory gene families have shown that these multigenic families follow a birth-and-death evolution and that this might be correlated with ecological characteristics such as lifestyle or diet (Li and Zhang, 2014; Hughes *et al*, 2018). Indeed, dietary adaptations seem to shape the evolution of gene families as shown, for instance, for the chitinase gene family in placentals with gene losses occurring disproportionally in non-insectivorous species (Emerling *et al*, 2018; Janiak *et al*, 2018).

Gene losses might indeed be adaptive (in response to a change in selective pressure) or neutral if the function is dispensable in the new environment (Albalat and Cañestro, 2016; Sharma *et al*, 2018a). Several studies have focused on specific gene losses linked with morphological, physiological, or metabolic adaptations in response to environmental, lifestyle and diet changes (*e.g.*, Sharma *et al*, 2018b; Hecker *et al*, 2019; Jiao *et al*, 2019; Springer *et al*, 2019). Gene losses related to independent trait losses are increasingly studied and methods have been developed to identify them (Hiller *et al*, 2012; Prudent *et al*, 2016). For instance, the adaptation to the subterranean life has impacted gene repertoires linked with this peculiar lifestyle. Several genes involved in eyes development, light perception, and circadian rhythm have been lost in subterranean blind mammals that lost vision (Prudent *et al*, 2016). In subterranean beetles, losses occurred in olfactory and gustatory genes and a duplication was identified in a gene involved in thermic and humidity perception (Balart-García *et al*, 2021).

Additionally, when working within a phylogenetic context, gene losses and gains can be dated (for example using dN/dS ratio and inactivating mutations to identify pseudogenes) enabling us to make hypotheses about ancestral states and the number of events that occurred in some lineages and whether they might correlate with specific adaptations. This has been done, for instance, to study convergent evolution of osmotrophic lifestyles among

opisthokonts (Torruella *et al*, 2015), or transitions towards insectivory in placentals (Emerling *et al*, 2018).

All together, these studies highlight the importance of analyzing global evolution of gene families, as well as more specific patterns of gene losses, and incorporate species ecological characteristics to fully understand how gene families are involved in the adaptation of organisms to environmental and lifestyle changes.

In the specific case of convergent dietary adaptations in myrmecophagous mammals, two gene families will be studied: taste receptor genes (*TASRs*; see part I.2) and chitinase genes (*CHIAs*; see part I.3), to respectively investigate how myrmecophagous mammals perceive and digest their prey. These studies will help us understand the evolution of these gene families within placentals and highlight molecular mechanisms underlying the adaptation to myrmecophagy, and whether similar mechanisms were involved between the different myrmecophagous species. To answer these questions, gene repertoires will be studied combining different approaches like comparative genomics, reconciliation methods, and comparative transcriptomics between several placental species including myrmecophagous ones and their closely related species. Comparing the number of pseudogenized and functional genes between myrmecophagous species and their non-myrmecophagous relatives will potentially reveal patterns of convergence in gene repertoires among ant- and termite-eating mammals with similar gene losses and/or gains. Differences in gene repertoires could also highlight different adaptive mechanisms involved in the adaptation to myrmecophagy with potential specific losses and/or gains. Adding transcriptomic data will further help understanding the role these genes might have played in the adaptation to myrmecophagy by shedding light on their potential functions in the different focal species.

I.2. Evolution of taste perception in myrmecophagous mammals

I.2.1. Introduction: taste perception in mammals

Taste perception

Taste is an important sense to consider when studying specific dietary adaptations as it allows an organism to identify food sources and avoid toxins, and is thus influenced by dietary changes (Luca *et al*, 2010). Five tastes are recognized: sour, salt, sweet, umami, and bitter (Chandrashekar *et al*, 2006; Yarmolinsky *et al*, 2009; Chaudhari and Roper, 2010; Roper and Chaudhari, 2017; Fig I.5 A). Sour taste is elicited by the presence of organic acids and is important to maintain the acid-base balance in an organism and to avoid spoiled food that is often acidic. Salt taste is activated by the presence of Na+ ions and helps in maintaining water balance. Sweet taste is triggered by carbohydrates and is important for energy intake. The presence of amino acids and nucleotides in protein-rich food activates the umami taste. Finally, bitter taste is important to detect toxins (Chandrashekar *et al*, 2006; Yarmolinsky *et al*, 2009; Chaudhari and Roper, 2010).

Taste cells enable the perception of tastes. They contain taste receptors and are classified in three types (type I, II, and III) (Chaudhari and Roper, 2010; Roper and Chaudhari, 2017). These taste cells are elongated cells clustered within taste buds (50-100 depending on the species). Taste buds are located in papillae, which are classified in three main types (circumvallate, foliate, and fungiform), distinguished by their form and location (Chandrashekar *et al*, 2006; Fig I.5 B). In humans, around 5 000 taste buds are present in the oral cavity on the tongue, palate, and epiglottis (Yarmolinsky *et al*, 2009; Chaudhari and Roper, 2010). Taste cells carry different taste receptors and are involved in the recognition of different tastes (Chaudhari and Roper, 2010; Roper and Chaudhari, 2017). Within taste buds, the most abundant taste cell type is the type I, which plays a role in synapses (Chaudhari and Roper, 2010; Roper and Chaudhari, 2017). About one third of a taste bud is composed of type II cells in which taste receptors for the umami, sweet, and bitter tastes are present; they are also called receptor cells (Chaudhari and Roper, 2010; Roper and Chaudhari, 2017; Fig I.5 B).

Finally, two to 20% of a taste bud is composed of type III taste cells, which also have a role in synapses and respond to a broad range of tastes as they receive signals from the receptor cells and might also be involved in sour taste perception; they are also called pre-synaptic cells (Chaudhari and Roper, 2010; Roper and Chaudhari, 2017). Taste perception originates in the apical position of the taste bud where taste cells form the taste pore in which interactions with tastants occurs (Chandrashekar *et al*, 2006; Fig I.5 B). The three types of cells are involved in the perception of taste and the transmission of the signal to the brain via sensory neurons connected to taste buds and elicited by one or several tastes (Chaudhari and Roper, 2010; Roper and Chaudhari, 2017).

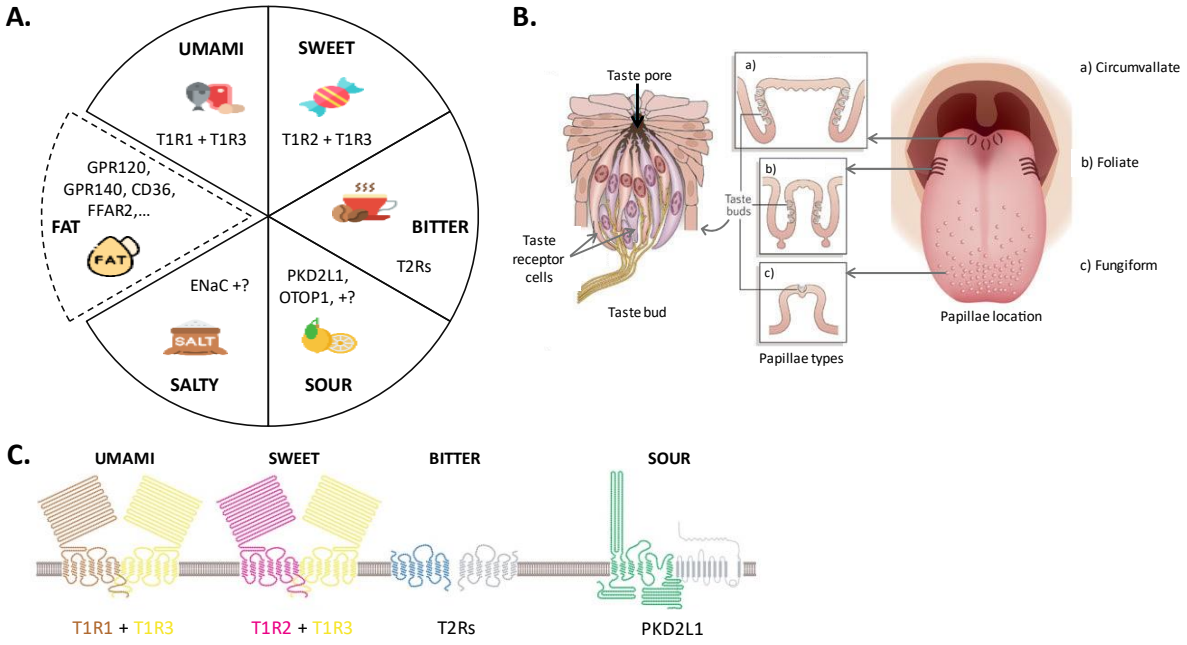


Figure I.5. Taste perception in mammals. A. The five recognized tastes with their associated gustatory receptors. Fat perception is also represented (dashed lines). B. Papillae and taste buds. C. Transmembrane taste receptors for sweet, umami, bitter and, sour perception (which are the focus of this thesis chapter). Adapted from Chandrashekar *et al* (2006) and Chaudhari and Roper (2010).

Proteins involved in taste perception

The five tastes are recognized by different taste receptors (Fig I.5 A). Salt and sour tastes are thought to be detected thanks to ion channels. Salt taste is recognized by an apical Na+ ion channel, ENaC (epithelial sodium channel), composed of three subunits (Yarmolinsky *et al*, 2009; Chaudhari and Roper, 2010). It has been demonstrated to be involved in the detection of salts in mice and potentially in humans (Shigemura *et al*, 2008; Chandrashekar *et al*, 2010; Oka *et al*, 2013; Bigiani, 2020). Other receptors or ion channels might be involved but are not yet fully described (Chaudhari and Roper, 2010).

Mechanisms of sour taste detection are still poorly known. It has been proposed that sour taste is detected thanks to a transmembrane receptor, PKD2L1 (polycystic kidney disease 2-like-1), which has been determined as a marker of type III taste cells (Chandrashekar *et al*, 2006; Huang *et al*, 2006; Ishimaru *et al*, 2006; Lopez Jimenez *et al*, 2006; Kataoka *et al*, 2008; Roper, 2014; Fig I.5). Ion channels are also potentially involved in sour taste perception (Chandrashekar *et al*, 2006). For instance, OTOP1 (Otopetrin 1) is a proton-selective ion channel recently recognized in sour taste perception as it is found expressed in type III taste cells and its knock-out in mice reduces the response to acids (Ramsey and DeSimone, 2018; Tu *et al*, 2018; Teng *et al*, 2019).

Mechanisms underlying sweet, umami, and bitter taste perception are better known and well studied. These tastes are perceived thanks to transmembrane taste receptors (TASRs) composed of seven helices and coupled with G-proteins. They are classified in two types: TAS1Rs (taste receptor type 1) are involved in the perception of sweet and umami tastes, and TAS2Rs (taste receptor type 2) of bitter taste (Yarmolinsky *et al*, 2009; Chaudhari and Roper, 2010; Roper and Chaudhari, 2017; Fig I.5). Genes coding for those receptors are divided in two paralogous classes: *Tas1rs* coding for sweet and umami taste receptors, and *Tas2rs* for bitter taste receptors (Hoon *et al*, 1999; Adler *et al*, 2000; Chandrashekar *et al*, 2000; Roper and Chaudhari, 2017). These two types of receptors will be the focus of this thesis chapter and are described below. Linked with the perception of these three tastes, TRPM5 (transient receptor potential cation channel subfamily M member 5), an ion channel found in umami, sweet, and bitter taste cells (type II taste cells, receptor cells), acts as a transduction element (Chaudhari and Roper, 2010; Roper and Chaudhari, 2017) and thus participates in taste detection.

Finally, fatty taste is not recognized as a taste as its detection might involve somatosensory perception (*i.e*., texture) (Chaudhari and Roper, 2010; Fig I.5 A). Yet, membrane receptors (*i.e*., CD36, GPR120, GRP40, GPR84, FFAR2, FFAR4) involved in the detection of fatty acids are present on taste bud cells and might be required for gustatory fat perception (Sclafani *et al*, 2007; Chaudhari and Roper, 2010; Galindo *et al*, 2012; Liu *et al*, 2018).

TAS1Rs for sweet and umami taste perception

Tas1Rs comprise three genes: *Tas1r1*, *Tas1r2*, and *Tas1r3* coding for TAS1R receptors. TAS1R1 receptor forms a heterodimer with TAS1R3 to make the umami receptor, which binds to amino acids, while TAS1R2 and TAS1R3 form a heterodimer to make the sweet receptor that binds to carbohydrates (Chandrashekar *et al*, 2006; Chaudhari and Roper, 2010; Bachmanov *et al*, 2014: Kinnamon, 2016; Fig I.5 C). *Tas1r1, Tas1r2,* and *Tas1r3* have respectively six, six, and one exon(s) in *Homo sapiens*.

The evolution of these three genes in placentals is characterized by gene losses specific to certain lineages losing one of the two tastes or both. Evolution of sweet and umami taste perception seems to be linked with dietary adaptations and notably toward specialized diets. Indeed, many losses have been reported within Carnivora, especially sweet perception in carnivorous species through the loss of *Tas1r2*, for instance, in Felidae (Li *et al*, 2005, 2006, 2009), and other carnivorous non-Felidae species (Jiang *et al*, 2012), consistent with the fact that meat contains more proteins and few carbohydrates. The transition toward an herbivorous bamboo-based diet likely influenced the loss of umami in panda probably due to relaxed selective pressures on *Tas1r1* whose inactivation matches evidence from the fossil record about this dietary shift (Zhao *et al*, 2010a). Marine mammals also lost taste perception abilities. For example, in pinnipeds *Tas1r1* has been lost and thus these species cannot detect umami taste (Sato and Wolsan, 2012) as well as in whales which lack umami and sweet perception (Feng *et al*, 2014; Kishida *et al*, 2015). It has been hypothesized that these losses could be influenced by the specific characteristics of the marine environment such as dietary changes including more fish which contains less umami tastants, the lack of mastication, and

the high concentration of salt in sea water that could mask umami taste (Sato and Wolsan, 2012; Feng *et al*, 2014).

In bats, losses of sweet and umami taste genes have also been reported (Zhao *et al*, 2010b; Zhao *et al*, 2012). For instance, a comparative analysis of *Tas1r2* and *Tas1r3* gene sequences in 34 frugivorous and insectivorous bat species showed that frugivorous species could detect sugars, as opposed to insectivorous ones (Jiao *et al*, 2021). Sweet perception has also been lost in blood-feeding vampire bats (Zhao *et al*, 2010b) that also lack *Tas1r3* required for umami perception (Zhao *et al*, 2012).

These studies all suggest potential links between the evolution of *Tas1r* gene repertoires in mammals and their diet. Yet, several incongruities have been reported between patterns of *Tas1rs* losses and species diet (Zhao and Zhang, 2012). For instance, contrary to pandas, other herbivorous species, like horses and cows, have intact *Tas1r1*, and some carnivorous species present functional *Tas1r2* genes suggesting other factors besides diet shape the evolution of *Tas1r* genes (Zhao and Zhang, 2012).

TAS2Rs for bitter taste perception

Tas2r genes are mono-exonic genes of a length of ~1 000 bp and include much more gene copies (~30 copies) than the *Tas1r* family (Chandrashekar *et al*, 2000, 2006; Chaudhari and Roper, 2010). The evolution of this gene family seems to follow a birth-and-death model of evolution with pseudogenizations and gene duplications occurring in specific clades and resulting in high variability in the number of functional gene copies among species (Dong *et al*, 2009; Bachmanov *et al*, 2014; Hayakawa *et al*, 2014; Li and Zhang, 2014; Liu *et al*, 2016).

In Euarchontoglires, between 16 and 40 gene copies have been identified among 28 species and 26 intact copies have been inferred in their last common ancestor from which numerous gene losses occurred, and gene duplications were mainly identified in anthropoids (Hayakawa *et al*, 2014). Between zero and 52 *Tas2r* gene copies have for instance been identified in a study conducted on 41 laurasiatherian placental species (Liu *et al*, 2016). Their evolution might be linked with the species ecological characteristics, notably their foraging habits with mammals having narrow diets and reduced mastication, like marine mammals or pangolins, having small *Tas2r* repertoires (Feng *et al*, 2014; Liu *et al*, 2016). Similarly to *Tas1r*

genes, many *Tas2r* losses occurred in some carnivorous species, which present fewer intact *Tas2rs* compared to non-carnivorous species (Hu and Shi, 2013). The shift toward eating bamboo in the red and giant pandas also influenced the evolution of their *Tas2r* repertoires with retention of more functional *Tas2rs* than other carnivores but with different *Tas2r* losses between the two species (Shan *et al*, 2018). In vampire bats, numerous gene losses were identified suggesting reduced bitter taste perception (Hong and Zhao, 2014) in addition to sweet and umami perception loss (Zhao *et al*, 2010b, 2012). *Tas2r* losses have also been identified in frugivorous bats and could be linked with their diet (Wang *et al*, 2020).

These studies therefore highlight potential links between the birth-and-death evolution of *Tas2r* genes in placentals and species ecological characteristics. Similar to *Tas1r* genes, evolution of *Tas2rs* is characterized by many gene losses likely due to relaxed selective pressures on genes not needed for bitter taste, which might be the result of dietary shifts.

Case study: evolution of taste receptors in myrmecophagous mammals

TAS1R, TAS2R, and PKD2L1 receptors will be studied to understand how myrmecophagous mammals perceive sweet, umami, bitter, and sour tastes that are crucial to find nutritive food and avoid poisonous food sources. Studying their evolution in convergent myrmecophagous species would allow identifying the underlying genomic adaptations of taste perception in these species and help to understand whether similar adaptations were involved by comparing gene repertoires between myrmecophagous and non-myrmecophagous species. Ant- and termite-eating mammals might rely less on taste than other mammals. Indeed, their tongue, comporting few taste buds and lacking certain types of papillae, is not adapted to taste perception, especially in anteaters and pangolins (*e.g.*, Doran and Allbrook, 1973; Casali *et al*, 2017). Ant-eating species have a very narrow diet and do not masticate their preys. These characteristics could be linked with losses of taste receptors in pangolins and marine mammals (Sato and Wolsan, 2012; Feng *et al*, 2014; Liu *et al*, 2016). In non-placental ant-eating mammals, losses of bitter taste receptors have been found in the short-beaked echidna (*Tachyglossus aculeatus*, Monotremata) where only three *Tas2rs* seem to be functional (Zhou *et al*, 2021), and in the numbat (*Myrmecobius fasciatus*, Marsupials) where sweet taste loss was also observed (Peel *et al*, 2022). This reduction of *Tasr* gene repertoires in these

specialized insectivorous species might be linked to their specialized diet (Zhou *et al*, 2021; Peel *et al*, 2022). Therefore, a reduction of taste receptor gene repertoires in myrmecophagous species compared to other placentals might be expected with potentially similar gene losses and therefore taste perception losses in these species. In the case of *Tas2r* genes, another non mutually exclusive hypothesis also resides in the fact that some copies might be retained in ant-eating mammals to enable them to detect toxins present in their food, as some termites and ants can produce toxins (Schmidt, 1986; Lopez and Morgan, 1997).

The following subparts of this chapter will focus on the evolution of the three *Tas1r* and *PKD2L1* genes (part I.2.2), and of the *Tas2r* gene family (part I.2.3).

I.2.2. Oral regression and evolution of sweet and umami taste perception in myrmecophagous mammals

Because of their feeding habits, myrmecophagous mammals have convergently evolved morphological adaptations (Reiss, 2001). Most species share an elongated skull with reduction or even complete loss of teeth (Reiss, 2001; Ferreira-Cardoso *et al*, 2019, 2022). Their hypertrophied salivary glands produce a lot of viscous saliva and their long and extensible tongue is used to catch and rapidly ingest prey without mastication (Patterson, 1975; Griffiths, 1968; Reiss, 2001). The absence of mastication is characterized by reduced jaw muscles (Ferreira-Cardoso *et al*, 2020) and is compensated by morphological adaptations of their usually muscular stomach (Griffiths, 1968; Nisa' *et al*, 2010; De Oliveira Firmino *et al*, 2020). Their tongue also presents specific characteristics with a complex musculature, a reduction of the number of taste buds and the absence of certain types of papillae (*e.g.*, Doran and Allbrook, 1973; Reiss, 2001; Casali *et al*, 2017; de Oliveira Firmino *et al*, 2020) making it useful to catch prey but not particularly suited for taste perception.

Myrmecophagous mammals are thus characterized by the regression of anatomical traits of their oral apparatus: teeth, jaw muscles, and taste buds. To better understand the evolution of these traits, studying the evolution of the associated genes that might be involved in such traits can help determine whether they have been lost in concert with each other, that

is whether genomic adaptations correlate with morphological adaptations and more specifically phenotypic losses, and reveal the underlying adaptive mechanisms involved. As myrmecophagous species all present morphological adaptations linked to this diet (Reiss, 2001) with some species, such as pangolins and anteaters, having striking adaptations occurring together (*i.e.*, teeth reduction, tongue modification with taste bud loss, masticatory muscle reduction; Reiss, 2001; Ferreira-Cardoso *et al*, 2019, 2020, 2022), we might expect similar patterns of gene losses in these species.

The following article focuses on specific genes whose evolution within myrmecophagous mammals might reflect the regression of several oral phenotypic traits linked to this diet. By comparing gene repertoires between myrmecophagous species and their non-myrmecophagous sister-species, this study aims at identifying potential gene losses that might be correlated with phenotypic trait losses and/or reductions. Regarding taste perception, four genes mentioned previously were investigated: *Tas1r1*, *Tas1r2*, and *Tas1r3* for sweet and umami perception, and *PKD2L1* for sour taste. Moreover, the evolution of 11 genes involved in teeth development and enamel formation have already been studied in Xenarthra and revealed that several parallel losses occurred within this group leading to enamel loss in sloths and armadillos and tooth loss in anteaters (Emerling *et al*, 2023). These genes notably include the *ODAM* and *ENAM* genes which have already been shown to be lost in enamelless and toothless mammals (Meredith *et al*, 2009; Springer *et al*, 2019). The following study aims at investigating patterns of gene losses in other myrmecophagous species. Among genes involved in dental and enamel development, *MEPE* and *DMP1* are also involved in bone formation and do not present any pattern of pseudogenization (Bardet *et* al, 2010; Silvent *et al*, 2013; Emerling *et al*, 2023) and were thus not included in the present study. A total of ten genes involved in enamel and dentin development and function were thus selected (*i.e.*, *AMELX*, *ENAM*, *AMBN*, *MMP20*, *KLK4*, *AMTN*, *ODAM*, *ACP4*, *DSPP*, *ODAPH*). Finally, regarding jaw muscle reduction, patterns of pseudogenization in *MYH16* (myosin heavy chain 16) were investigated. This gene encodes a myosin protein involved in muscle contraction and found in jaw muscles. In humans, *MYH16* was lost after the divergence with chimpanzees, which is consistent with reduced needs for mastication with the apparition of cooking as opposed to other primates that retained powerful masticatory abilities (Stedman

et al, 2004; Perry *et al*, 2005, 2015). These results suggest that the loss of this gene might be linked to dietary changes in humans and therefore, similar patterns might be observed in myrmecophagous species lacking mastication and having reduced jaw muscles.

The following study thus aims at identifying functional genes and pseudogenes in myrmecophagous species compared to other closely related placentals by extracting genes from genome assemblies, aligning them, and looking for inactivating mutations. Investigating patterns of gene losses in myrmecophagous species should highlight the underlying molecular mechanisms involved in the regression of oral traits in these species and reveal potential patterns of convergence in gene repertoires. More generally, it will further help us understanding the link between morphological and genomic adaptations in response to dietary shift and emphasizing on the importance to study both levels to fully understand the phenomenon of convergence.

Manuscript. Genomic signals of oral regression in myrmecophagous mammals

Personal contribution

I contributed, as a second author, as follow:

- Identification of *Tas1r*, *PKD2L1, and MYH16* genes in genome assemblies of myrmecophagous species generated as part of the ConvergeAnt project.
- Reading and editing.

This following manuscript is a first draft with preliminary results and discussion and is intended for submission after the PhD.

Emerling, C.A., **Teullet, S**., Allio, R., Delsuc, F., *in prep*. Genomic signals of oral regression in myrmecophagous mammals.

Genomic Signals of Oral Regression in Myrmecophagous Mammals

Christopher A. Emerling[1,2,*], Sophie Teullet[2], Rémi Allio[2,3], and Frédéric Delsuc[2,*]

[1]Biology Department, Reedley College, Reedley, CA, USA
[2]ISEM, Univ. Montpellier, CNRS, IRD, Montpellier, France
[3]CBGP, INRAE, CIRAD, IRD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France

***Correspondence to** christopher.emerling@reedleycollege.edu and frederic.delsuc@umontpellier.fr

ABSTRACT

Adaptation to ant and/or termite consumption (myrmecophagy) in mammals constitutes a textbook example of convergent evolution with at least five independent apparitions in placentals in the aardvark, anteaters, armadillos, pangolins, and aardwolves. Ant-eating species are characterized by striking convergent morphological adaptations such as skull elongation, teeth reduction or even complete loss, reduction of masticatory muscles, and long and protrusive tongues. To gain insights into the molecular adaptations underlying the regression of these traits, we investigated the functionality of the associated dental (*DSPP*, *ODAPH*), enamel (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, *MMP20*, *ODAM*, *KLK4*), taste receptor (*TAS1R1*, *TAS1R2*, *TAS1R3, PKD2L1*), and masticatory myosin (*MYH16*) genes in myrmecophagous species and their non-myrmecophagous closest relatives. Our results highlighted numerous taste receptor gene inactivations suggesting loss of sweet, umami, and/or sour taste perception in several myrmecophagous species, which is compatible with their narrow diet and tongues better suited for catching prey than for taste perception. Most myrmecophagous lineages also lost their masticatory myosin gene, consistent with their reduced mastication abilities. Our results confirmed major dental and enamel gene losses in Xenarthra with toothless anteaters losing almost all genes, and enamelless armadillos losing most enamel genes. This study additionally revealed losses of all dental and enamel genes in the toothless pangolins whereas more recently diverged species, which do not present marked morphological adaptations, such as aardwolves, have retained all functional genes. Finally, we highlight numerous shared inactivating mutations among myrmecophagous species suggesting ancient gene losses. Together these results help us to better understand the link between morphological and genomic adaptations in the context of the convergent regression of the oral apparatus linked to the specialization toward myrmecophagy.

INTRODUCTION

Convergent evolution is a widespread phenomenon in which distantly-related organisms independently develop highly similar phenotypic traits as a result of experiencing similar selection pressures. Deriving similar anatomy, physiology, and behavior can result from modifications to identical, distinct, or partially overlapping sets of genes, depending on the trait and its underlying molecular basis. The frequency with which convergent evolution is demonstrably repeatable at the genetic level, particularly among highly divergent taxa, is a topic of ongoing interest (Blount et al., 2018; Cerca, 2023).

One prominent example of convergent evolution can be found in myrmecophagous mammals, species that specialize in eating ants and/or termites. Mammalian taxa with some degree of myrmecophagy are phylogenetically widespread (Redford, 1987), with some of the most specialized myrmecophages being found among the xenarthran anteaters (Vermilingua), pangolins (Pholidota), the aardvark (Tubulidentata, *Orycteropus afer*), tolypeutine armadillos (Cingulata, Tolypeutinae: three-banded [*Tolypeutes* spp.], giant [*Priodontes maximus*], and naked-tailed [*Cabassous* spp.] armadillos), the aardwolves (Carnivora, Hyaenidae: *Proteles* spp.), the numbat (Dasyuromorphia, *Myrmecobius fasciatus*), and the short-beaked echidna (Monotremata, *Tachyglossus aculeatus*). Other mammals with a tendency towards myrmecophagy include other armadillos such as the fairy (Chlamyphorinae) and long-nosed armadillos (Dasypodidae), as well as some carnivorans, such as the bat-eared fox (Canidae*, Otocyon megalotis*), and sloth bear (Ursidae, *Melursus ursinus*).

Together these taxa present a suite of convergent traits tied to consuming copious amounts of social insects, albeit varying in degree, including a reduction and/or simplification of the dentition, thinning of the mandible, an elongated tongue, the production of a high volume of sticky saliva, and the reorganization of jaw musculature (Reiss, 2001; Ferreira-Cardoso et al., 2019, 2020). As these taxa have become adapted to consuming social insects, there has been a shift from a more typical mammalian masticatory apparatus, involving specialized teeth and robust jaw muscles and skeletal elements, towards maximizing the efficiency of consuming ants and termites with their modified tongues and saliva. These adaptations are particularly exemplified in xenarthran anteaters (“anteaters” throughout the remaining text) and pangolins, whose extreme specialization to myrmecophagy and accompanying anatomical modifications had led systematists to formerly conclude that they belong to a monophyletic group (Edentata) (Glass, 1985; Reiss, 2001). It was not until molecular phylogenetics was applied to this question towards the turn of the 21st century that it was firmly demonstrated that their supposedly unifying morphology was actually the result of convergent evolution (Murphy et al., 2001; Delsuc et al., 2002; Springer et al., 2019).

Given that the adaptation to myrmecophagy can lead to substantial modifications of the anatomy of the oral apparatus, it raises the question as to whether this convergent morphological evolution is detectable at the molecular level, as well. One way to test for convergent molecular evolution is tied to the phenomenon of regressive evolution, whereby phenotypic traits are reduced or lost over evolutionary

time (Fong et al., 1995; Lahti et al., 2009; Albalat and Cañestro, 2016). For instance, the narrowing of the mandible, the loss of any jaw musculature, and especially the reduction in dentition could all be interpreted as examples of such evolutionary regression or vestigialization. When such traits are tied to specific genes, and those genes are limited in their expression to specific traits (*i.e*., no or minimal pleiotropy), relaxation of selection for their functional maintenance can lead to the accumulation of inactivating mutations that render underlying genes non-functional (Albalat and Cañestro, 2016). The end result is unitary pseudogenes or deleted genes, with no paralogs to rescue their function. A number of examples of such gene loss linked to vestigialization have been documented, ranging from the loss of claw keratins in snakes and legless lizards (Dalla Valle et al., 2011; Emerling, 2017), the loss of insect-digesting chitinase and trehalase genes in mammals that have shifted from insectivory to carnivory and herbivory (Emerling et al., 2018; Janiak et al., 2018; Jiao et al., 2019) and the loss of melatonin synthesis and receptor pathway genes in mammals that have lost their pineal glands (Emerling et al., 2021; Valente et al., 2021; Yin et al., 2021).

We set out to test whether myrmecophagous mammals show similar signals of molecular regression in independent traits related to the oral apparatus, by examining genes related to dental development, tastant detection (*i.e*., gustation), and a gene expressed in jaw muscles. The development of vertebrate teeth relies on coordination between a suite of genes associated with the formation of dentin and overlying enamel. There has been extensive documentation of pseudogenization of these genes in a number of edentulous (toothless) and enamelless vertebrates, including turtles (Meredith et al., 2013), birds (Meredith et al., 2014), toads (Shaheen et al., 2021), and baleen whales (Randall et al., 2022) among others. As already shown for anteaters within xenarthrans (Emerling et al., 2023), dental simplification and tooth loss in myrmecophagous mammals is expected to lead to the accumulation of inactivating mutations in these genes, which should vary depending on the degree of dental reduction. Taste perception varies by taste modality, with distinct genes either being linked to detection of specific tastants or at least being located to specific gustatory cells (Chandrashekar et al., 2006; Yarmolinsky et al., 2009; Chaudhari and Roper, 2010; Roper and Chaudhari, 2017). Reductions in taste perception from the hypothesized vertebrate complement have been observed in certain species exhibiting dietary specializations, particularly hypercarnivorous mammals and the giant and red pandas (Zhao et al., 2010; Jiang et al., 2012), as well as in taxa that appear to have completely lost taste buds, such as cetaceans (Feng et al., 2014; Kishida et al., 2015) and snakes (Emerling, 2017). Given the extreme degree of dietary specialization of myrmecophagous mammals, changes in sweet and umami taste receptor genes (*TAS1Rs*) are predicted. Finally, a particular myosin protein (myosin heavy chain 16; MYH16) has been found to be expressed exclusively in the jaw muscles of multiple vertebrate lineages, and specifically appears to be associated with species with a strong bite force (Hoh, 2002; Lee et al., 2019). In myrmecophagous mammals, in which mastication has been greatly reduced or even lost (Ferreira-Cardoso et al., 2020), this gene would appear to be a good candidate for pseudogenization.

Using these sets of candidate genes, we addressed two main questions. First, do these sets of genes show evidence of pseudogenization in all or most myrmecophagous mammals? Second, despite being expressed in distinctly different tissues (teeth, taste buds, and muscles), is the pattern and therefore the timing of gene loss in these three categories of genes similar? If so, this may be pointing to a concomitant series of regressive events associated with the convergent evolution of myrmecophagy. We tested these hypotheses by interrogating genomic data in placental myrmecophages and closely related species, including three anteaters, six armadillos, five pangolins, the aardvark, two aardwolves, the bat-eared fox, and the sloth bear.

MATERIALS AND METHODS

Candidate genes

The genes selected for this study were based on studies that have demonstrated a strong link between gene functions and phenotypes, which similarly have been shown to become pseudogenes in at least some instances. For example, numerous genes are known to participate in tooth development, but many, presumably due to pleiotropic functions, do not degrade into pseudogenes. Accordingly, the tooth genes we examined can be categorized based on their expression patterns and functions inferred from association with human congenital diseases, mouse knockout studies, and interspecies comparisons (Meredith et al., 2013, 2014; Smith et al., 2017; Randall et al., 2022; Emerling et al., 2023): 1) Enamel-development: *AMELX* (amelogenin), *ENAM* (enamelin), *AMBN* (ameloblastin), *MMP20* (enamelysin), *KLK4* (kallikrein-related peptidase-4) (Lagerström et al., 1991; Rajpar et al., 2001; Kim et al., 2005; Meredith et al., 2009; Poulter et al., 2014; Seymen et al., 2015); 2) Expression during both enamel development and throughout adulthood at the gingiva-tooth junction: *AMTN* (amelotin), *ODAM* (odontogenic ameloblast-associated) (Nishio et al., 2010; Nakayama et al., 2015; Wazen et al., 2015; Smith et al., 2016; Springer et al., 2019); 3) Unknown enamel function: *ACP4* (acid phosphatase 4) (Seymen et al., 2016); 4) Dentin development: *DSPP* (dentin sialophosphoprotein) (Xiao et al., 2001); and 5) Tooth retention: *ODAPH* (odontogenesis-associated phosphoprotein) (Parry et al., 2012; Springer et al., 2016).

While multiple genes have been tied to gustation, we examined four genes with clear orthologs that are pseudogenized in at least in some vertebrates (Zhao et al., 2010; Zhao et al., 2012; Jiang et al., 2012; Feng et al., 2014; Emerling, 2017) and known to be tied to specific taste modalities: *TAS1R1* (taste receptor type 1 member 1), *TAS1R2* (taste receptor type 1 member 2), *TAS1R3* (taste receptor type 1 member 3), and *PKD2L1* (polycystic kidney disease 2-like 1). The protein TAS1R1 pairs with TAS1R3 to make the umami taste receptor, responding to certain amino acids and leading to the perception of savory or umami flavors in humans. TAS1R2 similarly forms a heterodimer with TAS1R3, which instead detects sweet tastants, especially monosaccharides and disaccharides (Yarmolinsky et al., 2009; Roper and Chaudhari, 2017). While *PKD2L1*'s function is unclear, it is expressed uniquely in type III cells

within taste buds, which specialize in transducing sour tastants (Tu et al., 2018; Liman and Kinnamon, 2021).

Finally, we examined the jaw myosin gene, *MYH16* (myosin heavy chain 16). While vertebrate jaw muscles can express several types of myosin proteins, MYH16, also known as the masticatory myosin, appears to be uniquely expressed in jaw muscles. While the gene itself has not been examined in much detail, outside of being a pseudogene in humans (Stedman et al., 2004), the protein does appear to be lost in a number of mammal species (Hoh, 2002; Lee et al., 2019).

Genomic dataset

We used a mixture of publicly available genomes and *de novo* generated data to obtain gene sequences in our focal species. The following data were pulled from whole genome assemblies on NCBI's RefSeq and/or WGS (whole genome shotgun) databases: Linnaeus's two-fingered sloth, *Choloepus didactylus* (GCF_015220235.1, 57x), Hoffmann's two-toed sloth, *Choloepus hoffmanni* (GCA_000164785.2, 65x); Brown-throated three-fingered sloth, *Bradypus variegatus* (GCA_004027775.1, 58x); Southern naked-tailed armadillo, *Cabassous unicinctus* (GCA_029593785.1, 60x); Southern three-banded armadillo, *Tolypeutes matacus* (GCA_026826555.1, 70x); nine-banded armadillo, *Dasypus novemcinctus* (GCF_000208655.3, 6x); Sunda pangolin, *Manis javanica* (GCF_014570535.1, 412x); Chinese pangolin, *Manis pentadactyla* (GCF_030020395.1, 30x); Indian pangolin, *Manis crassicaudata* (GCA_016801295.1, 44x)*;* white-bellied tree pangolin, *Phataginus tricuspis* (GCA_029783875.1, 100x); aardvark, *Orycteropus afer* (GCF_000298275.1, 44x); Southern aardwolf, *Proteles cristatus* (GCA_017311185.1, 100x); and bat-eared fox, *Otocyon megalotis* (GCA_017311455.1, 100x). Gene sequences for the following species were extracted from unpublished whole genome assemblies generated by the ConvergeAnt project: the pale-throated three-fingered sloth, *Bradypus tridactylus* (108x); giant anteater, *Myrmecophaga tridactyla* (119x); Southern tamandua, *Tamandua tetradactyla* (118x); silky anteater, *Cyclopes didactylus* (91x); giant armadillo, *Priodontes maximus* (71x); pink fairy armadillo, *Chlamyphorus truncatus* (92x); six-banded armadillo, *Euphractus sexcinctus* (71x); and giant pangolin, *Smutsia gigantea* (80x). Finally, we used short reads sequences downloaded from NCBI's SRA (Sequence Read Archive) database for the Eastern aardwolf (*Proteles septentrionalis*; SRX9615643; Allio et al., 2021) and the sloth bear (*Melursus ursinus*; ERX1025771; Kumar et al., 2017).

In addition to our myrmecophagous focal taxa, we obtained gene sequences from a variable number of outgroup species. We emphasized outgroup taxa that were more closely related to our focal species to determine if mutations were shared or autapomorphic, but included others for better mammalian taxonomic representation as needed. Gene sequences for nearly all outgroup species for comparison were derived from NCBI's RefSeq and WGS databases.

Gene assembly

For gene assembly, we first obtained reference mRNA sequences from NCBI (Genbank), typically for *Homo sapiens*. When this was not possible, we used alternative curated mammal mRNA sequences or gene models derived from the NCBI eukaryotic genome annotation pipeline (Thibaud-Nissen et al., 2016). We imported the full mRNA sequence into Geneious Prime (v2019.2.3; Kearse et al., 2012) and used the annotations provided by GenBank to identify exon/intron boundaries and coding sequence structure. Since we did not evaluate non-coding DNA in this study, we did not analyze the structure of untranslated regions. As such, exon numbering throughout this study is based on coding exons only.

Once we obtained the coding regions of the reference mRNA sequences, we used these to obtain sequences derived from whole genome assemblies. We began by obtaining sequences from NCBI's RefSeq and WGS databases by performing similarity search with BLAST (discontiguous megablast) using the mRNA reference sequence against target assemblies. After obtaining the first few hits, we downloaded the contig or scaffold regions encompassing the entire coding sequence (CDS) and imported them into Geneious. Then, we computed automated sequence alignments using MUSCLE (ver 3.8.425; Edgar, 2004) within Geneious Prime. In cases where sequences had large amounts of unknown bases (*e.g.*, more than 50 Ns), we deleted all except for about 10 Ns so as to facilitate better sequence alignments. As we obtained whole gene sequences from some representative species, we then used those sequences for subsequent BLAST searches in related taxa. Once we had obtained all relevant taxa from NCBI, we used these whole gene sequences to obtain genes from local databases.

We supplemented sequences on NCBI with some derived from novel hybrid whole genome assemblies produced by the ConvergeAnt project: Southern tamandua (*Tamandua tetradactyla*), giant anteater (*Myrmecophaga tridactyla*), silky anteater (*Cyclopes didactylus*), pale-throated sloth (*Bradypus tridactylus*), giant armadillo (*Priodontes maximus*), six-banded armadillo (*Euphractus sexcinctus*), pink fairy armadillo (*Chlamyphorus truncatus*), and giant pangolin (*Smutsia gigantea*; Heighton et al., 2023). Hybrid genome assemblies were produced using MaSuRCA v3.2.9 (Zimin et al., 2017) by combining Nanopore long reads with short Illumina reads as detailed in Allio et al. (2021) for the southern aardwolf (*Proteles cristatus*) and the bat-eared fox (*Otocyon megalotis*). *De novo* genome annotation was done using the MAKER v3 pipeline (Holt and Yandell, 2011) following the strategy designed by the DNA Zoo (www.dnazoo.org) to combine different annotation analyses: i) two rounds of *ab initio* gene prediction with SNAP (Korf, 2004) and Augustus (Stanke et al., 2006), ii) use of transcriptomic information from publicly available and newly generated transcriptomes, and iii) protein sequences from the Uniprot/SWISSPROT database (Bairoch and Apweiler, 2000). More details on the hybrid assembly and genome annotation steps can be found in Allio (2021). The Southern tamandua (*T. tetradactyla*), giant anteater (*M. tridactyla*), and bat-eared fox (*O. megalotis*) hybrid genome assemblies were later upgraded to chromosome-length by the DNA Zoo using complementary Hi-C data (Dudchenko et al., 2017, 2018) and re-annotated using the same strategy. Similarly,

we used the DNA Zoo annotations of the Hi-C genome assembly for the white-bellied tree pangolin (*Phataginus tricuspis*). The MAKER-predicted coding sequences of *TAS1Rs*, *PKD2L1*, and *MYH16* for these 11 species were then mapped against their corresponding genome assemblies using Minimap2 v2.24 (Li, 2018) using the long-read spliced alignment mode (with default parameters) as implemented in Geneious Prime v2022.1.1. This splice-aware alignment enables the mapping of CDS, cDNA, or mRNA sequences against reference genomic sequences containing introns by splitting the query sequence to its different mapping locations. The corresponding genomic regions including mapping exons and intercalating introns were extracted for downstream analyses.

In some instances, we used short read data derived from whole genome sequencing to assemble genes. For the sloth bear (*Melursus ursinus*), we downloaded whole genome sequencing reads from SRA and imported them into Geneious. We used reference gene sequences from other ursid species and mapped the paired-end short reads using the Geneious Prime mapper, with settings at Medium-Low Sensitivity/Fast with no iterations. Each mapping result was then examined by eye, and in cases of abundant erroneous mappings (*i.e*., nonhomologous sequences) we remapped those same reads at Low Sensitivity/Fastest and iterated up to five times. Again, we examined each mapping alignment by eye and manually adjusted as needed.

For the Eastern aardwolf (*Proteles septentrionalis*), paired-end short reads from SRA were mapped to the Southern aardwolf (*P. cristatus*) orthologs using bowtie2 v2.3.4.3 (Langmead and Salzberg, 2012), implementing default parameters. Mapped reads were extracted from the bam mapping files and converted to fastq files with SAMtools v1.10 (Danecek et al., 2021). These reads were imported into Geneious Prime and remapped against reference gene sequences for visualization.

Sequence alignments and gene functionality analyses

Once we had obtained all orthologs for a single gene, we performed successive DNA sequence alignments with MUSCLE in Geneious Prime, starting by aligning two closely related taxa, then aligning these two to a third, those three to a fourth, and so on. By doing this, we were able to anchor the alignment using highly similar sequences and then progressively add more divergent ones, thereby minimizing alignment errors. In addition, this allowed us to examine each alignment by eye to search for errors, a common issue when aligning divergent intronic sequences.

After these global alignments were complete, we examined the gene structure for every ortholog to ensure gene completeness. When data appeared to be missing, based on exon predictions from the reference mRNA, we attempted additional BLAST searches with more relaxed parameters and/or alternative reference sequences. In some cases, there was evidence of partial or whole gene deletion, which required us to obtain additional sequence data upstream or downstream from our original BLAST searches.

Once our gene alignments were complete, we searched for evidence of the following categories of inactivating mutations: start codon mutations, frameshift insertions and deletions, splice site mutations, alterations to the ancestral stop codon, exon-intron boundary deletions, whole exon deletions, and premature stop codons. To search for premature stop codons and mutations in the ancestral start and stop codons, we generated alignments encompassing the CDSs only. To do so we removed all introns, removed frameshift insertions, and inserted Ns to restore the correct reading frame in frameshift deletions.

In some instances, we needed to validate the presence of a mutation, which we did using data from at least one of three different sources. In some cases, alternative genome assemblies, sometimes from separate individuals, were available either in NCBI's WGS database or on local assemblies. For some NCBI-derived sequences, we searched using BLAST (megablast) the relevant regions against available experiments in SRA and mapped them in Geneious Prime. Finally, for others (*Cyclopes didactylus*, *Proteles cristatus*), we mapped short reads using *de novo* sequence data on the gene sequence with bowtie2 as described above for *P. septentrionalis*. Two individuals were used for *P. cristatus* (NMB12641 and NMB12667) and one for *C. didactylus* (M2300).

In addition to searching for inactivating mutations, we examined whether any were shared among two or more species. Such shared mutations are suggestive of loss of gene function in the lineage leading to a common ancestor. The only putative shared mutations were found in xenarthrans and pangolins, so we referred to the phylogenies of Gibb et al. (2016) for the former and Heighton et al. (2023) for the latter.

RESULTS

Overall patterns

Each of the examined genes was pseudogenized in at least some of our focal species (Figure 1). *TAS1R3* was inactivated the least frequently (2/18, 11.1%) with *ACP4* showing evidence of pseudogenization in the vast majority of the focal taxa (15/18, 83.3%). The most anatomically extreme myrmecophagous species, pangolins and anteaters, have the highest proportion of pseudogenes, with 86.7 to100% of the genes inactivated in pangolins, and 78.6 to 85.7% inactivated in anteaters. The white-bellied tree pangolin (*Phataginus tricuspis*) was the sole species to present evidence of inactivation in every examined gene. By contrast, myrmecophagous carnivorans, which present far less anatomical regression in the oral apparatus, ranged from 0 to 6.7% gene inactivation. Even still, when disabling mutations were present in these species, they were always heterozygous (functional/nonfunctional), suggesting that these mutations have not been fixed in their respective species. Some of our non-myrmecophagous outgroup species also possessed pseudogenes for some of the genes we studied (*e.g.*, sloths [Folivora]), and these are due in part to a combination of shared history of gene loss with myrmecophagous species and in other cases may represent adaptations tied to other dietary niches (see below).

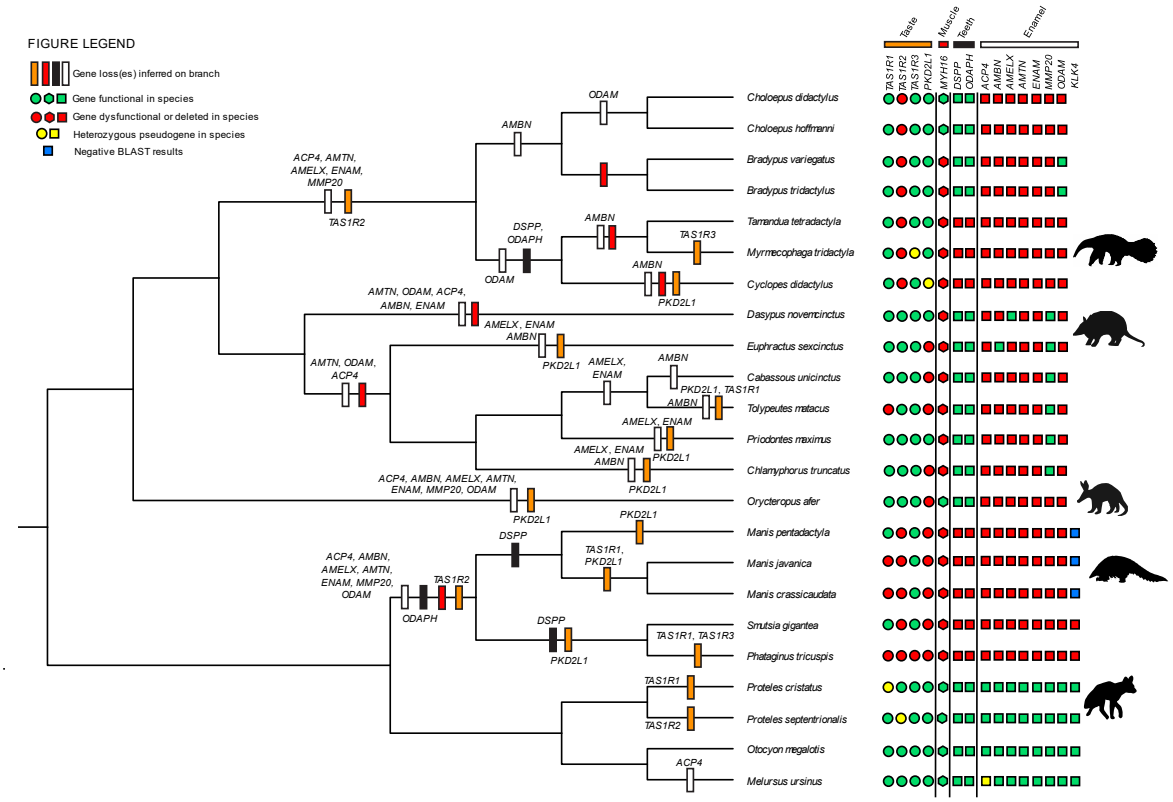


Figure 1: Phylogenetic patterns of gene loss linked to oral regression in myrmecophagous placental mammals and closely related species. Vertical bars on the phylogeny indicate taste receptor, myosin, teeth, and enamel gene losses respectively in orange, red, black, and white. Gene functionality for each species is indicated on the right of the graph with functional genes in green and pseudogenes in red. Genes for which polymorphism was detected within a species are indicated in yellow. Negative BLAST results are indicated in blue. Silhouettes of myrmecophagous species were downloaded from phylopic.org.

Taste genes

The taste genes we examined were quite variable in their patterns of functional loss, being inactivated almost never (*TAS1R1, TAS1R3*) to frequently (*TAS1R2*, *PKD2L1*). The umami receptor gene, *TAS1R1*, had clear positive evidence of inactivation in only the Southern three-banded armadillo (*Tolypeutes matacus*) and the white-bellied tree pangolin (*Phataginus tricuspis*). The former has eight inactivating mutations across the gene, validated by two different genome assemblies, and the latter has a single 1-bp insertion in exon 1, also supported by two different assemblies (Figure 2). *TAS1R1* in the Sunda (*Manis javanica*) and Indian (*M. crassicaudata*) pangolins may likewise be a pseudogene, based on an 8-bp insertion exon 6. However, this is near the ancestral stop codon and results in an additional five residues before the next stop codon.The Southern aardwolf (*Proteles cristatus*) assembly has an 8-bp deletion in exon 1, which is supported by short read data. However, one individual (NMB12641) is homozygous for this mutation, whereas another individual (NMB12667) is heterozygous. Its sister-species, the Eastern aardwolf (*P. septentrionalis*), does not possess this mutation.

The sweet receptor gene, *TAS1R2*, is much more commonly pseudogenized in our focal taxa. Among xenarthrans, it appears functionally intact in all six armadillos we examined, whereas it is clearly a pseudogene in all three anteaters. Not only do anteaters have eight inactivating mutations shared among them, their sister-group, sloths, also have a pseudogenized *TAS1R2* ortholog. Notably, the two clades of two-fingered (*Choloepus* spp.) and three-fingered sloths (*Bradypus* spp.) share a 19-bp deletion of the exon 4-intron 4 boundary and an 8-bp deletion in exon 5 with anteaters (Figure 2). *TAS1R2* is also inactivated in all five pangolins, with 10 disabling mutations shared between all these species (*e.g*., Figure 2). Finally, the Eastern aardwolf has a 1-bp insertion in exon 3, though the single individual examined is heterozygous for this mutation, which is not found in the Southern aardwolf.

TAS1R3, whose protein makes heterodimers with *TAS1R1* and *TAS1R2*, only rarely showed evidence of inactivation. The giant anteater (*Myrmecophaga tridactyla*) presents a premature stop codon in exon 3 of one genome assembly, but this was not reproduced in a second. Furthermore, the short read data suggests that this may be polymorphic, as this individual is heterozygous for the stop codon. By contrast, the white-bellied tree pangolin (*Phataginus tricuspis*) presents nine inactivating mutations, nearly all of which are validated by two different genome assemblies.

Finally, the sour taste gene, *PKD2L1*, shows a variable pattern of inactivation. The silky anteater (*Cyclopes didactylus*) has a 2-bp deletion in exon 13, which appears polymorphic based on short read data mapping. Multiple armadillos present inactivating mutations, including *Tolypeutes matacus* (three, two supported by two different assemblies), *Cabassous unicinctus* (Southern naked-tailed armadillo; two), *Chlamyphorus truncatus* (pink fairy armadillo; one), and *Euphractus sexcinctus* (six-banded armadillo; one). All pangolins have inactivating mutations in *PKD2L1* (*e.g*., Figure 2), and although none of these are unambiguously shared among all five species, four mutations are shared between the African pangolins (*Smutsia gigantea* and *Phataginus tricuspis*) and a putative splice acceptor mutation in intron 8 (AG → GG/CG) is shared by the Asian pangolins (*Manis* spp.). Finally, two inactivating mutations were found in the aardvark (*Orycteropus afer*), which were both validated by a second genome assembly.

Masticatory myosin

MYH16 is pseudogenized quite frequently among the focal taxa. Among xenarthrans, it is quite commonly inactivated, being a pseudogene in all three anteaters and all six armadillos examined. Among the anteaters, we found three shared disabling mutations between the Southern tamandua (*Tamandua tetradactyla*) and the giant anteater (*Myrmecophaga tridactyla*) (Figure 2). While the silky anteater (*Cyclopes didactylus*) also has a nonfunctional ortholog for *MYH16* (10 mutations), there are no mutations shared between this species and the other anteaters. Among the armadillos, we found seven mutations shared among all chlamyphorid armadillos (*e.g*., Figure 2). While within dasypodids, the nine-banded armadillo (*Dasypus novemcinctus*) also has an *MYH16* pseudogene, it is not clear if any of its mutations are shared with the five

chlamyphorids. Among the sloth outgroups, the two-fingered sloths (*Choloepus* spp.) have an intact *MYH16*, while it is a pseudogene in both three-fingered species (*Bradypus* spp.). Among the pangolins, all five species have a pseudogenized *MYH16*, sharing 12 inactivating mutations among them (*e.g.*, Figure 2), including a large deletion encompassing exons 19–34. By contrast, this gene appears functionally intact in *Orycteropus afer* and the myrmecophagous carnivorans.

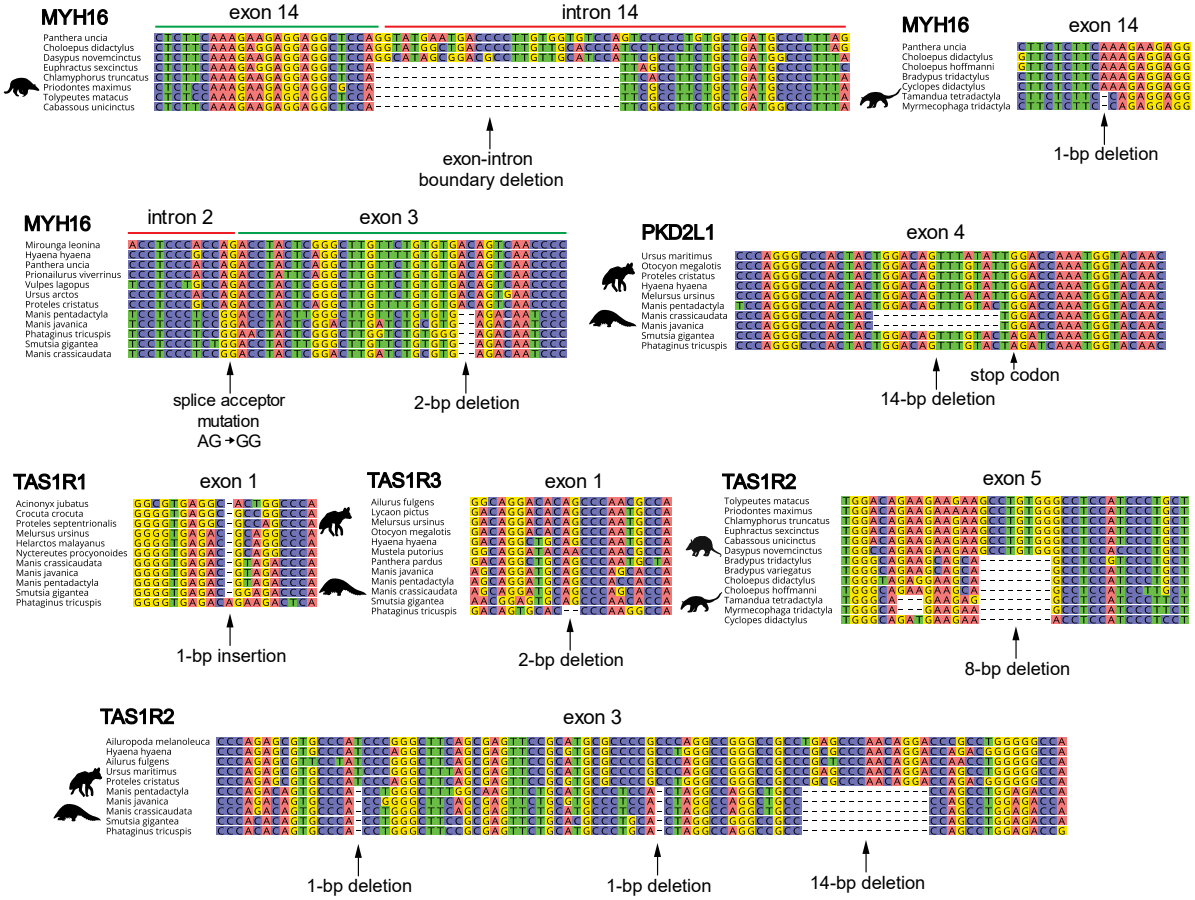


Figure 2: Examples of shared inactivating mutations of taste receptor genes and *MYH16* in myrmecophagous species.

Tooth genes

The tooth genes represent the largest set of genes that we examined. The xenarthran data have already been reported by Emerling et al. (2023). In short, all nine examined genes are pseudogenes in anteaters, with their sister group, the folivorous and enamelless sloths, possessing pseudogenic orthologs for nearly all the genes except *ODAPH* and *DSPP* in both genera, and *ODAM* in three-fingered sloths (*Bradypus* spp.). Notably, five genes (*ACP4*, *AMELX*, *AMTN*, *ENAM*, *MMP20*) present shared inactivating mutations between anteaters and sloths. The enamelless chlamyphorid armadillos also have inactivated tooth genes, with the exceptions of *MMP20*, *ODAPH*,

and *DSPP*, with the six-banded armadillo (*Euphractus sexcinctus*) also having an intact *AMBN*. The nine-banded armadillo (*Dasypus novemcinctus*), in which juveniles possess a thin layer of enamel, has pseudogenes for *ACP4*, *AMBN*, *AMTN*, *ENAM*, and *ODAM*. Note that *KLK4* being specific to boreoeutherian mammals (Laurasiatheria + Euarchontoglires), no orthologs exist for xenarthrans and afrotherians.

Tooth genes have been described in some detail for three species of pangolins (Meredith et al., 2009, 2014; Choo et al., 2016; Springer et al., 2016, 2019), but here we expand upon these results by characterizing nine genes with three Asian pangolins and two African pangolins, representing the two major subclades within Pholidota. All nine genes, plus the Boreoeutheria-specific *KLK4*, are pseudogenes or show evidence of whole gene deletion in all five pangolin species investigated. In almost every case, there is evidence of inactivating mutations shared between Asian and African pangolins (Figure 3). For the enamel-development genes, *AMELX* has a single shared splice acceptor mutation, *ENAM* has four shared mutations, including a deletion of exons 1–4, *AMBN* has three shared mutations, as does *MMP20* (Figure 3). *KLK4* returned no BLAST results for all three *Manis* spp., and the flanking regions for this gene appeared to be on separate contigs or scaffolds in these assemblies. For the African pangolins, we were only able to recover the first two exons of *KLK4*, with exon 2 presenting two shared premature stop codons. For the enamel and gingiva-tooth junction genes, *AMTN* exons 1–4 have been deleted in all pangolins, and *ODAM* has eight shared mutations (Figure 3). *ACP4*, whose function in enamel development is unclear, has 11 shared mutations. The dentin development gene *DSPP* has been completely deleted in Asian pangolins, and African pangolins have 10 shared mutations, including a massive deletion in the large repetitive region in exon 4. Finally, *ODAPH*, whose function is tied to tooth retention, has six mutations shared between all five species (*e.g.*, Figure 3).

The enamelless aardvark's pseudogene distribution has been described previously (Meredith et al., 2014; Springer et al., 2016, 2019), which we confirm here, having inactivated orthologs for all of the enamel associated genes while retaining functional *DSPP* and *ODAPH*. By contrast, the enameled myrmecophagous carnivorans almost exclusively have intact dental genes. The single genetic evidence of dental regression was found in the sloth bear (*Melursus ursinus*) whose *ACP4* is heterozygous for a splice donor mutation (GT → GG) in intron 3. Besides this, both *Proteles* spp. have a splice donor mutation in intron 8 of *ODAM* (GT → TT), but this is shared with the striped (*Hyaena hyaena*) and spotted (*Crocuta crocuta*) hyaenas. Otherwise, no other evidence of inactivating mutations is found in these hyaenids, suggesting there may be an alternative splice donor and/or the exon structure is distinct for this family, with one option presenting 9-bp upstream into exon 8.

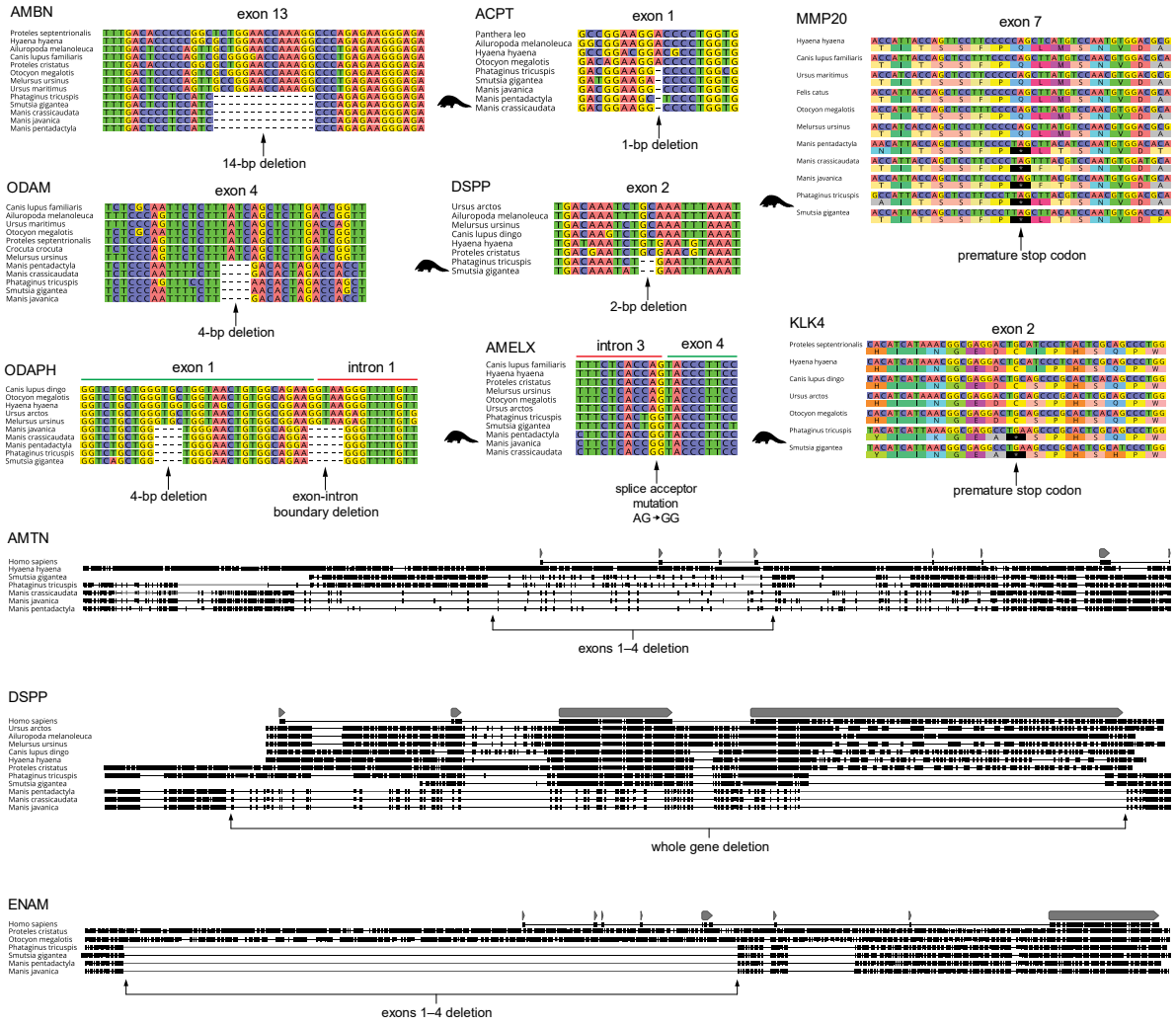


Figure 3: Examples of shared inactivating mutations in dental genes of pangolins and aardwolves.

DISCUSSION

Mammalian adaptations for consuming social insects include the regression of some anatomical elements in the oral apparatus, and here we provide evidence that this is partly reflected in the loss of function of protein-coding genes tied to taste reception, mastication, and dentition. In addition to the patterns of gene loss, the phylogenetic distribution of shared inactivating mutations helps to provide a sense of timing in these regressive evolutionary events. Below, we discuss the implications of each category of gene loss and follow up with an evaluation of the relative timing of these gene inactivations in the broader context of the convergent evolution of myrmecophagy in placentals.

Myrmecophagy is associated with some degree of taste loss

One of the more remarkable convergent adaptations of myrmecophagous mammals involves the elongation of the tongue, paired with rapid protrusive movements and the

production of sticky saliva, which together allow for the rapid consumption of large quantities of ants and termites. At its most extreme, including within pangolins and anteaters, there has been a remodeling of the masticatory muscles (Ferreira-Cardoso et al., 2020) and modifications of the mandible (Ferreira-Cardoso et al., 2019). With the modification of this organ into a prey-capturing structure, there seems to be some degree of reduction in the density of taste buds and/or the papillae with which they are associated (Kubota et al., 1962a, b; Doran and Allbrook, 1973; Abayomi et al., 2009; Casali et al., 2017). This may be the result of a modification of the tongue from an aid to the mechanical digestion associated with mastication, a reduced need for gustation given the simplification of the diet and rapid consumption of prey, or perhaps both.

While this anatomical reorganization of the tongue may predict a parallel reduction in the genes underlying gustatory pathways, we found that the result was rather mixed. At one extreme, some myrmecophagous species retained functional orthologs for the four genes of interest, suggesting the preservation of the sweet, umami, and sour taste modalities. These included species with a presumably more recent and weaker commitment to this dietary habit (sloth bear, bat-eared fox) but also a species with a seemingly long history of myrmecophagy (giant armadillo). At the opposite extreme stands the white-bellied tree pangolin, which appears to have lost all three gustatory pathways, suggesting a reduction in taste perhaps only matched by cetaceans (whales, dolphins) among placental mammals. The remainder of the myrmecophagous taxa are along a spectrum, including species that appear to lack sweet taste only (*TAS1R2*; anteaters), sour only (*PKD2L1*; aardvark, pink fairy and Southern naked-tailed armadillos), umami and sour (*TAS1R1*, *PKD2L1*; Southern three-banded armadillo), sweet and sour (*TAS1R2*, *PKD2L1*; Chinese and giant pangolins), and possibly umami, sweet and sour (*TAS1R1*, *TAS1R2*, *PKD2L1*; Indian and Sunda pangolins). Furthermore, based on evidence of nonfunctional alleles, some species appear to potentially be in the process of losing their first (Southern aardwolf, *TAS1R1*; Eastern aardwolf, *TAS1R2*), or second taste genes (giant anteater, *TAS1R3*; silky anteater, *PKD2L1*). The loss of *TAS1R3* in the giant anteater may further indicate incipient degradation of the umami taste pathway.

Based on the distribution of unique inactivating mutations, as well as species possessing pseudogenic alleles, it appears that some of these events of taste loss may be very recent. However, other species share mutations, suggesting pseudogenization events tracing back ancestral lineages. One of the oldest among these is the loss of *TAS1R2* in the common ancestor of all pangolin species, with 10 inactivating mutations, suggesting pseudogenization well before the origin of crown Pholidota dated around 41 million years ago (Mya; Heighton et al., 2023). Another very early loss of *TAS1R2* can be more confidently dated, with anteaters (Vermilingua) sharing three inactivating mutations with sloths (Folivora). The node for the most recent common ancestor of their clade (Pilosa) dates to around 55 Mya (Gibb et al., 2016). Notably, this occurred during a relatively narrow time frame of about nine million years (Myr) alongside the probable loss of enamel in this lineage (Emerling et al., 2023). We will further discuss the potential implications of this below.

The loss of taste genes has been relatively well described in placental mammals, with the most striking examples of gene loss being found in cetaceans (Feng et al., 2014), which appear to lack functional orthologs for sweet, umami, and sour taste genes. The semi-aquatic pinnipeds also lack sweet and umami receptor genes (Sato and Wolsan, 2012; Wolsan and Sato, 2020), perhaps suggesting that secondary marine adaptations somehow impact taste perception, a pattern also found in otters (Lutrinae; Wolsan and Sato, 2022) and penguins (Sphenisciformes; Zhao et al., 2015; Cole et al., 2022). However, some of these taste losses may be related to dietary shifts, given that the sweet receptor gene was lost in some hypercarnivorous species (Jiang et al., 2012) and the umami receptor gene was convergently lost in the giant and red pandas (Zhao et al., 2010; Hu et al., 2017). Curiously, the umami receptor has been lost widely in bats (Chiroptera; Zhao et al., 2012), though consistent with predictions from specialized diets, the sanguivorous vampire bats have additionally lost the sweet receptor gene (Zhao et al., 2012).

A narrowing of diet may largely explain the reduction of taste observed in most myrmecophages, but there are other possible explanations. First, it may be related to the modification of the tongue to a prey-capturing organ rather than being used to sample tastants. An analogous situation has been described in a few squamate lineages, including snakes, varanids, and teiids, who have modified tongues adapted to sensation via the vomeronasal organ (Schwenk, 1985; Young, 1997). Most, or even perhaps all, snakes have lost their sweet and umami receptor genes (Emerling et. al., 2017), though this may be in part due to an early history of carnivory (Emerling, 2022). A third hypothesis is that food consumption in some myrmecophages is so rapid that it minimizes the need for gustation. This hypothesis has also been put forth to explain taste reduction in cetaceans and penguins (Feng et al., 2014; Cole et al., 2022). All three hypotheses, diet specialization, tongue modification for prey capture, and rapid consumption of prey, are certainly not mutually exclusive to explain the pattern in myrmecophagous mammals and indeed may be reinforcing each other. Yet, the retention of some taste modalities in even some of the most specialized myrmecophagous species (*e.g*., anteaters, most pangolins) point to the possibility that complete taste gene loss may be rarely adaptive. The fact that some of these genes have other functional roles, such as carbohydrate metabolism (Roper and Chaudhari, 2017), may signify that they are unlikely to be lost with regularity.

Myrmecophagy is associated with loss of masticatory myosin

Myosin proteins, along with actin, tropomyosin and troponin, drive muscle contraction in mammals. Myosins are heterohexamer proteins, whose genes belong to two families (MYHs [myosin heavy chains] and MYLs [myosin light chains]) that are expressed in various muscles, with most being expressed in more than one muscle type (Hoh, 2002). However, the masticatory myosin, encoded by *MYH16*, appears to be exclusively expressed in vertebrate jaw muscles (Hoh, 2002; Lee et al., 2019). This myosin is distinctive in that it appears to allow for a powerful bite for species that possess it, which makes it all the more notable that it is not found in all jawed vertebrates. More specifically, among mammals, it is distinctly absent in various

lineages of herbivorous taxa, such as kangaroos, cows, and most rodents (Hoh, 2002). Indeed, we found that it is similarly inactivated in the folivorous three-fingered sloths (*Bradypus* spp.) but remains fully functional in the two-fingered sloths (*Choloepus* spp.). While these and other herbivorous mammals have jaw-closing muscles, their composition is made up of other myosin classes (Hoh, 2002). Given that masticatory myosin provides power to those that retain it, perhaps shifting from faunivory to dietary habits that relies more on lateral, grinding movements renders *MYH16* superfluous.

Moreover, our results show that numerous lineages of myrmecophagous mammals possess inactivated *MYH16* genes. While faunivorous, species that consume social insects have reduced masticatory muscles. Indeed, the dentition of many myrmecophages is simplified or even completely absent, as in the cases of anteaters and pangolins. Furthermore, in the case of the latter taxa, much of the jaw musculature has been rearranged such that it is able to facilitate rapid extrusion of the tongue (Endo et al., 1998; Ferreira-Cardoso et al., 2020). Accordingly, in both anteaters and pangolins, *MYH16* has been pseudogenized, with pangolins sharing at least 12 inactivating mutations, pointing to an early loss in their ancestral lineage, and anteaters losing the gene more recently, following the split of *Cyclopes* from *Tamandua* and *Myrmecophaga*. Notably, armadillos lost *MYH16* probably even earlier than their anteaters, with seven shared mutations found in chlamyphorid armadillos, pointing to inactivation at least 37 Mya (Gibb et al., 2016). Still, some myrmecophagous taxa have retained a functional *MYH16*. In this regard, the aardvark appears particularly distinctive given its specialization to the myrmecophagous lifestyle, but the retention of a seemingly functional masticatory myosin gene suggests sufficient bite force in this species. The carnivoran myrmecophages, including the two strictly termitivorous aardwolf species, likewise retain a functional *MYH16* ortholog, which may reflect more recent adaptations to eating social insects and phylogenetic constraints associated to an ancestral carnivorous diet.

Contrasts in dental pseudogenes between recent and ancient myrmecophagous species

One of the best characterized examples of convergent evolution in the context of mammalian myrmecophagy concerns what is apparently an ubiquitous simplification of the dentition. Beyond the completely edentulous jaws of pangolins and anteaters and the loss of enamel in the aardvark and armadillos, various taxa display a reduction in the number of teeth, supernumerary teeth, a reduction in dental complexity and even variation in tooth number on the left and right side of the jaws (Koyasu, 1993; Davit-Béal et al., 2009; Charles et al., 2013). The association of tooth and enamel loss with dental pseudogenes has been well-documented up to this point, ranging from birds, turtles, baleen whales, and toads, to some of the species in this study (Meredith et al., 2013, 2014; Emerling et al., 2023). Indeed, the evidence is so extensive that it has been suggested that these species may aid in the discovery of loci tied to congenital dental diseases in humans (Emerling et al., 2017).

What we have added here are examples from opposite extremes of dental gene loss, namely pangolins and myrmecophagous carnivorans. Unsurprisingly,

pangolins have pseudogenes for all nine dental genes we examined, with nearly all of them providing evidence of a complete degradation of the dental development pathway prior to the common ancestor of Pholidota. This is consistent with the description of Eocene fossil stem pangolins (*Eomanis*, *Eurotamandua*) with already edentulous jaws (Storch, 1978, 1981; Gaudin et al., 2009). As for the myrmecophagous carnivorans, all four species showed evidence of functional retention for all nine genes, consistent with their enameled dentition, with a single exception. The sloth bear (*Melursus ursinus*) appears to have a pseudogenic allele of *ACP4*, a gene that participates in enamel development and whose inactivation leads to hypoplastic (thin) enamel (Kim et al., 2022; Liang et al., 2022). Notably, when the pseudogenization pattern of this gene has been compared with the other genes in detailed taxonomic datasets, it appeared to be one of the first to become inactivated during dental regression, including in baleen whales (Mysticeti), certain toothed whales (Odontoceti; Randall et al., 2022), in pilosans, and both chlamyphorid and dasypodid armadillos (Emerling et al., 2023).

Pseudogenes of the oral apparatus and their implications for the evolution of myrmecophagy

Convergent evolution, while incredibly common during the history of life, frequently does not take the same evolutionary path. Organisms can derive analogous solutions to the same adaptive problem, but even when such solutions are highly similar, they may take very different roads to arrive there. Anteaters and pangolins are remarkably alike in many aspects of their anatomy and behavior, so much so that anatomical systematic studies almost invariably grouped them together in a taxon known as Edentata (Glass, 1985; Reiss, 2001). It was only with the advent of molecular phylogenetics that it became clear that anteaters and pangolins evolved convergently, with pangolins being recognized as the closest living relatives of carnivorans (Murphy et al., 2001; Delsuc et al., 2002; Springer et al., 2019). Given their distinctive evolutionary histories, it raises the question of whether these independent lineages arose through a similar sequence of evolutionary events. What we found is that while there is some overlap, our analyses are limited in the resolution with which we can pinpoint the precise timing of these events.

For anteaters (Vermilingua), we have the benefit of possessing a sister-group (sloths; Folivora) that shares some of the regressive events, suggesting loss in their common ancestor (Pilosa). Furthermore, this branch length is relatively short (9 Myr), allowing for relatively fine temporal precision (Gibb et al., 2016). As reported in Emerling et al. (2023), five genes tied with enamel development (*ACP4*, *AMELX*, *AMTN*, *ENAM*, *MMP20*) possess shared inactivating mutations, strongly suggesting that enamel was lost on the stem pilosan branch. To this we can add evidence that the capacity for sweet taste perception was also likely lost on this branch, given shared mutations found in *TAS1R2*. While the xenarthran fossil record is silent on cranial anatomy during this era, comparative analyses of chitinase genes (*CHIA*s) suggest that the earliest pilosans were highly insectivorous (Emerling et al., 2018), suggesting that myrmecophagy may have been the dietary habit of these early xenarthrans.

Following the split of sloths (Folivora) and anteaters (Vermilingua), it would be about 21 Myr before the origin of crown vermilinguans (Gibb et al., 2016). In the intervening period, stem anteaters likely completely lost their teeth, as evidenced by the loss of *DSPP*, which encodes a dentin matrix protein, and *ODAPH*, a gene that appears to be uniquely linked to the retention of teeth (Springer et al., 2016; Emerling et al., 2023). Furthermore, a gene associated with the enamel-gingival junction (*ODAM*) was probably lost on this branch as well. After the split of vermilinguans into silky (Cyclopedidae) and other anteaters (Myrmecophagidae), each lineage lost the last remaining dental gene (*AMBN*) as well as the masticatory myosin (*MYH16*). Finally, in the terminal branches, evidence of further incipient taste loss is found in the polymorphic pseudogenes of *TAS1R3* in the giant anteater and *PKD2L1* in the silky anteater.

For pangolins, the temporal resolution is less precise given that crown Pholidota dates to approximately 41 Mya, but their closest living relatives are Carnivora, from which they split about 79 Mya (Heighton et al., 2023). On this 38 Myr stem branch, there were quite a few gene losses, including nearly every gene associated with enamel and the enamel-gingiva junction (*ACP4*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, *MMP20*) and *ODAPH*, whose inactivation points to tooth loss. The enamel gene *KLK4* and the dentin matrix gene *DSPP* may also have been lost on this branch, but due to missing data (*KLK4*) and a whole gene deletion (*DSPP*) in Asian pangolins, we cannot provide positive evidence of this. An early loss for these genes is to be expected, given that fossil pangolins from the Middle Eocene (47.8–38 Mya) were already edentulous (Storch, 1978, 1981; Gaudin et al., 2009). Whether enamel loss predated tooth loss, as suggested in anteaters, will need to be tested via other methods. Like anteaters, however, pangolins lost sweet gustation (*TAS1R2*) on the stem Pholidota branch, but unlike anteaters, they also lost the masticatory myosin (*MYH16*) prior to their last common ancestor. From there, individual lineages lost additional taste genes, associated with detection of umami (*TAS1R1*) and sour tastants (*PKD2L1*).

The results from pangolins and especially anteaters, paired with the data from armadillos and myrmecophagous carnivorans, point to another insight that we believe is worthy of further exploration. More specifically, these data seem to suggest that the regressive evolution that accompanies strict myrmecophagy appears to have occurred very gradually over a very extensive amount of evolutionary time. Crown Xenarthra dates to about 67 Mya and crown Pilosa to 58 Mya (Gibb et al., 2016). In this window of 9 Mya, the loss of enamel and sweet taste perception happened. In the following 58 Myr, there has been a gradual loss of teeth, then masticatory myosin; and it appears that only recently additional gustatory loss has begun. While the timing of the early events in pangolin history are currently shrouded in mystery, despite being fully edentulous, lacking sweet gustation and masticatory myosin by around 38 Mya, additional taste losses have occurred more recently, with perhaps only a single species, the white-bellied tree pangolin (*Phataginus tricuspis*), completely lacking sweet, umami, and sour taste detection.

This gradualistic trend towards strict myrmecophagy is further exemplified by the more intermediate states found in the aardvark and armadillos. The aardvark is the

only extant species from Tubulidentata, with a long branch splitting from other afrotherian mammals around 79 Mya (Meredith et al., 2011). However, fossil aardvarks exist, with the enamelless *Orycteropus minutus* dating to approximately 19 Mya (Pickford, 1974). Despite having such reduced teeth for at least 19 million years, reflected in the pseudogenization of underlying enamel genes, the aardvark only shows evidence of a single taste gene loss (*PKD2L1*), no evidence of masticatory myosin loss, and no trend towards complete tooth loss. Myrmecophagous tolypeutine armadillos within the family Chlamyphoridae have an even earlier history of dental reduction, with evidence of tooth simplification occurring by the origin of their clade (Cingulata) about 45 Mya (Emerling et al., 2023). This, in turn, was followed by the loss of two genes tied to the enamel-gingiva junction (*AMTN*, *ODAM*) and a gene associated with the thinning of enamel (*ACP4*). This corresponds well with the fossil armadillo *Utaetus buccatus* (42–39 Mya), which had thin enamel that wore easily (Simpson, 1932; Ciancio et al., 2014). On this same stem chlamyphorid branch, which dates between 45 and 37 Mya, the masticatory myosin (*MYH16*) was lost. From here, Emerling et al. (2023) inferred that complete enamel loss occurred in parallel within separate lineages. At most, a few species appear to have lost the capacity of sour taste (*PKD2L1*) and one (*Tolypeutes matacus*) lost umami taste (*TAS1R1*). Again, no evidence of trending towards complete tooth loss was found.

Finally, the myrmecophagous carnivorans give hints as to the state of very recent adaptations to this dietary habit. We found no evidence of pseudogene fixation in the three species examined, although more thorough sampling of individuals would be required to address this clearly. That said, pseudogene polymorphisms were found for the umami taste receptor in the Southern aardwolf (*TAS1R1*), the sweet taste receptor in the Eastern aardwolf (*TAS1R2*), and an enamel-associated gene in the sloth bear (*ACP4*).

Future Directions

While this study has begun to explore the genetic consequences of myrmecophagy, specifically in relation to the oral apparatus, there are a number of additions that we plan to make to more thoroughly test our hypotheses. First, we will include data from beyond placental mammals to consider a marsupial myrmecophage, the numbat (*Myrmecobius fasciatus*), and a monotreme myrmecophage, the short-beaked echidna (*Tachyglossus aculeatus*). These would give us extra replicates of the natural experiment of convergent evolution toward myrmecophagy to examine the underlying genetic consequences of adopting this highly specialized dietary habit.

A second shortcoming we acknowledge concerns the patterns of gene loss in bitter taste receptor genes (*TAS2R*s) that were not explored here. Existing data suggest that these may similarly become pseudogenized in myrmecophagous mammals (Liu et al., 2016; Peel et al., 2022), but given that they belong to a relatively complex gene family with dozens of paralogs, it will require careful analysis to understand the evolutionary trends for bitter taste in our focal taxa.

Finally, examining patterns of pseudogenization in a phylogenetic context provides insights into the timing and distribution of gene loss, in turn helping us to

understand trait regression. However, analyses of selective pressures based on dN/dS ratios, particularly in the context of molecular clocks, may help us to more precisely estimate the patterns and timing of gene loss. We believe this will be particularly helpful in the context of lineages and clades that are taxon poor (*e.g*., aardvark) and/or have species with highly similar phenotypes (*e.g*., pangolins). We intend to add such analyses to improve our resolution in studying the evolution of these genes.

ACKNOWLEDGEMENTS

This work has been supported by grants from the European Research Council (ConvergeAnt project: ERC-2015-CoG-683257) and Investissements d'Avenir of the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01; CEMEB: ANR-10-LABX-0004). This is contribution ISEM 2023-XXX of the Institut des Sciences de l'Evolution de Montpellier.

REFERENCES

Adeniyi T, Tijani A, Adekomi D, Abayomi T. 2012. A comparative study of the lateral geniculate body of rat (*Rattus norvegicus*), bat (*Eidolon helvum*) and pangolin (*Manis tricuspis*). *Glob J Health Sci* 4:118–125.

Albalat R, Cañestro C. 2016. Evolution by gene loss. *Nat Rev Genet* 17:379–391.

Allio R. 2021. Phylogenomics and comparative genomics in ant-eating mammals. *PhD Thesis*. Université de Montpellier.

Allio R, Tilak M-K, Scornavacca C, Avenant NL, Kitchener AC, Corre E, Nabholz B, Delsuc F. 2021. High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox. *Elife* 10:e63167.

Blount ZD, Lenski RE, Losos JB. 2018. Contingency and determinism in evolution: Replaying life's tape. *Science* 362:eaam5979.

Casali DM, Martins-Santos E, Santos ALQ, Miranda FR, Mahecha GAB, Perini FA. 2017. Morphology of the tongue of Vermilingua (Xenarthra: Pilosa) and evolutionary considerations. *J Morph* 278:1380–1399.

Cerca J. 2023. Understanding natural selection and similarity: Convergent, parallel and repeated evolution. *Mol Ecol* in press.

Chandrashekar J, Hoon MA, Ryba NJ, Zuker CS. 2006. The receptors and cells for mammalian taste. *Nature* 444:288–294.

Charles C, Solé F, Rodrigues HG, Viriot L. 2013. Under pressure? Dental adaptations to termitophagy and vermivory among mammals. *Evolution* 67:1792–1804.

Chaudhari N, Roper SD. 2010. Review series: The cell biology of taste. *J Cell Biol* 190:285.

Choo SW, Rayko M, Tan TK, Hari R, Komissarov A, Wee WY, Yurchenko AA, Kliver S, Tamazian G, Antunes A, et al. 2016. Pangolin genomes and the evolution of mammalian scales and immunity. *Genome Res* 26:1312–1322.

Ciancio MR, Vieytes EC, Carlini AA. 2014. When xenarthrans had enamel: insights on the evolution of their hypsodonty and paleontological support for independent evolution in armadillos. *Naturwissenschaften* 101:715–725.

Cole TL, Zhou C, Fang M, Pan H, Ksepka DT, Fiddaman SR, Emerling CA, Thomas DB, Bi X, Fang Q, et al. 2022. Genomic insights into the secondary aquatic transition of penguins. *Nat Commun* 13:3912.

Dalla Valle L, Benato F, Rossi C, Alibardi L, Tschachler E, Eckhart L. 2011. Deleterious mutations of a claw keratin in multiple taxa of reptiles. *J Mol Evol* 72:265–273.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008.

Davit-Béal T, Tucker AS, Sire J-Y. 2009. Loss of teeth and enamel in tetrapods: Fossil record, genetic data and morphological adaptations. *J Anat* 214:477–501.

Delsuc F, Scally M, Madsen O, Stanhope MJ, de Jong WW, Catzeflis FM, Springer MS, Douzery EJP. 2002. Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. *Mol Biol Evol* 19:1656–1671.

Doran GA, Allbrook DB. 1973. The tongue and associated structures in two species of African pangolins, *Manis gigantea* and *Manis tricuspis*. *J Mammal* 54:887–899.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356:92–95.

Dudchenko O, Shamim M, Batra S, Durand N, Musial N, Mostofa R, Pham M, Glenn St Hilaire B, Yao W, Stamenova E, et al. 2018. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. bioRxiv https://doi.org/10.1101/254797

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nuc Acid Res* 32:1792–1797.Emerling CA. 2017. Genomic regression of claw keratin, taste receptor and light-associated genes provides insights into biology and evolutionary origins of snakes. *Mol Phylogenet Evol* 115:40–49.

Emerling CA. 2022. The evolution of squamate chitinase genes (*CHIAs*) supports an insectivory–carnivory transition during the early history of snakes. In: Gower DJ, Zaher H, editors. The Origin and Early Evolutionary History of Snakes. Systematics Association Special Volume Series. Cambridge: Cambridge University Press. p. 235–247.

Emerling CA, Delsuc F, Nachman MW. 2018. Chitinase genes (*CHIAs*) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. *Sci Adv* 4:eaar6478.

Emerling CA, Springer MS, Gatesy J, Jones Z, Hamilton D, Xia-Zhu D, Collin M, Delsuc F. 2021. Genomic evidence for the parallel regression of melatonin synthesis and signaling pathways in placental mammals. *Open Research Europe* 1:75.

Emerling CA, Widjaja AD, Nguyen NN, Springer MS. 2017. Their loss is our gain: regressive evolution in vertebrates provides genomic models for uncovering human disease loci. *J Med Genet* 54:787–794.

Endo H, Nishiumi I, Kurohmaru M, Nabhitabhata J, Chan-Ard T, Nadee N, Agungpriyono S, Yamada J. 1998. The functional anatomy of the masticatory muscles of the Malayan pangolin, *Manis javanica*. *Mammal Study* 23:1–8.

Feng P, Zheng J, Rossiter SJ, Wang D, Zhao H. 2014. Massive losses of taste receptor genes in toothed and baleen whales. *Genome Biol Evol* 6:1254–1265.

Ferreira-Cardoso S, Delsuc F, Hautier L. 2019. Evolutionary tinkering of the mandibular canal linked to convergent regression of teeth in placental mammals. *Curr Biol* 29:468–475.

Ferreira-Cardoso S, Fabre P-H, Thoisy B de, Delsuc F, Hautier L. 2020. Comparative masticatory myology in anteaters and its implications for interpreting morphological convergence in myrmecophagous placentals. *PeerJ* 8:e9690.

Fong DW, Kane TC, Culver DC. 1995. Vestigialization and loss of nonfunctional characters. *Annu Rev Ecol Syst* 26:249–268.

Gaudin TJ, Emry RJ, Wible JR. 2009. The phylogeny of living and extinct pangolins (Mammalia, Pholidota) and associated taxa: a morphology based analysis. *J Mammal Evol* 16:235–305.

Gibb GC, Condamine FL, Kuch M, Enk J, Moraes-Barros N, Superina M, Poinar HN, Delsuc F. 2016. Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans. *Mol Biol Evol* 33:621–642.

Glass BP. 1985. History of classification and nomenclature in Xenarthra (Edentata). In *The evolution and ecology of armadillos, sloths, and vermilinguas*:51–64.

Hoh JF. 2002. Superfast'or masticatory myosin and the evolution of jaw-closing muscles of vertebrates. *J Exp Biol* 205:2203–2210.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinf* 12:491.

Hu Y, Wu Q, Ma S, Ma T, Shan L, Wang X, Nie Y, Ning Z, Yan L, Xiu Y, et al. 2017. Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proc Natl Acad Sci USA* 114:1081–1086.

Janiak MC, Chaney ME, Tosi AJ. 2018. Evolution of acidic mammalian chitinase genes (*CHIA*) is related to body mass and insectivory in primates. *Mol Biol Evol* 35:607–622.

Jiang P, Josue J, Li X, Glaser D, Li W, Brand JG, Margolskee RF, Reed DR, Beauchamp GK. 2012. Major taste loss in carnivorous mammals. *Proc Natl Acad Sci USA* 109:4956–4961.

Jiao H, Zhang L, Xie H-W, Simmons NB, Liu H, Zhao H. 2019. Trehalase gene as a molecular signature of dietary diversification in mammals. *Mol Biol Evol* 36:2171–2183.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.

Kim YJ, Lee Y, Kasimoglu Y, Seymen F, Simmer JP, Hu JC-C, Cho E-S, Kim J-W. 2022. Recessive mutations in *ACP4* cause amelogenesis imperfecta. *J Dent Res* 101:37–45.

Kishida T, Thewissen J, Hayakawa T, Imai H, Agata K. 2015. Aquatic adaptation and the evolution of smell and taste in whales. *Zoological Lett* 1:9.

Korf I. 2004. Gene finding in novel genomes. *BMC Bioinf* 5:59.

Koyasu K. 1993. Variability of dentition in the bat-eared fox, *Otocyon megalotis*. *J Dent Sci* 31:63–104.

Kubota K, Kubota J, Fukuda N, Asakura S. 1962. Comparative anatomical and neurohistological observations on the tongue of the great anteater (*Myrmecophaga Jubata* Linne). *Anat Rec* 143:15–25.

Kubota K, Kubota J, Nakamura T, Fukuda N, Asakura S, Nakagawa S, Masui M. 1962. Comparative anatomical and neurohistological observations on the tongue of the pangolin (*Manis pentadactyla*, Linneus). *Anat Rec* 144:43–55.

Kumar V, Lammers F, Bidon T, Pfenninger M, Kolter L, Nilsson MA, Janke A. 2017. The evolutionary history of bears is characterized by gene flow across species. *Sci Rep* 7:46487.

Lagerström M, Dahl N, Nakahori Y, Nakagome Y, Bäckman B, Landegren U, Pettersson U. 1991. A deletion in the amelogenin gene (*AMG*) causes X-linked amelogenesis imperfecta (AIH1). *Genomics* 10:971–975.

Lahti DC, Johnson NA, Ajie BC, Otto SP, Hendry AP, Blumstein DT, Coss RG, Donohue K, Foster SA. 2009. Relaxed selection in the wild. *Trends Ecol Evol* 24:487–496.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359.

Lee LA, Karabina A, Broadwell LJ, Leinwand LA. 2019. The ancient sarcomeric myosins found in specialized muscles. *Skeletal Muscle* 9:7.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.

Liang T, Wang S-K, Smith C, Zhang H, Hu Y, Seymen F, Koruyucu M, Kasimoglu Y, Kim J-W, Zhang C, et al. 2022. Enamel defects in *Acp4R110C/R110C* mice and human *ACP4* mutations. *Sci Rep* 12:16477.

Liman ER, Kinnamon SC. 2021. Sour taste: receptors, cells and circuits. *Curr Opin Physiol* 20:8–15.

Liu Z, Liu G, Hailer F, Orozco-terWengel P, Tan X, Tian J, Yan Z, Zhang B, Li M. 2016. Dietary specialization drives multiple independent losses and gains in the bitter taste gene repertoire of Laurasiatherian Mammals. *Front Zool* 13:28.

Meredith RW, Gatesy J, Murphy WJ, Ryder OA, Springer MS. 2009. Molecular decay of the tooth gene Enamelin (*ENAM*) mirrors the loss of enamel in the fossil record of placental mammals. *PLoS Genet* 5:e1000634.

Meredith RW, Gatesy J, Springer MS. 2013. Molecular decay of enamel matrix protein genes in turtles and other edentulous amniotes. *BMC Evol Biol* 13:20.

Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simao TLL, Stadler T, et al. 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* 334:521–524.

Meredith RW, Zhang G, Gilbert MTP, Jarvis ED, Springer MS. 2014. Evidence for a single loss of mineralized teeth in the common avian ancestor. *Science* 346:1254390.

Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, Jong WW de, et al. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348–2351.

Parry DA, Brookes SJ, Logan CV, Poulter JA, El-Sayed W, Al-Bahlani S, Al Harasi S, Sayed J, Shore RC, Dashash M. 2012. Mutations in *C4orf26*, encoding a peptide with in vitro hydroxyapatite crystal nucleation and growth activity, cause amelogenesis imperfecta. *Am J Human Genet* 91:565–571.

Peel E, Silver L, Brandies P, Hayakawa T, Belov K, Hogg CJ. 2022. Genome assembly of the numbat (*Myrmecobius fasciatus*), the only termitivorous marsupial. *GigaByte* 2022:gigabyte47.

Pickford M. 1974. New fossil Orycteropodidae (Mammalia, Tubulidentata) from East Africa. *Orycteropus minutus* sp. nov. and *Orycteropus chemeldoi* sp. nov. *Netherlands J Zool* 25:57–88.

Poulter JA, Murillo G, Brookes SJ, Smith CE, Parry DA, Silva S, Kirkham J, Inglehearn CF, Mighell AJ. 2014. Deletion of ameloblastin exon 6 is associated with amelogenesis imperfecta. *Human Mol Genet* 23:5317–5324.

Rajpar MH, Harley K, Laing C, Davies RM, Dixon MJ. 2001. Mutation of the gene encoding the enamel-specific protein, enamelin, causes autosomal-dominant amelogenesis imperfecta. *Human Mol Genet* 10:1673–1677.

Randall JG, Gatesy J, Springer MS. 2022. Molecular evolutionary analyses of tooth genes support sequential loss of enamel and teeth in baleen whales (Mysticeti). *Mol Phylogenet Evol* 171:107463.

Redford KH. 1987. Ants and termites as food. In: Genoways HH, editor. Current Mammalogy. Boston, MA: Springer US. p. 349–399.

Reiss KZ. 2001. Using phylogenies to study convergence: the case of the ant-eating mammals. *Am Zoologist* 41:507–525.

Roper SD, Chaudhari N. 2017. Taste buds: cells, signals and synapses. *Nat Rev Neuro* 18:485–497.

Sato JJ, Wolsan M. 2012. Loss or major reduction of umami taste sensation in pinnipeds. *Naturwissenschaften* 99:655–659.

Schwenk K. 1985. Occurrence, distribution and functional significance of taste buds in lizards. *Copeia* 1985:91–101.

Seymen F, Kim YJ, Lee YJ, Kang J, Kim T-H, Choi H, Koruyucu M, Kasimoglu Y, Tuna EB, Gencay K. 2016. Recessive mutations in *ACPT*, encoding testicular acid phosphatase, cause hypoplastic amelogenesis imperfecta. *Am J Human Genet* 99:1199–1205.

Seymen F, Park J-C, Lee K-E, Lee H-K, Lee D-S, Koruyucu M, Gencay K, Bayram M, Tuna EB, Lee ZH, et al. 2015. Novel *MMP20* and *KLK4* mutations in amelogenesis imperfecta. *J Dent Res* 94:1063–1069.

Shaheen J, Mudd AB, Diekwisch TG, Abramyan J. 2021. Pseudogenized amelogenin reveals early tooth loss in true toads (Anura: Bufonidae). *Integ Comp Biol* 61:1933–1945.

Simpson GG. 1932. Enamel on the teeth of an Eocene edentate. *Am Mus Novit* 567.

Smith CE, Murillo G, Brookes SJ, Poulter JA, Silva S, Kirkham J, Inglehearn CF, Mighell AJ. 2016. Deletion of amelotin exons 3–6 is associated with amelogenesis imperfecta. *Human Mol Genet* 25:3578–3587.

Smith CE, Poulter JA, Antanaviciute A, Kirkham J, Brookes SJ, Inglehearn CF, Mighell AJ. 2017. Amelogenesis imperfecta; genes, proteins, and pathways. *Frontiers Physiol* 8:435.

Springer MS, Emerling CA, Gatesy J, Randall J, Collin MA, Hecker N, Hiller M, Delsuc F. 2019. Odontogenic ameloblast-associated (*ODAM*) is inactivated in toothless/enamelless placental mammals and toothed whales. *BMC Evol Biol* 19:31.

Springer MS, Meredith RW, Teeling EC, Murphy WJ. 2013. Technical comment on “The placental mammal ancestor and the post–K-Pg radiation of placentals.” *Science* 341:613–613.

Stedman HH, Kozyak BW, Nelson A, Thesier DM, Su LT, Low DW, Bridges CR, Shrager JB, Minugh-Purvis N, Mitchell MA. 2004. Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature* 428:415–418.

Storch G. 1978. Messel fossil finds part 14: *Eomanis waldi* (new genus, new species), a pangolin from the middle Eocene of the messel pit near darmstadt, West Germany (Mammalia, Pholidota). *Senckenb Leth* 59:503–530.

Storch G. 1981. *Eurotamandua joresi*, ein Myrmecophagidae aus dem Eozan der Grubbe Messel bei Darmstadt (Mammalia, Xenarthra). *Senckenb Leth* 61:247–269.

Thibaud-Nissen F, DiCuccio M, Hlavina W, Kimchi A, Kitts PA, Murphy TD, Pruitt KD, Souvorov A. 2016. P8008 the NCBI eukaryotic genome annotation pipeline. *J Anim Sci* 94:184–184.

Tu Y-H, Cooper AJ, Teng B, Chang RB, Artiga DJ, Turner HN, Mulhall EM, Ye W, Smith AD, Liman ER. 2018. An evolutionarily conserved gene family encodes proton-selective ion channels. *Science* 359:1047–1050.

Valente R, Alves F, Sousa-Pinto I, Ruivo R, Castro LFC. 2021. Functional or vestigial? The genomics of the pineal gland in Xenarthra. *J Mol Evol* 89:565–575.

Wazen RM, Moffatt P, Ponce KJ, Kuroda S, Nishio C, Nanci A. 2015. Inactivation of the Odontogenic ameloblast-associated gene affects the integrity of the junctional epithelium and gingival healing. *Eur Cell Mater* 30:187–199.

Wolsan M, Sato JJ. 2020. Parallel loss of sweet and umami taste receptor function from phocids and otarioids suggests multiple colonizations of the marine realm by pinnipeds. *J Biogeogr* 47:235–249.

Wolsan M, Sato JJ. 2022. Role of feeding specialization in taste receptor loss: insights from sweet and umami receptor evolution in Carnivora. *Chem Senses* 47:bjac033.

Xiao S, Yu C, Chou X, Yuan W, Wang Y, Bu L, Fu G, Qian M, Yang J, Shi Y. 2001. Dentinogenesis imperfecta 1 with or without progressive hearing loss is associated with distinct mutations in DSPP. *Nat Genet* 27:201–204.

Yarmolinsky DA, Zuker CS, Ryba NJ. 2009. Common sense about taste: from mammals to insects. *Cell* 139:234–244.

Yin D, Zhou R, Yin M, Chen Y, Xu S, Yang G. 2021. Gene duplication and loss of *AANAT* in mammals driven by rhythmic adaptations. *Mol Biol Evol* 38:3925–3937.

Young BA. 1997. On the absence of taste buds in monitor lizards (*Varanus*) and snakes. *J Herpetol* 31:130–137.

Zhao H, Li J, Zhang J. 2015. Molecular evidence for the loss of three basic tastes in penguins. *Curr Biol* 25:R141-142.

Zhao H, Xu D, Zhang S, Zhang J. 2012. Genomic and genetic evidence for the loss of umami taste in bats. *Genome Biol Evol* 4:73–79.

Zhao H, Yang J-R, Xu H, Zhang J. 2010. Pseudogenization of the umami taste receptor gene *Tas1r1* in the giant panda coincided with its dietary switch to bamboo. *Mol Biol Evol* 27:2669–2673.

I.2.3. Evolution of bitter taste perception in myrmecophagous mammals

Context and short introduction

Myrmecophagous mammals represent a textbook example of dietary convergent adaptation. Their independent evolution toward the same highly specialized diet raises questions regarding the mechanisms involved in their convergent adaptation. Among their remarkable morphological adaptations, their elongated tongue does not seem to be particularly suited for taste perception and is rather used to catch prey suggesting that myrmecophagous species might rely less on taste perception (*e.g.*, Doran and Allbrook, 1973; Casali *et al*, 2017). Because of their convergent adaptation to the same specialized diet, these species might share similar underlying genomic adaptations and therefore may have undergone a reduction of their *Tas2r* gene repertoires, which is involved in bitter taste perception, an important taste, notably for toxic food identification. Few studies have focused on *Tas2r* gene repertoires of myrmecophagous mammals but gene losses have been reported in the Chinese pangolin, (*Manis pentadactyla*; Liu *et al*, 2016), the short-beaked echidna (*Tachyglossus aculeatus*; Zhou *et al*, 2021), and the numbat (*Myrmecobius fasciatus*; Peel *et al*, 2022), the latter two being respectively a monotreme and marsupial ant-eating species. These results suggest that other myrmecophagous species might also have experienced a reduction of their TAS2R repertoires. Besides, some ant and termite species produce toxic compounds (Schmidt, 1986; Lopez and Morgan, 1997) that could be detected by TAS2R receptors. Therefore, gene repertoires of myrmecophagous species might have evolved toward a loss of receptors as a result of their low need to taste their food with their tongue, and/or a retention of TAS2Rs enabling the detection of bitter tastants. In order to test these hypotheses, we will study the evolution of the *Tas2r* gene family in placentals as a follow up of the Master Project of Kathleen Garland (2018) who studied *Tas2r* genes in 177 placental species.

The aim of this study is to i) reconstruct the evolutionary history of the *Tas2r* gene family in placentals, ii) compare gene repertoires among myrmecophagous species to understand whether similar genomic adaptations (*i.e.*, shared gene losses or duplications) were involved in the different focal species, and iii) contrast myrmecophagous species with

their closest non-myrmecophagous relatives to highlight a potential convergent pattern of gene repertoire evolution. The dataset was finalized at the end of my PhD and is presented below. An overview of the analyses that could be done after my PhD is given afterwards. The questions they should help answer is discussed in the light of previous studies on myrmecophagous taste receptors, notably of Garland (2018) for which a summary of her main results is given.

Preliminary material and methods: genome selection

High-quality genomes of myrmecophagous and non-myrmecophagous placental species that will be used in this study were generated as part of the ConvergeAnt project (n = 8) with some HiC-upgraded by the DNA Zoo (n = 3; Dudchenko *et al*, 2017; Table S I.1), and downloaded from publicly available databases based on their quality (NCBI Genbank: n = 97, and the DNA Zoo: n = 122; Table S I.1).

Genomes from the ConvergeAnt project were assembled from Nanopore long reads and short Illumina reads using the hybrid assembler MaSuRCA v3.2.9 (Zimin *et al*, 2017). *De novo* genome annotation was done using the MAKER v3 pipeline (Holt and Yandell, 2011) following the strategy designed by the DNA Zoo (www.dnazoo.org) to use different annotation approaches (*i.e*., gene prediction, transcriptomics, and protein sequences). A detailed description of these steps is given in Allio (2021) and Allio *et al* (2021). ConvergeAnt genome assemblies are available for the common silky anteater (*Cyclopes didactylus*), the giant anteater (*Myrmecophaga tridactyla*; HiC updated), the southern tamandua (*Tamandua tetradactyla*; HiC updated), the pale-throated sloth (*Bradypus tridactylus*), the pink fairy armadillo (*Chlamyphorus truncatus*), the six-banded armadillo (*Euphractus sexcinctus*), the giant armadillo (*Priodontes maximus*), the giant pangolin (*Smutsia gigantea*), the southern aardwolf (*Proteles cristatus*), and the bat-eared fox (*Otocyon megalotis*; HiC updated), an insectivorous species considered partially myrmecophagous. The Discovar draft genome assembly of the southern naked-tailed armadillo (*Cabassous unicinctus*) generated by Emerling *et al* (2023), although very fragmented, will also be used in this study as it is the only genome assembly available for this species.

Genomes from the DNA Zoo and Genbank were selected based on their quality, and to ensure the dataset was representative of the placental diversity, as follow:

- Genome statistics were downloaded from NCBI for all placental assemblies; the same was done for the placental DNA Zoo assemblies.
- One genome per genera was selected (except for myrmecophagous species for which all available assemblies were selected) as follow:
    - When an HiC assembly (*i.e.*, DNA Zoo assembly) was available, this assembly was selected.
    - Otherwise, the selection was based on the N50 of the scaffolds with a threshold of at least 100 000 bp.
- After this selection and for families with more than 10 genomes selected, one genome was selected per subfamily using the same criteria.

In total, the dataset includes 230 genomes (the selection was done in May 2023; it does not include genomes that have been published later) (Fig I.6, Table S I.1).

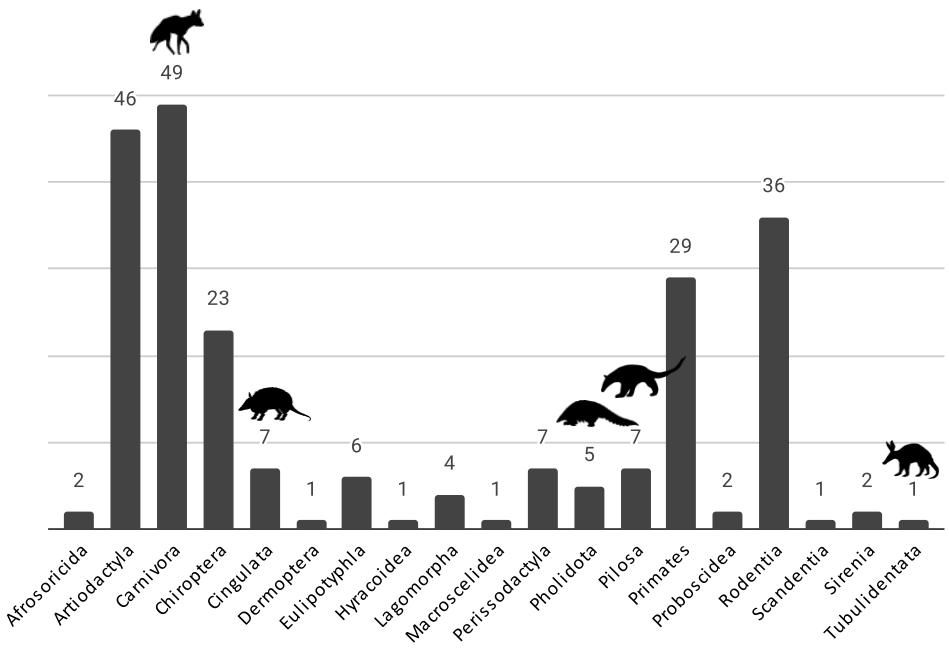


Figure I.6. Number of placental genomes per order used in this study. Numbers above bars indicate the number of genomes. Silhouettes were downloaded from phylopic.org to illustrate myrmecophagous placental orders.

Preliminary exploration of the dataset: *Tas2r* gene tree reconstruction

To start exploring this dataset, a preliminary *Tas2r* gene tree was reconstructed. When available, annotation files were downloaded (n = 170; Table S I.1; Fig S I.1). They correspond to MAKER annotation files for the ConvergeAnt and DNA Zoo genomes (n = 126) and files with CDS sequences extracted from the genomic assemblies for the Genbank genomes (n = 44). Annotated *Tas2r* sequences were extracted from these files (n = 1 692). These sequences were then aligned with MAFFT v7.490 (Katoh *et al*, 2002; Katoh and Standley, 2013) used with default parameters (gap opening penalty = 1.53; offset = 0, maximum number of iterative refinements = 0). Vomeronasal type 1 receptor gene (*VN1R*) sequences of *Homo sapiens* (n = 5) served as outgroup sequences. VN1Rs are closely related to TASRs and are also transmembrane chemosensory receptors coupled with G-proteins present in the vomeronasal organ and involved in the perception of pheromones (Adler *et al*, 2000; Pantages and Dulac, 2000). A gene-tree (Fig I.7) was then inferred using RAxML v8 (Stamatakis, 2014) with the GTR+GAMMA model and the rapid hill-climbing algorithm model and otherwise default parameters. This tree was rooted on the VN1R clade. Additional cleaning was done to remove sequences corresponding to very long-branches and not well aligned resulting in a total number of sequences of 1 673.

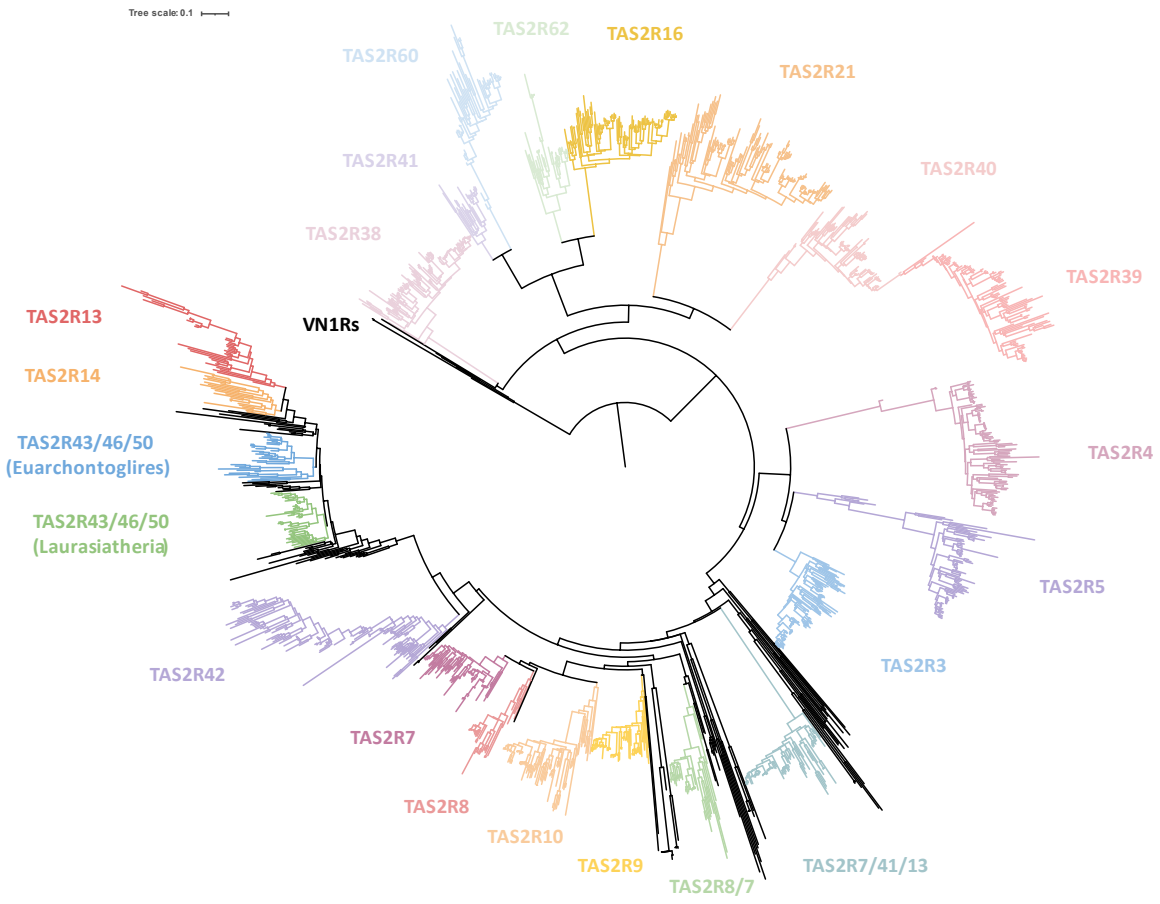


Figure I.7. *Tas2r* gene tree of placental mammals. This tree was obtained by using a maximum-likelihood approach on the set of 1 673 *Tas2r* annotated sequences extracted from 170 high-quality selected placental genomes and the VN1Rs sequences of *Homo sapiens* used as outgroup sequences. Clades including representants of at least two of the four placental super-orders are distinguished by different colors.

Some sequences were not well placed in the phylogeny and represented by long branches (Fig I.7). Importantly, this gene tree was reconstructed based on nucleotide sequences, which are less conserved than protein ones. Therefore, future analyses that will be done in the purpose of doing a gene-tree/species-tree reconciliation will be conducted on amino acid sequences. Additional data cleaning will also be done, for instance to remove mis-annotated sequences. In certain species, no *Tas2r* was identified, it might be due to annotation errors; this will also need additional verification notably by checking for the presence of the gene directly in the genome assembly to assess whether it is pseudogenized or not. Overall, 22 clades can be distinguished including representants of at least two of the

four placental super-orders (Fig I.7). Clade nomenclature is based on the annotation of the sequences composing the clade (Fig I.7) but is not clear for all clades (*e.g*., TAS2R43/46/50 clade). This problem can notably result from annotation errors with sequences similar to specific paralogues (for instance because of concerted evolution between genes) that can be mis-annotated. Adding more sequences and better resolving phylogenetic relationships (*e.g*., using amino acid sequences) will help clarify their nomenclature.

Preliminary exploration of the dataset: numbers of *Tas2r* annotated sequences in the selected placental genomes

This first preliminary analysis revealed that, on average, 9.8 *Tas2r* annotated genes were identified per species. Between one and 21 genes were identified (Table S I.1): the southern giant pouched rat (*Cricetomys ansorgei*, Nesomyidae), the broad-toothed rat (*Mastacomys fuscus*, Muridae), the Mongolian gerbil (*Meriones unguiculatus*, Muridae), the cactus mouse (*Peromyscus eremitus*, Cricetidae), and the indo-pacific bottlenose dolphin (*Tursiops aduncus*, Delphinidae) having only one annotated *Tas2r* gene, and the Angolan colobus monkey (*Colobus angolensis*, Cercopithecidae) and the Sumatran orangutan (*Pongo abelii*, Hominidae) having 21 annotated *Tas2r* genes (Table S I.1). Marine mammals seem to carry few *Tas2r* genes, between one in *Tursiops aduncus* and 11 in the northern fur seal (*Callorhinus ursidae*, Otariidae) (Table S I.1). Those having the most annotated genes (between seven and 11) are marine mammals mostly belonging to the Phocidae and Otariidae families whereas whales, dolphins, and porpoises present in our dataset have the least annotated genes (between one and six).

On the contrary, primates include species with the most annotated *Tas2r* genes with a number of identified genes varying between 12 (Philippine tarsier, *Carlito syrichta,* Tarsiidae) and 21 (*Pongo abelii*, *Colobus angolensis*) (Table S I.1). Several species have 19 annotated genes like the pygmy marmoset (*Cebuella pygmaea*, Cebidae), the squirrel monkey (*Saimiri boliviensis*, Cebidae), and the Coquerel's sifaka (*Propithecus coquereli*, Indriidae), or 20 genes such as the Goeldi's monkey (*Callimico goeldii*, Cebidae) and brown greater galago (*Otolemur crassicaudatus*, Galagidae). Some rodent species and herbivorous species belonging to the Perissodactyla also have high numbers of *Tas2r* annotated genes. For instance, 20 genes were

identified in the domestic guinea pig (*Cavia porcellus*, Caviidae) and the mountain zebra (*Equus zebra*, Equidae), 18 in the Asian elephant (*Elephas maximus,* Elephantidae), and 17 in the capybara (*Hydrochoerus hydrochaeris*) (Table S I.1). Among Carnivora, numbers of annotated *Tas2r* genes vary between five in the harbor porpoise (*Phocoena phocoeana*, Phocoenidae) and 17 in the dingo (*Canis lupus dingo*, Canidae), the maned wolf (*Chrysocyon brachyurus*, Canidae), and the clouded leopard (*Neofelis nebulosa*, Felidae) (Table S I.1).

Finally, in myrmecophagous species few annotated genes were identified with numbers varying between three in the Giant pangolin (*Smutsia gigantea*, Manidae) and the Sunda pangolin (*Manis javanica*, Manidae) and 12 in the aardvark (*Orycteropus afer*, Orycteroporidae) and the bat-eared fox (*Otocyon megalotis*, Canidae) (Table S I.1). Four and five genes were identified in the two other pangolin species studied here (respectively the Chinese pangolin, *Manis pentadactyla*, and the African tree pangolin, *Phataginus tricuspis*). In the three anteater species present in our dataset four, six, and seven genes were identified respectively in the silky anteater (*Cyclopes didactylus*), the giant anteater (*Myrmecophaga tridactyla*), and the southern tamandua (*Tamandua tetradactyla*). The southern aardwolf (*Proteles cristatus*), the other myrmecophagous carnivoran species, has six genes. In armadillos, numbers vary a bit more, between five for the six-banded armadillo (*Euphractus sexcinctus*) and nine for the nine-banded armadillo (*Dasypus novemcinctus*), which are generalist species and considered insectivorous/omnivorous. The giant armadillo (*Priodontes maximus*), a fully myrmecophagous armadillo, has eight annotated *Tas2r* genes.

Future analyses and prospects

➔ **Evolution of *Tas2r* gene repertoires in placentals**

Overall, the first exploration of this dataset highlights the fact that *Tas2r* gene repertoires in placental mammals can vary greatly between species and suggests an evolution of this gene family following a birth-and-death model with numbers of gene duplications and losses varying between lineages, consistent with the literature (Hayakawa *et al*, 2014; Liu *et al*, 2016). Garland (2018) identified 25 clades suggesting the common ancestor of placentals had at least 25 *Tas2r* functional gene paralogues and varying numbers of genes between species. Notably, she found high number of duplications within the subfamilies *Tas2r4* and *Tas2r14*

and defined a new clade (*Tas2Rnew*). In this clade, five main duplications lead to several gene duplications notably in primates, with for instance, *H. sapiens* having 21 copies within this new subfamily. Using a subset of 34 placental species to study in more detail patterns of gene inactivation and the number of functional genes and pseudogenes, Garland (2018) found an average number of 14.02 absent or inactive genes. The tree-shrew (*Tupaia belangeri*) had the fewest number of gene losses with 19 gene copies retained and *Balaenoptera acutorostrata*, *Orcinus orca*, and *Enhydra lutris* were the species with the highest number of gene losses (22 paralogues lost) (Garland, 2018). These results are consistent with previous studies also showing high number of gene losses in marine mammals (Feng *et al*, 2014; Kishida *et al*, 2015). These results are also consistent with our preliminary analyses showing that marine mammals seem to carry small *Tas2r* repertoires whereas primates seem to have much more *Tas2r* annotated functional genes.

➔ ***Tas2r* gene losses in myrmecophagous mammals**

Myrmecophagous species are expected to present a reduction of their *Tas2r* gene repertoires and our exploratory analyses showed few annotated *Tas2rs* in their genomes suggesting they might indeed have lost some copies. Besides, Garland (2018) showed a reduction of the number of functional *Tas2r* genes in the four myrmecophagous species she analyzed (*C. didactylus*, *M. javanica*, *O. afer*, and *T. tetradactyla*). More specifically, *M. javanica*, *T. tetradacyla*, and *C. didactylus* respectively lost 21, 20, and 19 out of 25 genes and the aardvark 14 out of 25, and no gene duplication was found in myrmecophagous mammals (Garland, 2018). These results are consistent with previous studies who found reduced *Tas2r* repertoires in ant-eating mammals such as the Chinese pangolin in which only two functional genes were retained (Liu *et al*, 2016), the short-beaked echidna retaining only three intact genes (Zhou *et al*, 2021), and the numbat having 11 pseudogenes out of 22 identified *Tas2rs* (Peel *et al*, 2022). These gene losses in myrmecophagous species might be due to their tongue better suited to capture prey rather than taste (*e.g.*, Casali *et al*, 2017) inducing relaxed selection on the associated genes. This was also suggested by Emerling *et al* (2017) for snakes which also lost *Tas2r* genes and for which the tongue with fewer numbers of taste buds seems to be used preferentially for vomeronasal sensation and not so much for taste perception.

Despite their higher number of gene losses, Garland (2018) found that myrmecophagous species retained systematically two bitter taste receptor genes, *Tas2r1* and *Tas2r4*, further highlighting the convergent evolution of myrmecophagous placental taste receptor gene repertoires. It has been found that these receptors are receptive to a broad range of bitter tastants in humans (Meyerhof *et al*, 2010). They are notably elicited by phenylalanine and leucin (Kohl *et al*, 2013) and two cyclic dipeptides having bitter tasting properties and found in ant venom (Lopez and Morgan, 1997). Therefore, retaining these two genes might have been convergently selected in myrmecophagous mammals due to their diet (Garland, 2018). Here, these two genes were both identified in the annotated sequences of *M. tridactyla*, *T. tetradactyla, O. afer*, *M. javanica*, *P. tricuspis*, *O. megalotis*, and *D. novemcinctus*.

➔ **Future analyses**

To infer numbers and rates of gene losses and duplications in placentals, a reconciliation approach could be conducted between a gene tree and species tree. This could be done with GeneRax, which uses a probabilistic approach and infers the maximum likelihood of a reconciled gene tree using the species tree and the multiple gene alignment (Morel *et al*, 2020). Using amino acid sequences instead of nucleotides should help clarify paralogue relationships. This analysis should help us to better understand the overall evolutionary history of *Tas2rs* in placentals. Synteny analyses could also be conducted, as *Tas2r* genes are usually clustered together in the genome, for instance by remapping identified sequences on genome assemblies and should give further insights on the evolutionary dynamic of this gene family.

To fully characterize and compare *Tas2r* gene repertoires between species, the aim will then be to identify functional and pseudogenized genes in the 230 selected genomes using HMM models that will be built from the annotated *Tas2r* sequences previously extracted. Pseudogenes could be identified and aligned using MACSE (Ranwez *et al*, 2011) as this program is specifically designed to align nucleotide sequences with respect to their codon sequence. Numbers of functional genes and pseudogenes per species could then be computed and compared between myrmecophagous species and their non-myrmecophagous sister-

species to highlight potential patterns of convergence among species gene repertoires and among myrmecophagous species to assess whether similar genomic adaptations occurred between the different species (*i.e*., similar gene losses and duplications). Additionally, patterns of shared inactivating mutations (*i.e*., splice-site mutations, frameshifts, premature stop codons) could also be investigated to uncover the mechanisms involved in the reduction of the *Tas2r* gene repertoires in myrmecophagous mammals.

To better understand the link between bitter taste receptor genes and dietary shifts, one could try to correlated the number of functional genes with characteristics of a species diet like the percentage of invertebrates, as it has been done for example for chitinase genes in placentals (Emerling *et al*, 2018; Janiak *et al*, 2018). Garland (2018) did not find a clear correlation between placental bitter taste gene repertoires (*i.e*., the number of functional genes) according to their diet (*i.e*., the percentage of invertebrates or plants in their diet) but rather highlighted a phylogenetic effect with sloths clustering with anteaters and armadillos. Considering the breadth of the diet could be one possibility to better assess the link between the size of taste gene repertoires and the diet as narrow diets have been linked with a reduction of *Tas2r* repertoires in placentals (Liu *et al*, 2016).

Finally, to fully understand the evolution of *Tas2r* gene repertoires in placentals, their expression should be studied as their evolution might be influenced by their pleiotropic effects. Indeed, *Tas2r* genes are found expressed in several organs of the digestive system suggesting extra-gustatory functions such as gut absorption initiation or hormone secretion (Wu *et al*, 2002; Taniguchi, 2004; Dyer *et al*, 2005), and even in the human skin where they might play a role in keratinocyte differentiation (Wölfle *et al*, 2015). Therefore, studying their expression in different organs of ant- and termite-eating mammals and closely related species should help decipher whether gene copies that are not lost are used for taste perception or other functions. This will help us understand the role and evolution of the different *Tas2r* gene copies in myrmecophagous species. Transcriptomic data are already available for several digestive (*i.e*., salivary glands, tongue, stomach, liver, intestine, pancreas, and spleen samples) and non-digestive organs of the southern tamandua (*Tamandua tetradactyla*), the Malayan pangolin (*Manis javanica*), and the nine-banded armadillo (*Dasypus novemcinctus*), and for 30 salivary glands of 23 placental species (Allio *et al*, 2023). Additional samples of several digestive and non-digestive organs of *T. tetradactyla*, *M. tridactyla*, and the pale-throated

sloth (*Bradypus tridactylus*), and salivary glands of the Linnaeus's two-toed sloth (*Choloepus didactylus*) have been collected during fieldwork in French Guiana and stored in RNA later. These samples could be used to complete this dataset after RNA sequencing and transcriptome assembly. The presence and expression levels of *Tas2r* transcripts could then be compared between organs of the same species and between myrmecophagous species and non-myrmecophagous closely related species.

Supplementary data

Table S I.1. Detailed information on the 230 placental genomes selected for the analysis of *Tas2r* genes.

Table S I. 1. Detailed information on the 230 placental genomes selected for the analysis of *Tas2r* genes.

Species	Taxonomy						Diet	Genome source			Genome statistics			Annotated TAS2Rs sequences	
	Class	Super order	Order	Family	Genus	Common name	Percentage invertebrates	Genome name	Genome source	Genbank genome accession number	Total length (bp)	N50 contig (bp)	N50 scf (bp)	Annotation file available	Number of total T2Rs sequences identified
Acinonyx jubatus	Mammalia	Laurasiatheria	Carnivora	Felidae	Acinonyx	Cheetah	0	aciJub1_HiC	DNAZOO	NA	2373338770	28241	144637309	yes	13
Acomys percivali	Mammalia	Euarchontoglires	Rodentia	Muridae	Acomys	Percival's spiny mouse	NA	mAcoPer2_REL_1905	GENBANK	GCA_907169655.1	2302049166	26477	126270630	no	NA
Aeorestes cinereus	Mammalia	Laurasiatheria	Chiroptera	Vespertilionidae	Aeorestes	Hoary bat	100	L.cinereus_Cryan_1219_p1.0_HiC	DNAZOO	NA	2143505952	120604	201349205	yes	13
Aepyceros melampus	Mammalia	Laurasiatheria	Artiodactyla	Bovidae	Aepyceros	Impala	0	IMP	GENBANK	GCA_006408695.1	2631303056	82459	344542	no	NA
Ailuropoda melanoleuca	Mammalia	Laurasiatheria	Carnivora	Ursidae	Ailuropoda	Giant panda	0	ASM200744v2	GENBANK	GCF_002007445.1	2444060653	127380	129245720	yes	9
Ailurus fulgens styani	Mammalia	Laurasiatheria	Carnivora	Ailuridae	Ailurus	Red panda	10	ASM200746v1_HiC	DNAZOO	NA	2343308739	43058	143796361	yes	13
Alces alces	Mammalia	Laurasiatheria	Artiodactyla	Cevidae	Alces	Eurasian elk	0	GSC_moose_1.0	GENBANK	GCA_007570765.1	2743728988	45091	4131188	no	NA
Anoura caudifer	Mammalia	Laurasiatheria	Chiroptera	Phyllostomidae	Anoura	Tailed tailless bat	30	AnoCau_v1_BIUU	GENBANK	GCA_004027475.1	2206589520	143417	185021	no	NA
Antilocapra americana	Mammalia	Laurasiatheria	Artiodactyla	Antilocapridae	Antilocapra	Pronghorn	0	AntAmePen_v2_BIUU_UCD	GENBANK	GCA_004027515.2	2955306661	61698	1884506	no	NA
Aotus nancymaae	Mammalia	Euarchontoglires	Primates	Aotidae	Aotus	Ma's night monkey	20	Anan_1.0	GENBANK	GCF_000952055.1	2926565220	28503	8280397	yes	0
Arctocephalus townsendi	Mammalia	Laurasiatheria	Carnivora	Otariidae	Arctocephalus	Guadalupe fur seal	30	Arctocephalus_townsendi_HiC	DNAZOO	NA	2372376283	61084	127156334	yes	7
Artibeus jamaicensis	Mammalia	Laurasiatheria	Chiroptera	Phyllostomidae	Artibeus	Jamaican fruit bat	10	WHU_Ajam_v2_HiC	DNAZOO	NA	2208330307	41756	148939558	yes	12
Arvicola amphibius	Mammalia	Euarchontoglires	Rodentia	Cricetidae	Arvicola	Eurasian water vole	0	mArvAmp1.2	GENBANK	GCF_903992535.2	2297766297	5392280	158924400	yes	0
Ateles hybridus	Mammalia	Euarchontoglires	Primates	Atlelidae	Ateles	Brown spider monkey	10	ORGONE_01	GENBANK	GCA_916098195.1	2643274663	50515269	50515269	no	NA
Axis porcinus	Mammalia	Laurasiatheria	Artiodactyla	Cervidae	Axis	Hog deer	0	ASM379854v1_HiC	DNAZOO	NA	2676213324	67167	77075487	yes	12
Babyrousa celebensis	Mammalia	Laurasiatheria	Artiodactyla	Suidae	Babyrousa	North sulawesi babirusa	10	Babyrousa_celebensis_HiC	DNAZOO	NA	2571159929	52662	11896847	yes	12
Balaenoptera ricei	Mammalia	Laurasiatheria	Artiodactyla	Balaenopteridae	Balaenoptera	Rice's whale	50	Balaenoptera_ricei_HiC	DNAZOO	NA	2378502053	71244	99560599	yes	3
Bassariscus sumichrasti	Mammalia	Laurasiatheria	Carnivora	Procyonidae	Bassariscus	Cacomistle	20	Bassariscus_sumichrasti_HiC	DNAZOO	NA	2546126934	44964	125116104	yes	10
Bradypus variegatus	Mammalia	Xenarthra	Pilosa	Bradypodidae	Bradypus	Brown-throated sloth	0	BraVar_v1_BIUU	GENBANK	GCA_004027775.1	1502475299	1900	1900	no	NA
Bradypus_tridactylus	Mammalia	Xenarthra	Pilosa	Bradypodidae	Bradypus	Pale-throated sloth	0	Bradypus_tridactylus_V3450_25_02_2020_final_genome	CONVERGEANT	NA	3244835702	684256	NA	yes	9
Cabassous unicinctus	Mammalia	Xenarthra	Cingulata	Dasypodidae	Cabassous	Southern naked-tailed armadillo	100	Cabassous_unicinctus_MVZ155190_Discovar	CONVERGEANT	NA	NA	NA	NA	no	NA
Callimico goeldii	Mammalia	Euarchontoglires	Primates	Cebidae	Callimico	Goeldi's monkey	40	Callimico_goeldii_HiC	DNAZOO	NA	2887766742	46551	11553634	yes	20
Callithrix jacchus	Mammalia	Euarchontoglires	Primates	Cebidae	Callithrix	White-tufted-ear marmoset	20	mCalJac1.mat	GENBANK	GCA_011078405.1	2811151840	8609028	146897247	no	NA
Callorhinus ursinus	Mammalia	Laurasiatheria	Carnivora	Otariidae	Callorhinus	Northern fur seal	40	ASM326570v1	GENBANK	GCF_003265705.1	2706852204	133024	31506801	yes	11
Camelus dromedarius	Mammalia	Laurasiatheria	Artiodactyla	Camelidae	Camelus	Dromedary	0	PRJNA234474_Ca_dromedarius_V1.0_HiC	DNAZOO	NA	2004792918	69067	72916538	yes	9
Canis lupus dingo	Mammalia	Laurasiatheria	Carnivora	Canidae	Canis	Dingo	0	ASM325472v1_HiC	DNAZOO	NA	2436463757	5382508	63865217	yes	17
Capreolus pygargus	Mammalia	Laurasiatheria	Artiodactyla	Cervidae	Capreolus	Eastern roe deer	0	ASM1292296v1	GENBANK	GCA_012922965.1	2607832873	80310	6067221	no	NA
Caracal caracal	Mammalia	Laurasiatheria	Carnivora	Felidae	Caracal	Caracal	0	CarCar1.0	GENBANK	GCA_016801355.1	2420801777	32916	2085423	no	NA
Carlito syrichta	Mammalia	Euarchontoglires	Primates	Tarsiidae	Carlito	Philippine tarsier	100	Tarsius_syrichta-2.0.1	GENBANK	GCF_000164805.1	3453847770	38165	401181	yes	12
Carollia perspicillata	Mammalia	Laurasiatheria	Chiroptera	Phyllostomidae	Carollia	Seba's short-tailed bat	0	CarPer_v1_BIUU_HiC	DNAZOO	NA	2732657905	10340	96450962	yes	10
Castor canadensis	Mammalia	Euarchontoglires	Rodentia	Castoridae	Castor	Canadian beaver	0	C.can_genome_v1.0_HiC	DNAZOO	NA	2527266565	167197	136673807	yes	14
Catagonus wagneri	Mammalia	Laurasiatheria	Artiodactyla	Tayassuidae	Catagonus	Chacoan peccary	0	CatWag_v2_BIUU_UCD	GENBANK	GCA_004024745.2	2640067814	65965	19204659	no	NA
Cavia porcellus	Mammalia	Euarchontoglires	Rodentia	Caviidae	Cavia	Domestic guinea pig	0	Cavpor3.0_HiC	DNAZOO	NA	2723543632	80475	82743159	yes	20
Cebuella pygmaea	Mammalia	Euarchontoglires	Primates	Cebidae	Cebuella	Pygmy marmoset	20	Cebuella_pygmaea_HiC	DNAZOO	NA	2909843497	57333	118608034	yes	19
Cebus imitator	Mammalia	Euarchontoglires	Primates	Cebidae	Cebus	Panamanian white-faced capuchin	20	Cebus_imitator-1.0	GENBANK	GCF_001604975.1	2717703182	41196	5274112	yes	16
Cephalorhynchus commersonii	Mammalia	Laurasiatheria	Artiodactyla	Delphinidae	Cephalorhynchus	Commerson's dolphin	20	Cephalorhynchus_commersonii_HiC	DNAZOO	NA	2347727969	92178	104003720	yes	3
Ceratotherium simum simum	Mammalia	Laurasiatheria	Perissodactyla	Rhinocerotidae	Ceratotherium	Southern white rhinoceros	0	CerSimSim1.0_HiC	DNAZOO	NA	2463469153	92633	66082376	yes	13
Cervus elaphus hippelaphus	Mammalia	Laurasiatheria	Artiodactyla	Cervidae	Cervus	Red deer	0	CerEla1.0	GENBANK	GCA_002197005.1	3438623608	7944	107358006	yes	0
Chaetophractus vellerosus	Mammalia	Xenarthra	Cingulata	Chlamyphoridae	Chaetophractus	Screaming hairy armadillo	50	ChaVel_v1_BIUU	GENBANK	GCA_004027955.1	5335596729	1606	1606	no	NA
Cheirogaleus medius	Mammalia	Euarchontoglires	Primates	Cheirogaleidae	Cheirogaleus	Lesser dwarf lemur	20	ASM808673v1	GENBANK	GCA_008086735.1	2121890802	34905	48318266	no	NA
Chinchilla lanigera	Mammalia	Euarchontoglires	Rodentia	Chinchillidae	Chinchilla	Long-tailed chinchilla	0	ChiLan1.0_HiC	DNAZOO	NA	2387037043	61014	74430473	yes	12
Chlamyphorus_truncatus	Mammalia	Xenarthra	Cingulata	Chlamyphoridae	Chlamyphorus	Pink fairy armadillo	100	Chlamyphorus_truncatus_CT1_7_12_2019_final_genome	CONVERGEANT	NA	3213424224	64202	NA	yes	6
Choloepus didactylus	Mammalia	Xenarthra	Pilosa	Megalonychidae	Choloepus	Linnaeus two-toed sloth	0	mChoDid1.pri	GENBANK	GCF_015220235.1	3214686105	20994632	513103	yes	8
Choloepus hoffmanni	Mammalia	Xenarthra	Pilosa	Megalonychidae	Choloepus	Two-toed sloth	0	C_hoffmanni-2.0.1_HiC	DNAZOO	NA	3293892468	64321	140950122	yes	7
Chrysochloris asiatica	Mammalia	Afrotheria	Afrosoricida	Chrysochloridae	Chrysochloris	Cape golden mole	100	ChrAsi1.0	GENBANK	GCF_000296735.1	4210093806	19632	13470186	yes	0
Chrysocyon brachyurus	Mammalia	Laurasiatheria	Carnivora	Canidae	Chrysocyon	Maned wolf	10	Chrysocyon_brachyurus_HiC	DNAZOO	NA	2338125846	91425	60785871	yes	17
Coendou prehensilis	Mammalia	Euarchontoglires	Rodentia	Erethizontidae	Coendou	Brazilian porcupine	0	Coendou_prehensilis_HiC	DNAZOO	NA	2849766862	32746	42412415	yes	15
Colobus angolensis	Mammalia	Euarchontoglires	Primates	Cercopithecidae	Colobus	Angolan colobus monkey	0	Cang.pa_1.0_HiC	DNAZOO	NA	2970213929	38355	147317940	yes	21
Condylura cristata	Mammalia	Laurasiatheria	Eulipotypha	Talpidae	Condylura	Star-nosed mole	80	ConCri1.0	GENBANK	GCF_000260355.1	1769662895	46163	55520359	yes	0
Connochaetes taurinus	Mammalia	Laurasiatheria	Artiodactyla	Bovidae	Connochaetes	Blue wildebeest	0	BWD_HiC	DNAZOO	NA	2650143527	46500	98098359	yes	13
Corynorhinus rafinesquii	Mammalia	Laurasiatheria	Chiroptera	Vespertilionidae	Corynorhinus	Rafinesque's big-eared bat	100	Corynorhinus_rafinesquii_HiC	DNAZOO	NA	2114175942	35465	144810274	yes	13
Cricetomys ansorgei	Mammalia	Euarchontoglires	Rodentia	Nesomyidae	Cricetomys	Southern giant pouched rat	40	Cricetomys_ansorgei_HiC	DNAZOO	NA	2453730532	13581387	49543905	yes	1
Cricetulus griseus	Mammalia	Euarchontoglires	Rodentia	Cricetidae	Cricetulus	Chinese hamster	0	CriGri-PICR_HiC	DNAZOO	NA	2369202408	1954687	284526441	yes	0
Crocuta crocuta	Mammalia	Laurasiatheria	Carnivora	Hyaenidae	Crocuta	Spotted hyaena	0	Crocuta_crocuta_HiC	DNAZOO	NA	2548291099	60727	101812163	yes	9
Cryptoprocta ferox	Mammalia	Laurasiatheria	Carnivora	Eupleridae	Cryptoprocta	Fossa	10	Cryptoprocta_ferox_HiC	DNAZOO	NA	2373379534	124286	107277967	yes	12
Ctenodactylus gundi	Mammalia	Euarchontoglires	Rodentia	Ctenodactylidae	Ctenodactylus	Northern gundi	0	CteGun_v1_BIUU	GENBANK	GCA_004027205.1	2322471743	218543	354548	no	NA
Cyclopes_didactylus	Mammalia	Xenarthra	Pilosa	Myrmecophagidae	Cyclopes	Silky anteater	100	Cyclopes_didactylus_M2300_15_01_2020_final_genome	CONVERGEANT	NA	3551270084	555204	NA	yes	4
Cynomys gunnisoni	Mammalia	Euarchontoglires	Rodentia	Sciuridae	Cynomys	Gunnison's prairie dog	0	ASM1131664v1	GENBANK	GCA_011316645.1	2674371627	687762	824613	no	NA
Cynopterus brachyotis	Mammalia	Euarchontoglires	Rodentia	Pteropodidae	Cynopterus	Lesser short-nosed fruit bat	0	ASM979314v1	GENBANK	GCA_009793145.1	1758935687	16627	251278	no	NA
Dasypus novemcinctus	Mammalia	Xenarthra	Cingulata	Dasypodidae	Dasypus	Nine-banded armadillo	100	Dasnov3.0	GENBANK	GCF_000208655.1	3631505655	26277	1687935	yes	9
Daubentonia madagascariensis	Mammalia	Euarchontoglires	Primates	Daubentoniidae	Daubentonia	Aye-aye	40	Daubentonia_madagascariensis	DNAZOO	NA	2433754680	215378	211484450	yes	14
Delphinapterus leucas	Mammalia	Laurasiatheria	Artiodactyla	Monodontidae	Delphinapterus	Beluga whale	40	ASM228892v2_HiC	DNAZOO	NA	2356565923	158270	107969763	yes	4
Desmodus rotundus	Mammalia	Laurasiatheria	Chiroptera	Phyllostomidae	Desmodus	Common vampire bat	0	ASM294091v2	GENBANK	GCF_002940915.1	2063791738	80250	26869735	no	NA
Dicerorhinus sumatrensis harrissoni	Mammalia	Laurasiatheria	Perissodactyla	Rhinocerotidae	Dicerorhinus	Sumatran rhinoceros	0	NRM_Dsumatrensis_v1.fasta	GENBANK	GCA_014189135.1	2442438657	70892	54482381	no	NA
Diceros bicornis minor	Mammalia	Laurasiatheria	Perissodactyla	Rhinocerotidae	Diceros	African black rhinoceros	0	Diceros_bicornis_HiC	DNAZOO	NA	2604587929	87018	59591286	yes	14
Dipodomys ordii	Mammalia	Euarchontoglires	Rodentia	Heteromyidae	Dipodomys	Ord's kangaroo rat	20	Dord_2.0	GENBANK	GCF_000151885.1	2236368823	48087	11931245	yes	0
Dugong dugon	Mammalia	Afrotheria	Sirenia	Dugongidae	Dugong	Dugong	0	Dugong_dugon_HiC	DNAZOO	NA	3101568320	63948	118739814	yes	9
Echinops telfairi	Mammalia	Afrotheria	Afrosoricida	Tenrecidae	Echinops	Small Madagascar hedgehog	60	ASM31398v2	GENBANK	GCF_000313985.2	2947103070	20425	54422506	yes	8
Eidolon dupreanum	Mammalia	Laurasiatheria	Chiroptera	Pteropodidae	Eidolon	Madagascan fruit bat	0	Eidolon_dupreanum_HiC	DNAZOO	NA	2294643016	109242	101563129	yes	11
Elaphurus davidianus	Mammalia	Laurasiatheria	Artiodactyla	Cervidae	Elaphurus	Pere David's deer	0	Milu1.0	GENBANK	GCA_002443075.1	2584693296	59950	2844142	no	NA
Elephantulus edwardii	Mammalia	Afrotheria	Macroscelidea	Macroscelididae	Elephantulus	Cape elephant shrew	100	EleEdw1.0	GENBANK	GCF_000299155.1	3843982861	24219	15011382	yes	0
Elephas maximus	Mammalia	Afrotheria	Proboscidea	Elephantidae	Elephas	Asian elephant	0	Elephas_maximus_HiC	DNAZOO	NA	3212596588	57611	95956831	yes	18
Eonycteris spelaea	Mammalia	Laurasiatheria	Chiroptera	Pteropodidae	Eonycteris	Lesser dawn bat	0	Espe.v1	GENBANK	GCA_003508835.1	1966861576	8002591	13454942	no	NA
Eptesicus fuscus	Mammalia	Laurasiatheria	Chiroptera	Vespertilionidae	Eptesicus	Big brown bat	100	EptFus1.0_HiC	DNAZOO	NA	2017582869	21390	102216854	yes	10
Equus zebra	Mammalia	Laurasiatheria	Perissodactyla	Equidae	Equus	Mountain zebra	0	Equus_zebra_HiC	DNAZOO	NA	2505679032	182799	150600390	yes	20
Erethizon dorsatum	Mammalia	Euarchontoglires	Rodentia	Erethizontidae	Erethizon	North american porcupine	0	GSC_porc_1.0_HiC	DNAZOO	NA	2436132597	118744	126287952	yes	16
Erignathus barbatus	Mammalia	Laurasiatheria	Carnivora	Phocidae	Erignathus	Bearded seal	60	Erignathus_barbatus_HiC	DNAZOO	NA	2369341062	46151	133069726	yes	7
Erinaceus europaeus	Mammalia	Laurasiatheria	Eulipotypha	Erinaceidae	Erinaceus	Western European hedgehog	80	EriEur2.0	GENBANK	GCF_000296755.1	2715703478	21359	3264618	yes	5
Eschrichtius robustus	Mammalia	Laurasiatheria	Artiodactyla	Eschrichtiidae	Eschrichtius	Gray whale	90	Eschrichtius_robustus_HiC	DNAZOO	NA	2348811279	67147	103007579	yes	6
Eubalaena australis	Mammalia	Laurasiatheria	Artiodactyla	Balaenidae	Eubalaena	Southern right whale	100	RWref_HiC	DNAZOO	NA	2316908615	27668	112042483	no	NA
Eulemur flavifrons	Mammalia	Euarchontoglires	Primates	Lemuridae	Eulemur	Blue-eyed black lemur	0	Eflavifronsk33QCA_HiC	DNAZOO	NA	2119919085	16270	156661540	yes	16
Eumetopias jubatus	Mammalia	Laurasiatheria	Carnivora	Otariidae	Eumetopias	Steller sea lion	0	ASM402803v1	GENBANK	GCF_004028035.1	2418246527	242372	14018600	yes	10
Euphractus_sexcinctus	Mammalia	Xenarthra	Cingulata	Dasypodidae	Euphractus	Six-banded armadillo	50	Euphractus_sexcinctus_ESE1_3_11_2021_final_genome	CONVERGEANT	NA	3477096506	853561	NA	yes	5
Felis nigripes	Mammalia	Laurasiatheria	Carnivora	Felidae	Felis	Black-footed cat	0	Felis_nigripes_HiC	DNAZOO	NA	2454753549	50735	139656774	yes	11
Fukomys damarensis	Mammalia	Euarchontoglires	Rodentia	Bathyergidae	Fukomys	Damaraland mole-rat	0	DMR_v1.0_HiC	DNAZOO	NA	2334358650	44756	62586000	yes	8

Table S I. 1. Detailed information on the 230 placental genomes selected for the analysis of *Tas2r* genes.

Species	Taxonomy						Diet	Genome source			Genome statistics			Annotated TAS2Rs sequences	
	Class	Super order	Order	Family	Genus	Common name	Percentage invertebrates	Genome name	Genome source	Genbank genome accession number	Total length (bp)	N50 contig (bp)	N50 scf (bp)	Annotation file available	Number of total T2Rs sequences identified
Galemys pyrenaicus	Mammalia	Laurasiatheria	Eulipotyphla	Talpidae	Galemys	Pyrenean desman	90	Gpyr_1.0	GENBANK	GCA_019455555.1	1828347170	64544	8503682	yes	0
Galeopterus variegatus	Mammalia	Euarchontoglires	Dermoptera	Cynocephalidae	Galeopterus	Sunda flying lemur	0	GalVar_v2_BIUU_UCD	GENBANK	GCA_004027255.2	3349451543	34666	7885395	no	NA
Giraffa camelopardalis	Mammalia	Laurasiatheria	Artiodactyla	Giraffidae	Giraffa	Giraffe	0	ASM165123v1_HiC	DNAZOO	NA	2713320025	22965	154205505	yes	9
Glaucomys volans	Mammalia	Euarchontoglires	Rodentia	Sciuridae	Glaucomys	Southern flying squirrel	0	ASM2066280v1	GENBANK	GCA_020662805.1	2582196772	75533	452493	no	NA
Globicephala melas	Mammalia	Laurasiatheria	Artiodactyla	Delphinidae	Globicephala	Long-finned pilot whale	90	ASM654740v1_HiC	DNAZOO	NA	2333877532	332801	106927605	yes	2
Gorilla gorilla gorilla	Mammalia	Euarchontoglires	Primates	Hominidae	Gorilla	Western lowland gorilla	0	Kamilah_GGO_v0	GENBANK	GCF_008122165.1	3044855802	9522971	26116462	no	NA
Halichoerus grypus	Mammalia	Laurasiatheria	Carnivora	Phocidae	Halichoerus	Gray seal	20	Halichoerus_grypus_HiC	DNAZOO	NA	2413496209	62032	141565142	yes	9
Helarctos malayanus	Mammalia	Laurasiatheria	Carnivora	Ursidae	Helarctos	Malayan sun bear	50	Helarctos_malayanus_HiC	DNAZOO	NA	2486327345	73706	59704986	yes	14
Helogale parvula	Mammalia	Laurasiatheria	Carnivora	Herpestidae	Helogale	Dwarf mongoose	70	HelPar_v1_BIUU	GENBANK	GCA_004023845.1	2392471390	113567	179119	no	NA
Heterocephalus glaber	Mammalia	Euarchontoglires	Rodentia	Bathyergidae	Heterocephalus	Naked mole-rat	0	HetGla_female_1.0_HiC	DNAZOO	NA	2617997130	47773	100148522	yes	13
Hippopotamus amphibius	Mammalia	Laurasiatheria	Artiodactyla	Hippopotamidae	Hippopotamus	Hippopotamus	0	HipAmp_v2_BIUU_UCD	GENBANK	GCA_004027065.2	2733493772	76609	4444377	no	NA
Hipposideros armiger	Mammalia	Laurasiatheria	Chiroptera	Hipposideridae	Hipposideros	Great roundleaf bat	100	ASM189008v1	GENBANK	GCF_001890085.1	2236564388	39863	2328177	yes	8
Homo sapiens	Mammalia	Euarchontoglires	Primates	Hominidae	Homo	Human	10	CHM13 T2T v1.1	GENBANK	GCA_009914755.3	3054815472	154259566	154259566	no	NA
Hyaena hyaena	Mammalia	Laurasiatheria	Carnivora	Hyaenidae	Hyaena	Striped hyaena	10	ASM300989v1	GENBANK	GCF_003009895.1	2374716107	311202	2001327	yes	8
Hydrochoerus hydrochaeris	Mammalia	Euarchontoglires	Rodentia	Hydrochaeridae	Hydrochoerus	Capybara	0	Hydrochoerus_hydrochaeris_HiC	DNAZOO	NA	2950839073	78808	71052467	yes	17
Hydropotes inermis	Mammalia	Laurasiatheria	Artiodactyla	Cervidae	Hydropotes	Chinese water deer	0	ASM2022607v1	GENBANK	GCA_020226075.1	2540444434	131414	74967728	no	NA
Hylobates agilis	Mammalia	Euarchontoglires	Primates	Hylobatidae	Hylobates	Agile gibbon	10	Hylobates_agilis_HiC	DNAZOO	NA	2941325771	35590	93981261	yes	16
Hystrix brachyura	Mammalia	Euarchontoglires	Rodentia	Hystricidae	Hystrix	Malayan porcupine	0	DSBC_Hbra_1.0	GENBANK	GCA_016801275.1	2257332368	29439	2980431	no	NA
Ictidomys tridecemlineatus	Mammalia	Euarchontoglires	Rodentia	Sciuridae	Ictidomys	Thirteen-lined ground squirrel	20	HiC_Itri_2	GENBANK	GCF_016881025.1	2478949113	44127	193221680	yes	0
Jaculus jaculus	Mammalia	Euarchontoglires	Rodentia	Dipodidae	Jaculus	Lesser Egyptian jerboa	0	mJacJac1.mat.Y.cur	GENBANK	GCF_020740685.1	2863848715	22104564	158244790	yes	0
Kobus leche leche	Mammalia	Laurasiatheria	Artiodactyla	Bovidae	Kobus	Lechwe	0	Klec_v1.0	GENBANK	GCA_014926565.1	2771252237	66923	3233651	no	NA
Kogia breviceps	Mammalia	Laurasiatheria	Artiodactyla	Physeteridae	Kogia	Pygmy sperm whale	100	KogBre_v1_BIUU_HiC	DNAZOO	NA	2782744659	26116	69986930	yes	3
Lama glama chaku	Mammalia	Laurasiatheria	Artiodactyla	Camelidae	Lama	Llama	0	Lama_glama_HiC	DNAZOO	NA	2351763638	93318	57490894	yes	11
Lemur catta	Mammalia	Euarchontoglires	Primates	Lemuridae	Lemur	Ring-tailed lemur	0	mLemCat1.pri	GENBANK	GCA_020740605.1	2245584463	32529614	102162704	yes	16
Leontopithecus rosalia	Mammalia	Euarchontoglires	Primates	Cebidae	Leontopithecus	Golden lion tamarin	50	Leontopithecus_rosalia_HiC	DNAZOO	NA	2874536464	55757	12128971	yes	18
Leopardus geoffroyi	Mammalia	Laurasiatheria	Carnivora	Felidae	Leopardus	Geoffroy's cat	0	O.geoffroyi_Oge1_pat1.0	GENBANK	GCA_018350155.1	2426370816	104474415	152606360	yes	11
Leptonychotes weddellii	Mammalia	Laurasiatheria	Carnivora	Phocidae	Leptonychotes	Weddell seal	40	Leptonychotes_weddellii_HiC	DNAZOO	NA	2407118189	46840	131171632	yes	8
Lepus timidus	Mammalia	Euarchontoglires	Lagomorpha	Leporidae	Lepus	Mountain hare	0	CIBIO-ISEM_LeTim_1.1	GENBANK	GCA_009760805.1	2703257108	18458	116271063	no	NA
Lipotes vexillifer	Mammalia	Laurasiatheria	Artiodactyla	Lipotidae	Lipotes	Yangtze river dolphin	0	Lipotes_vexillifer_v1	GENBANK	GCF_000442215.1	2429195737	31902	2419148	yes	0
Lophiomys imhausi	Mammalia	Euarchontoglires	Rodentia	Muridae	Lophiomys	Crested rat	0	mLopImh1.curated_primary_1811	GENBANK	GCA_907164525.1	2920431132	35791	6438253	no	NA
Loxodonta africana	Mammalia	Afrotheria	Proboscidea	Elephantidae	Loxodonta	African savannah elephant	0	Loxafr3.0_HiC	DNAZOO	NA	3196344422	69012	117617189	yes	14
Lutra lutra	Mammalia	Laurasiatheria	Carnivora	Mustelidae	Lutra	Eurasian otter	30	mLutLut1.pri.cur.20190822	DNAZOO	NA	2438442342	30403456	149004807	yes	14
Lycaon pictus	Mammalia	Laurasiatheria	Carnivora	Canidae	Lycaon	African wild dog	0	sis2-181106_HiC	DNAZOO	NA	2352061016	100350	62683413	yes	13
Lynx canadensis	Mammalia	Laurasiatheria	Carnivora	Felidae	Lynx	Canada lynx	0	mLynCan4_v1.p	GENBANK	GCF_007474595.1	2408883772	7503561	146106016	yes	10
Macaca fuscata fuscata	Mammalia	Euarchontoglires	Primates	Cercopithecidae	Macaca	Japanese macaque	10	Macaca_fuscata_HiC	DNAZOO	NA	2843076980	90024	149352196	yes	17
Macroglossus sobrinus	Mammalia	Laurasiatheria	Chiroptera	Pteropodidae	Macroglossus	Long-tongued fruit bat	0	MacSob_v1_BIUU	GENBANK	GCA_004027375.1	1897644983	338389	453401	no	NA
Manis crassicaudata	Mammalia	Laurasiatheria	Pholidota	Manidae	Manis	India pangolin	100	DSBC_Mcra_1.0	GENBANK	GCA_016801295.1	2124325914	7447	14162	no	NA
Manis javanica	Mammalia	Laurasiatheria	Pholidota	Manidae	Manis	Sunda pangolin	100	ManJav1.0_HiC	DNAZOO	NA	2555637331	16350	131848799	yes	3
Manis pentadactyla	Mammalia	Laurasiatheria	Pholidota	Manidae	Manis	Chinese pangolin	100	M_pentadactyla-1.1.1_HiC	DNAZOO	NA	2215491672	20721	111940472	yes	4
Marmota marmota marmota	Mammalia	Euarchontoglires	Rodentia	Sciuridae	Marmota	Alpine marmot	0	marMar2.1	GENBANK	GCF_001458135.1	2510587379	66492	31340621	yes	0
Martes martes	Mammalia	Laurasiatheria	Carnivora	Mustelidae	Martes	Pine marten	0	mmar.min_150.pseudohap2.1_HiC	DNAZOO	NA	2425306698	327623	144638580	yes	16
Mastacomys fuscus	Mammalia	Euarchontoglires	Rodentia	Muridae	Mastacomys	Borad-toothed rat	20	Mastacomys_fuscus_wtdbg2_polished_HiC	DNAZOO	NA	2327498967	371733	104439433	yes	1
Megaptera novaeangliae	Mammalia	Laurasiatheria	Artiodactyla	Balaenopteridae	Megaptera	Humpback whale	80	Megaptera_novaeangliae_HiC	DNAZOO	NA	2540677902	55713	94294397	yes	4
Meriones unguiculatus	Mammalia	Euarchontoglires	Rodentia	Muridae	Meriones	Mongolian gerbil	20	ASM813125v1_HiC	DNAZOO	NA	2543403711	34209	116873421	yes	1
Mesoplodon europaeus	Mammalia	Laurasiatheria	Artiodactyla	Ziphiidae	Mesoplodon	Gervai's beaked whale	80	Mesoplodon_europaeus_HiC	DNAZOO	NA	2399299491	44972	90973712	yes	4
Microcebus murinus	Mammalia	Euarchontoglires	Primates	Cheirogaleidae	Microcebus	Gray mouse lemur	20	Mmur_3.0_HiC	DNAZOO	NA	2459610902	201416	109421721	yes	18
Mirounga angustirostris	Mammalia	Laurasiatheria	Carnivora	Phocidae	Mirounga	Northern elephant seal	60	Mirounga_angustirostris_HiC	DNAZOO	NA	2366206800	76189	139676048	yes	5
Molossus molossus	Mammalia	Laurasiatheria	Chiroptera	Molossidae	Molossus	Palla's mastiff bat	100	mMolMol1.p	GENBANK	GCF_014108415.1	2315568481	22174888	110665204	yes	12
Monodon monoceros	Mammalia	Laurasiatheria	Artiodactyla	Monodontidae	Monodon	Narwhal	40	NGI_Narwhal_1	GENBANK	GCF_005190385.1	2355574979	255327	107566389	yes	7
Mormoops blainvillei	Mammalia	Laurasiatheria	Chiroptera	Mormoopidae	Mormoops	Antillean ghost-faced bat	100	MorMeg_v1_BIUU	GENBANK	GCA_004026545.1	2111750309	142682	156292	no	NA
Moschus berezovskii	Mammalia	Laurasiatheria	Artiodactyla	Moschidae	Moschus	Forest musk deer	0	ls35.final.genome_HiC	DNAZOO	NA	2728698671	22671	102122392	yes	12
Mungos mungo	Mammalia	Laurasiatheria	Carnivora	Herpestidae	Mungos	Banded mongoose	80	Mungos_mungo_HiC	DNAZOO	NA	2494883117	60614	128516143	yes	11
Muntiacus muntjak	Mammalia	Laurasiatheria	Artiodactyla	Cervidae	Muntiacus	Indian muntjac	0	CMJ_HiC	DNAZOO	NA	2706609549	10910	705141453	yes	14
Mustela nigripes	Mammalia	Laurasiatheria	Carnivora	Mustelidae	Mustela	Black-footed ferret	0	musNig1_HiC	DNAZOO	NA	2498707582	148635	145433501	yes	8
Myocastor coypus	Mammalia	Euarchontoglires	Rodentia	Myocastoridae	Myocastor	Coypus	0	Myocastor_coypus_QM1153_Masurca.scf_HiC	DNAZOO	NA	2632872920	3975994	138704734	no	NA
Myotis septentrionalis	Mammalia	Laurasiatheria	Chiroptera	Vespertilionidae	Myotis	Northern long-eared bat	100	myse_ont_racon_pilon_HiC	DNAZOO	NA	1977996141	203380	96689946	yes	14
Myrmecophaga tridactyla	Mammalia	Xenarthra	Pilosa	Myrmecophagidae	Myrmecophaga	Giant anteater	100	Myrmecophaga_tridactyla_M3023_24_10_19_final_genome_HiC	DNAZOO	NA	3136179797	1102000	111305850	yes	6
Nanger dama ruficollis	Mammalia	Laurasiatheria	Artiodactyla	Bovidae	Nanger	Dama gazelle	0	gazelle.1_HiC	DNAZOO	NA	3013446760	270108	156272796	yes	12
Nannospalax galili	Mammalia	Euarchontoglires	Rodentia	Spalacidae	Nannospalax	Upper Galilee mountains blind mole rat	0	S.galili_v1.0	GENBANK	GCF_000622305.1	3061408210	30353	3618479	yes	0
Nasua narica	Mammalia	Laurasiatheria	Carnivora	Procyonidae	Nasua	White-nosed coati	10	Nasua_narica_HiC	DNAZOO	NA	2610175269	64559	119764535	yes	13
Neofelis nebulosa	Mammalia	Laurasiatheria	Carnivora	Felidae	Neofelis	Clouded leopard	0	Neofelis_nebulosa_HiC	DNAZOO	NA	2416186965	76417	147111411	yes	17
Neomonachus schauinslandi	Mammalia	Laurasiatheria	Carnivora	Phocidae	Neomonachus	Hawaiian monk seal	30	EXP_REFINEFINAL1_bppAdjust_cmap_10X_BNG_fasta_NGScontigs_HYBRID_SCAFFOLD_NCBI_HiC	DNAZOO	NA	2364932810	185965	149571140	yes	8
Neophocaena asiaeorientalis asiaeorientalis	Mammalia	Laurasiatheria	Artiodactyla	Phocoenidae	Neophocaena	Ynagtze finless porpoise	40	Neophocaena_asiaeorientalis_V1	GENBANK	GCF_003031525.2	2284609912	86003	6341296	yes	6
Neosciurus carolinensis	Mammalia	Euarchontoglires	Rodentia	Sciuridae	Neosciurus	Eastern gray squirrel	0	mSciCar1.2	GENBANK	GCA_902686445.2	2815397268	13975867	148229995	yes	0
Noctilio leporinus	Mammalia	Laurasiatheria	Chiroptera	Noctilionidae	Noctilio	Greater bulldog bat	50	NocLep_v1_BIUU	GENBANK	GCA_004026585.1	2098501394	135651	191494	no	NA
Nomascus leucogenys	Mammalia	Euarchontoglires	Primates	Hylobatidae	Nomascus	Northern white-cheeked gibbon	10	Nleu_3.0	GENBANK	GCF_000146795.2	2962060179	35148	52956880	no	NA
Nyctereutes procyonoides	Mammalia	Laurasiatheria	Carnivora	Canidae	Nyctereutes	Raccoon dog	20	NYPRO_anot_genome	GENBANK	GCA_905146905.1	2387080870	35077230	53959811	yes	13
Ochotona princeps	Mammalia	Euarchontoglires	Lagomorpha	Ochotonidae	Ochotona	American pika	0	OchPri4.0	GENBANK	GCF_014633375.1	2231476247	42119	75838078	yes	9
Octodon degus	Mammalia	Euarchontoglires	Rodentia	Octodontidae	Octodon	Degu	0	OctDeg1.0	GENBANK	GCF_000260255.1	2995872505	19847	12091372	yes	0
Odobenus rosmarus divergens	Mammalia	Laurasiatheria	Carnivora	Odobenidae	Odobenus	Walrus	80	Oros_1.0_HiC	DNAZOO	NA	2401122044	89829	154349455	yes	7
Odocoileus hemionus hemionus	Mammalia	Laurasiatheria	Artiodactyla	Cervidae	Odocoileus	Mule deer	0	Odocoileurs_hemionus_Hic	DNAZOO	NA	2609372263	28571243	72141738	yes	13
Okapia johnstoni	Mammalia	Laurasiatheria	Artiodactyla	Giraffidae	Okapia	Okapi	0	ASM166083v1_HiC	DNAZOO	NA	2890536570	12529	97427823	yes	10
Orcinus orca	Mammalia	Laurasiatheria	Artiodactyla	Delphinidae	Orcinus	Orca	25	Oorc_1.1_HiC	DNAZOO	NA	2373058624	70204	110405485	yes	2
Orycteropus afer afer	Mammalia	Afrotheria	Tubulidentata	Orycteropodidae	Orycteropus	Aardvark	100	OryAfe1.0_HiC	DNAZOO	NA	4423506777	17647	644001617	yes	12
Oryctolagus cuniculus	Mammalia	Euarchontoglires	Lagomorpha	Leporidae	Oryctolagus	European rabbit	0	OryCun2.0_HiC	DNAZOO	NA	2672010907	64634	135079528	yes	10
Oryx dammah	Mammalia	Laurasiatheria	Artiodactyla	Bovidae	Oryx	Scimitar-horned oryx	0	oryx.1_HiC	DNAZOO	NA	2720101635	373003	100398400	yes	13
Otocyon megalotis	Mammalia	Laurasiatheria	Carnivora	Canidae	Otocyon	Bat-eared fox	50	Otocyon_megalotis_TS305_17_09_2019_HiC	CONVERGEANT	NA	2377998532	617182	68620662	yes	12
Otolemur crassicaudatus	Mammalia	Euarchontoglires	Primates	Galagidae	Otolemur	Brown greater galago	20	Otolemur_crassicaudatus_HiC	DNAZOO	NA	2520588429	73591	91872397	yes	20
Ovis aries	Mammalia	Laurasiatheria	Artiodactyla	Bovidae	Ovis	Sheep	0	Oar_rambouillet_v1.0	GENBANK	GCF_002742125.1	2869897780	2572683	107697089	no	NA
Pan paniscus	Mammalia	Euarchontoglires	Primates	Hominidae	Pan	Bonobo	10	Mhudiblu_PPA_v0	GENBANK	GCF_013052645.1	3051884774	16579680	68246502	no	NA
Panthera pardus	Mammalia	Laurasiatheria	Carnivora	Felidae	Panthera	Leopard	0	PanPar1.0_HiC	DNAZOO	NA	2578248701	20986	155751443	yes	12
Perognathus longimembris pacificus	Mammalia	Euarchontoglires	Rodentia	Heteromyidae	Perognathus	Pacific pocket mouse	10	PPM_HiRise_rh_HiC	DNAZOO	NA	2212099196	7389774	72679016	yes	6
Peromyscus eremicus	Mammalia	Euarchontoglires	Rodentia	Cricetidae	Peromyscus	Cactus mouse	50	Peer2.0.1_fasta	DNAZOO	NA	2737829688	20445	122758979	yes	1
Phacochoerus africanus	Mammalia	Laurasiatheria	Artiodactyla	Suidae	Phacochoerus	Common warthog	0	ROS_Pafr_v1	GENBANK	GCA_016906955.1	2435083091	10602503	141887063	yes	14
Phataginus tricuspis	Mammalia	Laurasiatheria	Pholidota	Manidae	Phataginus	Tree pangolin	100	Jaziri_pseudohap2_scaffolds_HiC	DNAZOO	NA	2473187469	112908	46351422	yes	5

Table S I. 1. Detailed information on the 230 placental genomes selected for the analysis of *Tas2r* genes.

Species	Taxonomy						Diet	Genome source			Genome statistics			Annotated TAS2Rs sequences	
	Class	Super order	Order	Family	Genus	Common name	Percentage invertebrates	Genome name	Genome source	Genbank genome accession number	Total length (bp)	N50 contig (bp)	N50 scf (bp)	Annotation file available	Number of total T2Rs sequences identified
Phoca vitulina	Mammalia	Laurasiatheria	Carnivora	Phocidae	Phoca	Harbor seal	20	GSC_HSeal_1.0_HiC	DNAZOO	NA	2363280482	283887	152438930	yes	10
Phocoena phocoena	Mammalia	Laurasiatheria	Carnivora	Phocoenidae	Phocoena	Harbor porpoise	0	Phocoena_phocoena_HiC	DNAZOO	NA	2467183617	58076	97795164	yes	5
Phyllostomus discolor	Mammalia	Laurasiatheria	Chiroptera	Phyllostomidae	Phyllostomus	Pale spear-nosed bat	30	mPhyDis1.pri.v3	GENBANK	GCF_004126475.2	2108832841	6892556	171742863	yes	6
Physeter catodon	Mammalia	Laurasiatheria	Artiodactyla	Physeteridae	Physeter	Sperm whale	100	ASM283717v2	GENBANK	GCF_002837175.2	2512132974	42542	122182240	yes	7
Pipistrellus pipistrellus	Mammalia	Laurasiatheria	Chiroptera	Vespertilionidae	Pipistrellus	Common pipistrelle	100	mPipPip1.1	GENBANK	GCA_903992545.1	1763422308	4446752	94929986	no	NA
Pithecia pithecia	Mammalia	Euarchontoglires	Primates	Pitheciidae	Pithecia	White-faced saki	0	Pithecia_pithecia_HiC	DNAZOO	NA	3051948709	53327	103980617	no	NA
Pongo abelii	Mammalia	Euarchontoglires	Primates	Hominidae	Pongo	Sumatran orangutan	10	Susie_PABv2	GENBANK	GCF_002880775.1	3065035716	11074009	98475126	yes	21
Potos flavus	Mammalia	Laurasiatheria	Carnivora	Procyonidae	Potos	Kinkajou	10	Potos_flavus_HiC	DNAZOO	NA	2465741797	65113	121721779	yes	12
Priodontes maximus	Mammalia	Xenarthra	Cingulata	Chlamyphoridae	Priodontes	Giant armadillo	90	Priodontes_maximus_M844_28_01_2020_final_genome	CONVERGEANT	NA	4088061607	184893	no	yes	8
Prionailurus bengalensis	Mammalia	Laurasiatheria	Carnivora	Felidae	Prionailurus	Leopard cat	0	PriBen1.0.updated_2_HiC	DNAZOO	NA	2490745690	16121	147495894	yes	14
Procavia capensis	Mammalia	Afrotheria	Hyracoidea	Procaviidae	Procavia	Rock hyrax	0	Pcap_2.0_HiC	DNAZOO	NA	3605540188	35449	133724184	yes	14
Procyon lotor	Mammalia	Laurasiatheria	Carnivora	Procyonidae	Procyon	Common raccoon	40	pl-1k.fasta	DNAZOO	NA	2525715526	34230	114539748	yes	12
Prolemur simus	Mammalia	Euarchontoglires	Primates	Lemuridae	Prolemur	Greater bamboo lemur	0	Prosim_1.0	GENBANK	GCA_003258685.1	2411593676	47757	2710671	no	NA
Propithecus coquereli	Mammalia	Euarchontoglires	Primates	Indriidae	Propithecus	Coquerel's sifaka	0	Pcoq_1.0_HiC	DNAZOO	NA	2796124105	28031	148766566	yes	19
Proteles cristata cristata	Mammalia	Laurasiatheria	Carnivora	Hyaenidae	Proteles	Southern aardwolf	100	Proteles_cristatus_TS307_19_07_2019	GENBANK	GCA_017311185.1	2388965834	1258344	1308801	yes	6
Przewalskium albirostris	Mammalia	Laurasiatheria	Artiodactyla	Cervidae	Przewalskium	White-lipped deer	0	WLD	GENBANK	GCA_006408465.1	2692225130	39627	3769372	no	NA
Pteropus vampyrus	Mammalia	Laurasiatheria	Chiroptera	Pteropodidae	Pteropus	Large flying fox	0	Pvam_2.0_HiC	DNAZOO	NA	2198965418	21830	123279187	yes	9
Puma concolor	Mammalia	Laurasiatheria	Carnivora	Felidae	Puma	Cougar	0	PumCon1.0_HiC	DNAZOO	NA	2433007005	27164	148638334	yes	11
Rangifer tarandus	Mammalia	Laurasiatheria	Artiodactyla	Cervidae	Rangifer	Reindeer	0	RanTarSib_v1_BIUU_HiC	DNAZOO	NA	2912079834	77590	59320115	no	NA
Rhinoceros unicornis	Mammalia	Laurasiatheria	Perissodactyla	Rhinocerotidae	Rhinoceros	Indian rhinoceros	0	Rhinoceros_unicornis_HiC	DNAZOO	NA	2632041242	105810	56870999	yes	18
Rhinolophus sinicus	Mammalia	Laurasiatheria	Chiroptera	Rhinolophidae	Rhinolophus	Chinese rufous horseshoe bat	80	mRhiSin1.pri.cur	GENBANK	GCA_020740635.1	2338923724	34562834	185549907	no	NA
Rhizomys pruinosus	Mammalia	Euarchontoglires	Rodentia	Spalacidae	Rhizomys	Hoary bamboo rat	0	RhiPru_1.0	GENBANK	GCA_009823505.1	3711971808	103291	2203772	no	NA
Rousettus madagascariensis	Mammalia	Laurasiatheria	Chiroptera	Pteropodidae	Rousettus	Madagascan rousette	0	Rousettus_madagascariensis_HiC	DNAZOO	NA	2344105138	63903	85834856	yes	12
Saimiri boliviensis boliviensis	Mammalia	Euarchontoglires	Primates	Cebidae	Saimiri	Squirrel monkey	20	saiBolDis_HiC	DNAZOO	NA	2855086604	46512	110640503	yes	19
Sapajus apella	Mammalia	Euarchontoglires	Primates	Cebidae	Sapajus	Tufted capuchin	20	GSC_monkey_1.0	GENBANK	GCF_009761245.1	2729201088	144042	23742480	yes	17
Sciurus vulgaris	Mammalia	Euarchontoglires	Rodentia	Sciuridae	Sciurus	Red squirrel	0	mSciVul1.PB.asm1.purge2.scaff2_HiC	DNAZOO	NA	2880406725	11463051	153870512	yes	11
Sigmodon hispidus	Mammalia	Euarchontoglires	Rodentia	Cricetidae	Sigmodon	Hispid cotton rat	50	SigHis_v1_BIUU	GENBANK	GCA_004025045.1	2730600022	67983	101373	no	NA
Smutsia gigantea	Mammalia	Laurasiatheria	Pholidota	Manidae	Smutsia	Giant pangolin	100	Smutsia_gigantea_CAM011_21_11_2019_final_genome	CONVERGEANT	NA	2463847862	227038	no	yes	3
Solenodon paradoxus	Mammalia	Laurasiatheria	Eulipotyphla	Solenodontidae	Solenodon	Hispaniolan solenodon	30	SolPar_v1_BIUU	GENBANK	GCA_004363575.1	2109877870	236847	407682	no	NA
Sorex araneus	Mammalia	Laurasiatheria	Eulipotyphla	Soricidae	Sorex	European shrew	70	SorAra2.0	GENBANK	GCF_000181275.1	2423158183	22623	22794405	no	NA
Spermophilus dauricus	Mammalia	Euarchontoglires	Rodentia	Sciuridae	Spermophilus	Daurian ground squirrel	20	ASM240643v1	GENBANK	GCA_002406435.1	3106271744	34849	1761345	no	NA
Spilogale interrupta	Mammalia	Laurasiatheria	Carnivora	Mephitidae	Spilogale	Plains spotted skunk	30	polished_2_HiC	DNAZOO	NA	2367379711	16288691	84003853	yes	11
Sturnira hondurensis	Mammalia	Laurasiatheria	Chiroptera	Phyllostomidae	Sturnira	Honduran yellow-shouldered bat	0	WHU_Shon_v2.1	GENBANK	GCF_014824575.2	2096623354	2136421	10164808	yes	10
Suricata suricatta	Mammalia	Laurasiatheria	Carnivora	Herpestidae	Suricata	Meerkat	70	meerkat_22Aug2017_6uvM2_HiC	DNAZOO	NA	2353578805	75409	141453419	no	NA
Sus scrofa	Mammalia	Laurasiatheria	Artiodactyla	Suidae	Sus	Pig	10	ss10.2_mar2013	GENBANK	GCA_001292865.1	2611360562	17259	153651326	no	NA
Sylvicapra grimmia	Mammalia	Laurasiatheria	Artiodactyla	Bovidae	Sylvicapra	Bush duiker	0	CMD	GENBANK	GCA_006408735.1	3145094493	9720	110456461	no	NA
Sylvilagus bachmani	Mammalia	Euarchontoglires	Lagomorpha	Leporidae	Sylvilagus	Brush rabbit	0	Sylvilagus_bachmani_HiC	DNAZOO	NA	2686389842	58367	110456461	yes	16
Symphalangus syndactalus	Mammalia	Euarchontoglires	Primates	Hylobatidae	Symphalangus	Siamang	10	Symphalangus_syndactylus_HiC	DNAZOO	NA	2896093032	32309	87438310	yes	15
Talpa occidentalis	Mammalia	Laurasiatheria	Eulipotyphla	Talpidae	Talpa	Iberian mole	100	MPIMG_talOcc4	GENBANK	GCF_014898055.1	2098003508	2611711	119794413	yes	6
Tamandua tetradactyla	Mammalia	Xenarthra	Pilosa	Myrmecophagidae	Tamandua	Southern tamandua	100	Tamandua_tetradactyla_M3075_1_07_2020_final_genome_HiC	DNAZOO	NA	3319533610	2227613	124746971	yes	7
Tapirella bairdii	Mammalia	Laurasiatheria	Perissodactyla	Tapiridae	Tapirella	Baird's tapir	0	Tapirella_bairdii_HiC	DNAZOO	NA	2412987941	35886	43941894	yes	12
Tapirus indicus	Mammalia	Laurasiatheria	Perissodactyla	Tapiridae	Tapirus	Malayan tapir	0	Tapirus_indicus_HiC	DNAZOO	NA	2607247779	156144	111226585	yes	13
Tolypeutes matacus	Mammalia	Xenarthra	Cingulata	Chlamyphoridae	Tolypeutes	Southern three-banded armadillo	80	TolMat_v1_BIUU	GENBANK	GCA_004025125.1	4115028163	9441	10217	no	NA
Tonatia saurophila	Mammalia	Laurasiatheria	Chiroptera	Phyllostomidae	Tonatia	Striped-headed round-eared bat	90	TonSau_v1_BIUU	GENBANK	GCA_004024845.1	2105886965	141649	165561	no	NA
Trachops cirrhosus	Mammalia	Laurasiatheria	Chiroptera	Phyllostomidae	Trachops	Fringe-lipped bat	50	Trachops_cirrhosus_HiC	DNAZOO	NA	2179538887	60077	124458213	yes	16
Tragelaphus eurycerus isaaci	Mammalia	Laurasiatheria	Artiodactyla	Bovidae	Tragelaphus	Eastern bongo	0	barney_pseudo2.1_HiC	DNAZOO	NA	2967381369	79488	192009155	yes	15
Tragulus javanicus	Mammalia	Laurasiatheria	Artiodactyla	Tragulidae	Tragulus	Java mouse-deer	0	ASM402496v2	GENBANK	GCA_004024965.2	2589955488	80230	14082842	no	NA
Tremarctos ornatus	Mammalia	Laurasiatheria	Carnivora	Ursidae	Tremarctos	Spectacled bear	0	Tremarctos_ornatus_HiC	DNAZOO	NA	2342009703	117282	104741234	yes	13
Trichechus manatus latirostris	Mammalia	Afrotheria	Sirenia	Trichechidae	Trichechus	West indian manatee	0	TriManLat1.0_HiC	DNAZOO	NA	3103881405	37747	143724118	yes	8
Tupaia chinensis	Mammalia	Euarchontoglires	Scandentia	Tupaiidae	Tupaia	Chinese tree shrew	90	TupChi_1.0	GENBANK	GCF_000334495.1	2846580235	25938	3670124	yes	12
Tursiops aduncus	Mammalia	Laurasiatheria	Artiodactyla	Delphinidae	Tursiops	Indo-pacific bottlenose dolphin	20	ASM322739v1_HiC	DNAZOO	NA	2505817531	133491	111961311	yes	1
Urocitellus parryii	Mammalia	Euarchontoglires	Rodentia	Sciuridae	Urocitellus	Arctic ground squirrel	20	ASM342692v1	GENBANK	GCF_003426925.1	2520505282	91013	3964291	yes	0
Urocyon littoralis catalinae	Mammalia	Laurasiatheria	Carnivora	Canidae	Urocyon	Santa catalina island fox	10	Urocyon_littoralis_catalinae_HiC	DNAZOO	NA	2542767600	125395	65046675	no	NA
Ursus americanus	Mammalia	Laurasiatheria	Carnivora	Ursidae	Ursus	American black bear	0	ASM334442v1_HiC	DNAZOO	NA	2597179268	11694	71790186	yes	16
Urva auropunctata	Mammalia	Laurasiatheria	Carnivora	Herpestidae	Urva	Small indian mongoose	30	Urva_auropunctata_HiC	DNAZOO	NA	2548765194	79588	132835555	no	NA
Varecia variegata	Mammalia	Euarchontoglires	Primates	Lemuridae	Varecia	Black and white ruff lemur	0	Varecia_variegata_HiC	DNAZOO	NA	2424028280	40155	128811464	yes	16
Vicugna pacos	Mammalia	Laurasiatheria	Artiodactyla	Camelidae	Vicugna	Alpaca	0	Vicugna_pacos-2.0.1_HiC	DNAZOO	NA	2171730293	24594	73726673	yes	10
Vulpes vulpes	Mammalia	Laurasiatheria	Carnivora	Canidae	Vulpes	Red fox	20	VulVul2.2_HiC	DNAZOO	NA	2421674764	20740	139030359	yes	15
Xerus rutilus	Mammalia	Euarchontoglires	Rodentia	Sciuridae	Xerus	Unstriped ground squirrel	10	Xerus_rutilus_HiC	DNAZOO	NA	2750759463	58140	127260801	yes	8
Zalophus californianus	Mammalia	Laurasiatheria	Carnivora	Otariidae	Zalophus	California sea lion	40	mZalCal1.pri.v2	GENBANK	GCF_009762305.2	2409668595	32626160	147124152	yes	10

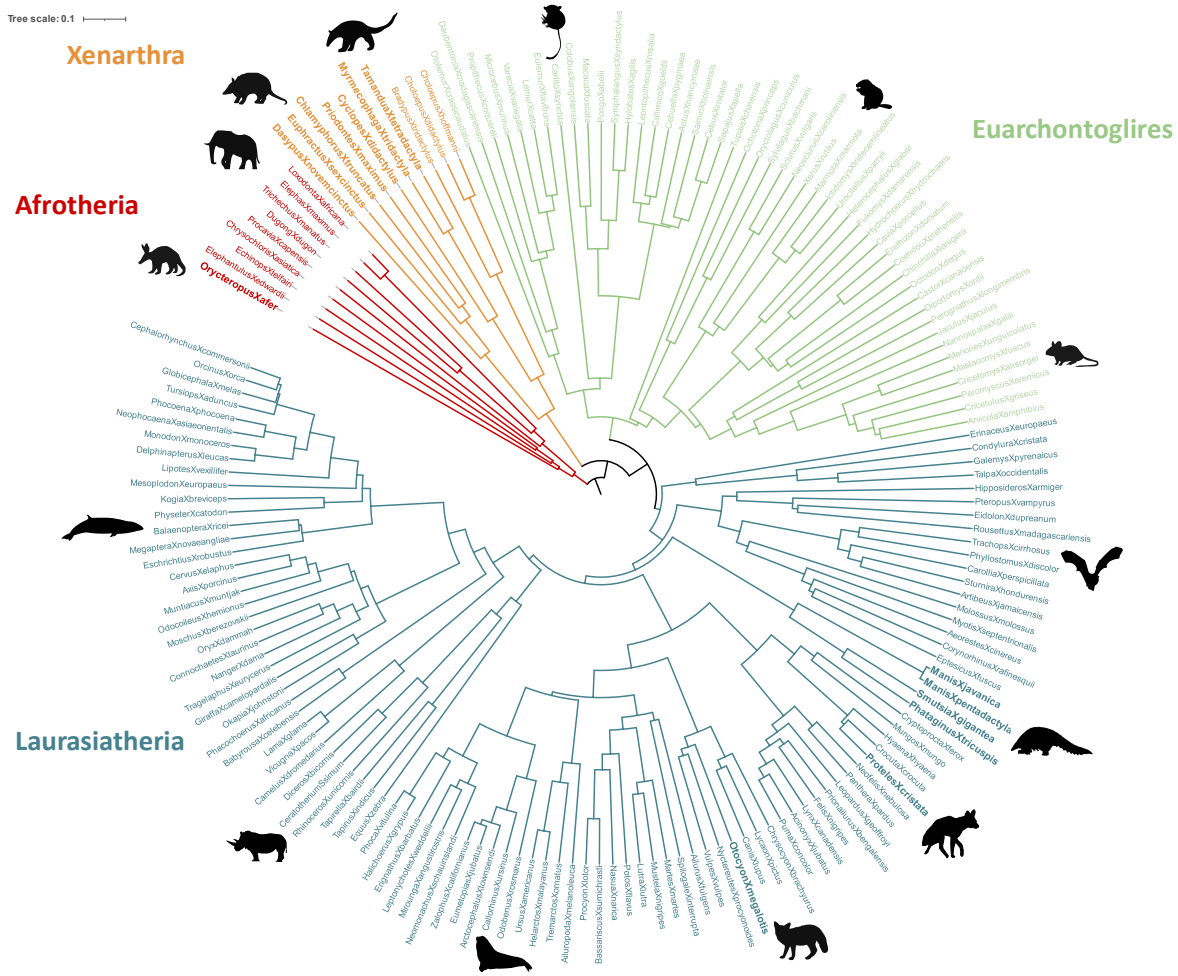


Figure S I.1 Phylogenetic relationships of the 170 species for which genome annotation files were available and used to extract *Tas2r* genes. This tree was obtained by pruning the species-level mammalian phylogeny of Álvarez-Carretero *et al* (2022) using the *drop.tip* function of the ape R package v5.3 (Paradis and Schliep, 2019) used in RStudio v3.5.3. Bold species names indicate myrmecophagous species. Silhouettes were downloaded from phylopic.org. Note that, contrary to most recent mammalian phylogenies, pangolins are not the sister group of carnivores.

I.2.4. Summary of *Tasr* gene evolution in myrmecophagous mammals

Analyzing taste receptor genes evolution in placentals revealed an evolution of this gene family marked by numerous gene losses. In myrmecophagous mammals, these analyses have highlighted different patterns of pseudogenization events. Recently diverged species such as Carnivoran myrmecophages seem to have retained *Tas1r* and *PKD2L1* genes whereas pangolins, more anciently diverged, seem to have lost sweet, umami, and sour taste perception. Among armadillos and anteaters, several taste losses have been reported with species unable to detect one or two of the three tastes studied. Moreover, further analyses should confirm a reduction of their bitter gene repertoires suggesting lower capabilities to detect bitter tastants except some potentially linked to ant and termite venom as specific bitter taste receptors could be retained (Garland, 2018).

Overall, these results suggest a reduction of taste perception in myrmecophagous species that might be linked to their adaptation toward this highly specialized diet and are consistent with previous studies (*e.g.*, Peel *et al*, 2015; Liu *et al*, 2016; Zhou *et al*, 2021). The narrow diet of ant- and termite-eating mammals, together with anatomical modifications of their tongue, which is not suited for taste perception but rather for rapid prey ingestion, have shaped their taste receptor gene repertoires. The pleiotropic effects of these genes will need further investigation, using comparative transcriptomics, to fully understand patterns of taste receptor gene losses in these species. Our analyses have also revealed shared and specific inactivating mutations among closely related myrmecophagous species for *Tas1r* genes. These results highlight the various underlying genomic adaptations involved in convergent myrmecophagous species in the light of their different timing of adaptation toward this diet (*i.e.*, ancient vs recent). More generally, these analyses further help us understand how myrmecophagous species perceive their prey, an important phenotypic trait to consider when trying to understand their convergent adaptation toward myrmecophagy.

I.3. Prey digestion in myrmecophagous mammals: insights from the study of the chitinase gene family

I.3.1. Introduction: evolution of mammalian chitinases

Chitin and chitinases

Chitin is a polymer of β-1,4-N-acetylglucosamin (Fig I.8) and the second most abundant polysaccharide on Earth after cellulose (El Knidri *et al*, 2018). It is an important component of several ecosystems, playing a central role in carbon and nitrogen cycles (Beier and Bertilsson, 2013), for instance in marine environments (Souza *et al*, 2011). Chitin is found in the exoskeleton of insects and the shell of crustaceans as well as in fungi, yeasts, and some algae (Gooday, 1990; Hamid *et al*, 2013; Rathore and Gupta, 2015; El Knidri *et al*, 2018).

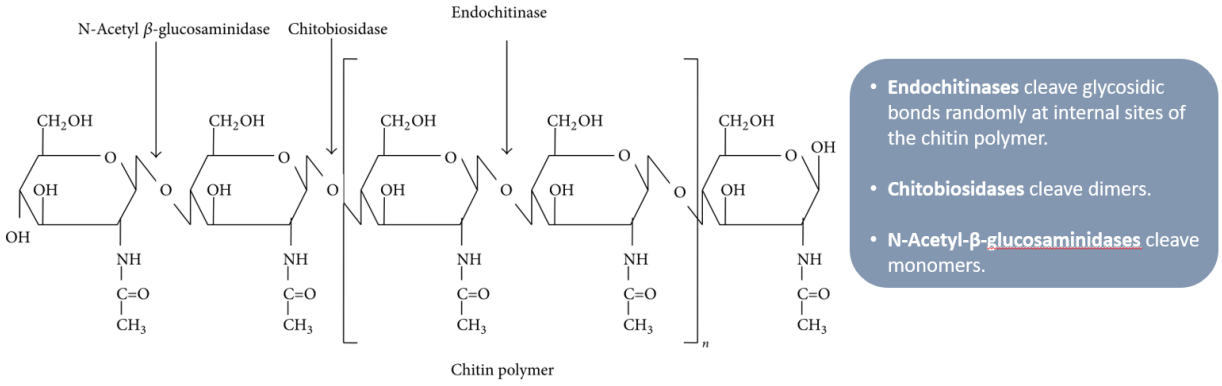


Figure I.8. Chitin structure and hydrolysis action of chitin-degrading enzymes. From Rathore and Gupta (2015).

Myrmecophagous mammals ingest high quantities of ants and/or termites notably because of the poor nutritive properties of their prey of which they need to digest the chitinous exoskeleton to access nutrients. To understand how myrmecophagous species convergently adapted to digest their prey and whether the same underlying mechanisms were involved, one could focus on studying the evolution of the associated digestive enzymes.

Chitinases are enzymes that degrade chitin by hydrolyzing it into smaller oligosaccharides thanks to their conserved active chitinolytic site composed of seven amino-acids, DXXDXDXE (D being aspartic acid, E glutamic acid, and X any amino acid) in which the glutamic acid in position 140 is the active proton donor site necessary for chitin hydrolysis (Olland *et al*, 2009; Hamid *et al*, 2013). Chitin binding domains are composed of six cysteines in C-terminal position (Tjoelker *et al*, 2000; Olland *et al*, 2009). Different types of chitinases hydrolyze the glycosidic bonds of the chitin polymer to produce small oligosaccharides with endochitinases cleaving bonds randomly at internal sites and exochitinases (*i.e*., N-acetyl-β-glucosaminidases, chitobiosidases) further cleaving the small oligomers produced by endochitinases (Rathore and Gupta, 2015; Fig I.7). Chitinases are mainly found in the Glycosyl Hydrolase enzyme family 18 (GH18) which is a Carbohydrate-Active enzymes family (CAZyme) containing chitinases and chitin-binding proteins (Huang *et al*, 2012; Rathore and Gupta, 2015). The GH18 family includes bacterial, viral, and fungal chitinases, as well as some plant, and mammalian chitinases (Rathore and Gupta, 2015). Other chitinases and chitin-binding proteins are found in the GH19 and GH20 families but mainly belong to plants, fungi, and bacteria (Rathore and Gupta, 2015).

Mammalian chitinases all have a common origin and constitute a multigenic family (Bussink *et al*, 2007; Funkhouser and Aronson, 2007; Hussain and Wilson, 2013). They are usually classified in two categories based on their enzymatic activity (Jeuniaux, 1959; Bussink *et al*, 2007; Funkhouser and Aronson, 2007; Hamid *et al*, 2013; Hussain and Wilson, 2013; Deeba *et al*, 2016):

- Chitinases (or active chitinases) with a catalytic function that can hydrolyze chitin and include Acidic Mammalian Chitinases (AMC or CHIA), chitotriosidase 1 (CHIT1), and chitobiase (CTBS). Active chitinases can participate in digestion and immunity.

- Inactive chitinases that cannot degrade chitin but can bind to it thanks to the chitin-binding domain. They are qualified as chitinase-like proteins, and are homologous to active chitinases. They ensure other functions, notably in immunity (*e.g*., Lee *et al*, 2011) or reproduction (*e.g*., Buhi, 2002), and include the oviduct-specific 1 enzyme (OVGP1), chitinase-3-like proteins (CHI3L1 and CHI3L2), and chitinase domain-containing 1 protein (CHID1).

Several studies have focused on deciphering the evolutionary history of these chitinase genes and try to understand how it might correlate with species ecological characteristics.

Evolution of chitinases in vertebrates and mammals

Jeuniaux (1961a, b) was one of the first to investigate chitinolytic activity in the digestive tract of vertebrates, including fishes, lizards, mammals, and birds. He identified gastric chitinases and highlighted a link between the presence of such enzymes in species having a chitin-rich diet (Jeuniaux, 1961a, b). Chitinase genes have also been identified in other vertebrate species consuming insects such as frogs (Dandrifosse, 1975; Fujimoto *et al*, 2002), lizards (Marsh *et al*, 2001), crustaceans-eating seabirds (Jackson *et al*, 1992), fish (Lindsay and Gooday, 1985), as well as carnivorous species like crocodiles (Siroski *et al*, 2014). These results suggest that chitinolytic enzymes are present in a diversity of species and are likely involved in chitin digestion. Jeuniaux (1971) hypothesized that the common ancestor of vertebrates had a diet composed of chitin and that from this ancestor, organisms not needing chitinases for digestion, due to a change in their diet, would not produce them anymore.

Recent phylogenetic analyses have then revealed the complex evolutionary history of the chitinase gene family in vertebrates in which several events of gene duplication led to the different known paralogues, except CTBS and CHID1, which diverged before all the other paralogues and did not further expand within vertebrate genomes (Bussink *et al*, 2007; Funkhouser and Aronson, 2007; Hussain and Wilson, 2013). This complex evolutionary history of chitinase genes in vertebrates was further confirmed by a study including much more data with 939 chitinase gene sequences from 242 different vertebrate species (see Chapter I annex). This work was done as part of my Master project and served as a basis to study in more details chitinase gene evolution in mammals (see part 1.3.2). Taken together these studies highlighted the global evolution of chitinases in vertebrates following a birth-and-death model of evolution with gene duplications giving birth to the different paralogues. Numerous gene losses occurred independently in several lineages, especially non-insectivorous ones, suggesting different functions for these paralogues with *CHIAs* probably playing important roles in digestion. Yet, questions remained regarding the specific evolution of these genes and their functions at finer evolutionary scales. Moreover, the link between

their evolution and the ecological diversification of vertebrate species, especially in the light of their dietary adaptations, needed further investigation.

To answer some of these questions, Emerling *et al* (2018) tested Jeuniaux's hypothesis (1971) by analyzing the evolutionary history of *CHIA* genes in 107 placental mammal species. They demonstrated that the common ancestor of placentals might have had at least five paralogues (*CHIA1-5*) suggesting that it was insectivorous, which would be consistent with the fossil record (Emerling *et al*, 2018). Besides, during the placental radiation, some paralogues were lost in certain lineages especially those shifting their diet toward herbivory or carnivory (Emerling *et al*, 2018). These losses resulted in a positive correlation between the number of functional *CHIA* paralogues carried by a species and the percentage of invertebrates in its diet (Emerling *et al*, 2018). This correlation was also found in primates (Janiak *et al*, 2018). Additionally, in birds only two functional paralogs of chitinase genes, *CHIA* and *CTBS*, were inferred in their common ancestor (Hussain and Wilson, 2013; Chen and Zhao, 2019). Consistent with the results found in mammals, chitinases of birds seem to have evolved in concert with the dietary diversification occurring within the group (Chen and Zhao, 2019).

To gain further insights into the evolution of chitinases and more specifically understand their role in the case of dietary adaptations, my PhD project focused on mammals and their chitinase genes, more specifically *CHIA* genes, following up on the exploration of chitinases in vertebrates (see Chapter I annex). The objectives were to i) reconstruct the evolutionary history of the nine paralogous chitinase genes in mammals using a reconciliation approach, ii) understand the functions of these paralogues by reconstructing ancestral sequences and comparing their chitinolytic and binding sites, and iii) unravel the role these genes played, especially *CHIAs*, in the adaptation to myrmecophagy in mammals by analyzing transcriptomic data to characterize their expression localization and hypothesize on the functions they might ensure. The manuscript in part 1.3.2 presents this work. In the following section (see "case study"), I briefly present the state of the art on chitinases of myrmecophagous mammals.

Case study: evolution of chitinases in myrmecophagous mammals

The positive correlation between the number of functional *CHIA* genes and the percentage of invertebrates in the diet of a species (Emerling *et al*, 2018) presents some discrepancies with species such as some marine mammals having one (common minke whale, *Balaenoptera acutorostrata*) or even no (walrus, *Odobenus rosmarus*) functional *CHIA* despite including 70% to 100% of invertebrates in their diet (Emerling *et al*, 2018). Within myrmecophagous species, these differences are also marked with species having one (*e.g.*, pangolins) to five (*e.g.*, aardvark) functional *CHIAs*. The aardvark (Tubulidentata) carries five functional *CHIA* genes in its genome (Emerling *et al*, 2018). The southern tamandua (*Tamandua tetradactyla;* Pilosa, Vermilingua) has four functional genes with only *CHIA5* being pseudogenized (Emerling *et al*, 2018). The pseudogenization of *CHIA5* in *T. tetradactyla* was estimated thanks to inactivating mutations and dated to 6.8 Mya, posterior to the origin of Vermilingua (34.2 Mya) and the divergence with *Myrmecophaga tridactyla* (11.3 Mya) (Emerling *et al*, 2018). Using BLAST to identify *CHIA* genes in genomes generated as part of the ConvergeAnt project, *CHIA5* was not found in the giant anteater (*M. tridactyla*) and silky anteater (*Cyclopes didactylus*) as opposed to the other four *CHIAs*, suggesting this gene might be absent and several independent *CHIA5* losses might have occurred in anteaters. Within armadillos (Cingulata), not all species are fully myrmecophagous but they can ingest substantial proportions of invertebrates. Four to five functional *CHIAs* have been identified in armadillo species, for instance, *D. novemcinctus* has lost *CHIA1* (Emerling *et al*, 2018). As most carnivorans, the aardwolf (*Proteles cristatus*) and the insectivorous bat-eared fox (*Otocyon megalotis*) have only one functional gene, *CHIA5*. A recent study has revealed that relaxed selective pressures are active on the only functional *CHIA* gene within carnivorans having a non-insectivorous diet leading to the loss of this gene by altering its structure and therefore functionality (Tabata *et al*, 2022). Species retaining it are those having an insectivorous diet such as the aardwolf or meerkat (Tabata *et al*, 2022). Pangolins (Pholidota) such as, the Malayan pangolin (*Manis javanica*), or the Chinese pangolin (*M. pentadactyla*), also have only *CHIA5* and have lost *CHIA1-4* (Emerling *et al*, 2018). Shared inactivating mutations were observed between carnivores and pangolins in *CHIA1* and date to at least 67 Mya, before the origin of Carnivora (46.2 Mya) and pangolins (26.5 Mya) (Emerling *et al*, 2018). This result suggests that *CHIA1-4* might have been lost in the common ancestor

of Pholidota and Carnivora, resulting in pangolins having only one functional *CHIA* gene despite a diet composed of 100% invertebrates.

In this context, questions arise regarding how these paralogues were involved in the adaptation to myrmecophagy in mammals and, more specifically, how ant- and termite-eating mammals use these different *CHIA* repertoires. Chitinases have been found expressed in the hypertrophied salivary glands and other digestive organs of *M. javanica*, with particularly high level of expression of the only functional *CHIA* gene (Ma *et al*, 2017; 2019; Cheng *et al*, 2023). These results were confirmed by proteomic analyses conducted on the saliva and intestinal juice of *M. javanica* (Zhang *et al*, 2019). This further raises questions on whether pangolin might compensate for having only one functional *CHIA* gene by overexpressing it in their digestive tract to ensure prey digestion, as opposed, for instance, to species like the southern tamandua having four functional *CHIAs*. Having several functional *CHIAs* could then allow expressing them equally along the digestive tract or use different paralogues depending on the tissue (*i.e.*, tissular specialization). Analyzing transcriptomic and/or proteomic data of chitinase genes in several digestive and non-digestive tissues of different myrmecophagous mammals would help understanding the role of these paralogues in adapting to this diet. More generally, it would shed light on the different molecular paths underlying the adaptation to myrmecophagy in mammals and how myrmecophagous species deal with differences in chitinase gene repertoires as a result of their independent evolutionary histories.

I.3.2. Manuscript. Comparative transcriptomics reveals divergent paths of chitinase evolution underlying dietary convergence in ant-eating mammals

The following article focuses on the underlying mechanisms involved in the convergent adaptation to myrmecophagy in placentals using comparative genomics and transcriptomics of chitinase genes to better understand prey digestion. The evolution of the nine chitinase paralogues (*CHIT1*, *OVGP1*, *CHI3L1*, *CHI3L2*, *CHIA1-5*) revealed an evolution of these paralogues toward different functions with *CHIAs* being involved in chitin degradation as their chitinolytic site is intact. In placentals, several *CHIA* gene losses occurred in non-insectivorous species (Emerling *et al*, 2018). Anteaters and pangolins have different *CHIA* repertoires: *Tamandua tetradactyla*, the southern tamandua, has four functional *CHIAs* (*CHIA1-4*) whereas *Manis javanica*, the Malayan pangolin, has only one functional *CHIA* (*CHIA5*) as a result of common ancestry with carnivorans (Emerling *et al*, 2018). This raises the question of how these genes contribute to chitin digestion in these two species. Here, comparative transcriptomics of salivary glands of 23 placental species and a comparison of chitinase gene expressions in several digestive and non-digestive organs of *T. tetradactyla* and *M. javanica* highlighted the importance of salivary glands in adapting to an insectivorous diet, and that chitinase genes are differently expressed in myrmecophagous species. Indeed, the only functional *CHIA* gene of *M. javanica* was found highly expressed in both its salivary glands and many digestive tissues (*i.e*., tongue, stomach, liver, pancreas, intestine), whereas only *CHIA3* and *CHIA4* were found highly expressed in the salivary glands and digestive tissues (*i.e*., tongue, liver, stomach) of *T. tetradactyla*, and *CHIA1* and *CHIA2* were found expressed in the pancreas and no other digestive organs. These results show that chitinase gene repertoires are used differently in these two myrmecophagous species as a result of historical contingency influencing the convergent adaptation to myrmecophagy in mammals.

Personal contribution

I contributed, as co first-author with R. Allio and D. Lutgen, as follow:

- Analyses of the evolution of the chitinase gene family: chitinase gene tree, reconciliation analysis, synteny analyses (Figs 1 and 2).
- Analyses of comparative transcriptomics of digestive and non-digestive organs of *Dasypus novemcinctus*, *Tamandua tetradactyla*, and *Manis javanica*: read cleaning, transcriptome assemblies, comparison of expression levels.
- Writing: part of the results and material and methods.
- Reading and editing.

This manuscript **is currently in revision for publication in Genome Biology and Evolution.**

The manuscript preprint can be found here:

Allio, R.[*], **Teullet, S**.[*], Lutgen, D.[*], Magdeleine, A., Koual, R., Tilak, M.-K., Thoisy, B. de, Emerling, C.A., Lefébure, T., Delsuc, F., 2023. Comparative transcriptomics reveals divergent paths of chitinase evolution underlying dietary convergence in ant-eating mammals. bioRxiv. https://doi.org/10.1101/2022.11.29.518312.

Comparative transcriptomics reveals divergent paths of chitinase evolution underlying dietary convergence in ant-eating mammals

Rémi Allio[1,2,§,*], Sophie Teullet[1,§], Dave Lutgen[1,3,4,§], Amandine Magdeleine[1], Rachid Koual[1], Marie-Ka Tilak[1], Benoit de Thoisy[5,6], Christopher A. Emerling[1,7], Tristan Lefébure[8], and Frédéric Delsuc[1,*]

[1]*ISEM, Univ. Montpellier, CNRS, IRD, Montpellier, France*

[2]*CBGP, INRAE, CIRAD, IRD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France*

[3]*Institute of Ecology and Evolution, University of Bern, Bern, Switzerland*

[4]*Swiss ornithological Institute, Sempach, Switzerland*

[5]*Institut Pasteur de la Guyane, Cayenne, French Guiana, France*

[6]*Kwata NGO, Cayenne, French Guiana, France*

[7]*Biology Department, Reedley College, Reedley, CA, USA*

[8]*Univ. Lyon, Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR 5023 LEHNA, F-69622, Villeurbanne, France*

[§]*Equal contribution*

***Correspondence**

Rémi Allio: remi.allio@inrae.fr

Frédéric Delsuc: frederic.delsuc@umontpellier.fr

Key words

Chitinases, Convergent evolution, Myrmecophagy, Mammals, Salivary glands, Transcriptomics

Abstract

Ant-eating mammals represent a textbook example of convergent evolution. Among them, anteaters and pangolins exhibit the most extreme convergent phenotypes with complete tooth loss, elongated skulls, protruding tongues, hypertrophied salivary glands producing large amounts of saliva, and powerful claws for ripping open ant and termite nests. However, comparative genomic analyses have shown that anteaters and pangolins differ in their chitinase gene (*CHIA*) repertoires, which potentially degrade the chitinous exoskeletons of ingested ants and termites. While the southern tamandua (*Tamandua tetradactyla*) harbors four functional *CHIA* paralogs (*CHIA1-4*), Asian pangolins (*Manis* spp.) have only one functional paralog

(*CHIA5*). Here, we performed a comparative transcriptomic analysis of salivary glands in 33 placental species, including 16 novel transcriptomes from ant-eating species and close relatives. Our results suggest that salivary glands play an important role in adaptation to an insect-based diet, as expression of different *CHIA* paralogs is observed in insectivorous species. Furthermore, convergently-evolved pangolins and anteaters express different chitinases in their digestive tracts. In the Malayan pangolin, *CHIA5* is overexpressed in all major digestive organs, whereas in the southern tamandua, all four functional paralogs are expressed, at very high levels for *CHIA1* and *CHIA2* in the pancreas, and for *CHIA3* and *CHIA4* in the salivary glands, stomach, liver, and pancreas. Overall, our results demonstrate that divergent molecular mechanisms underlie convergent adaptation to the ant-eating diet in pangolins and anteaters. This study highlights the role of historical contingency and molecular tinkering of the chitin-digestive enzyme toolkit in this classic example of convergent evolution.

Introduction

The phenomenon of evolutionary convergence is a fascinating process in which distantly related species independently acquire similar characteristics in response to the same selection pressures. A fundamental question famously illustrated by the debate between Stephen Jay Gould (Gould 2002) and Simon Conway Morris (Conway Morris 1999) resides in the relative contribution of historical contingency and evolutionary convergence in the evolution of biodiversity. While Gould (Gould 1990; 2002) argued that the evolution of species strongly depends on the characteristics inherited from their ancestors (historical contingency), Conway Morris (Conway Morris 1999) retorted that convergent evolution is one of the dominant processes leading to biodiversity evolution. Despite the huge diversity of organisms found on Earth and the numerous potential possibilities to adapt to similar conditions, the strong deterministic force of natural selection led to numerous cases of recurrent phenotypic adaptations (Losos 2011; McGhee 2011; Losos 2018). However, the role of historical contingency and evolutionary tinkering in convergent evolution has long been recognized, with evolution proceeding from available material through natural selection often leading to structural and functional imperfections (Jacob 1977). As first pointed out by François Jacob (Jacob 1977), molecular tinkering seems to be particularly frequent and has shaped the evolutionary history of a number of protein families (McGlothlin et al. 2016; Pillai et al. 2020; Xie et al. 2021). Indeed, if in some cases, convergent phenotypes can be associated with similar or identical mutations in the same genes occurring in independent lineages (Arendt and Reznick 2008), in other cases, they appear to arise by diverse molecular paths (*e.g.* Christin et al. 2010). Hence, both historical contingency and evolutionary convergence seems to have impacted the evolution of the current biodiversity and the major question relies on evaluating the relative impact of these two evolutionary processes (Blount et al. 2018).

A notable example of convergent evolution is the adaptation to the specialized ant- and/or termite-eating diet (*i.e.* myrmecophagy) in placental mammals (Reiss 2001). Within placental mammals, over 200 species include ants and termites in their regime, but only 22 of them can be considered as specialized myrmecophagous mammals, eating more than 90% of social insects (Redford 1987). Historically, based on shared morphological characteristics, ant-eating mammals were considered monophyletic (*i.e.* Edentata; Novacek 1992; O'Leary et al. 2013), but molecular phylogenetic evidence now strongly supports their polyphyly (*e.g.* Delsuc et al. 2002; Meredith et al. 2011; Springer et al. 2013). This highly-specialized diet has indeed independently evolved in five placental orders: armadillos (Cingulata), anteaters (Pilosa),

aardvarks (Tubulidentata), pangolins (Pholidota), and aardwolves (Carnivora). As a consequence of foraging for small-sized prey (Redford 1987), similar morphological adaptations have evolved in these mammalian species such as powerful claws used to dig into ant and termite nests, tooth reduction culminating in complete tooth loss in anteaters and pangolins (Ferreira-Cardoso et al. 2019), an elongated muzzle with an extensible tongue (Ferreira-Cardoso et al. 2020), and viscous saliva produced by hypertrophied salivary glands (Reiss 2001). Due to strong energetic constraints imposed by a nutritionally poor diet, myrmecophagous mammals also share relatively low metabolic rates and might thus require specific adaptations to extract nutrients from the chitinous exoskeletons of their prey (McNab 1984). It has long been shown that chitinase enzymes are present in the digestive tract of mammals and vertebrates more broadly (Jeuniaux 1961; Jeuniaux 1966; Jeuniaux 1971; Jeuniaux and Cornelius 1997). More recent studies have indeed shown that chitinase genes are present in the mammalian genome and may play an important digestive function in insectivorous species (Bussink et al. 2007; Emerling et al. 2018; Janiak et al. 2018; Wang et al. 2020; Cheng et al. 2023). Elevated levels of digestive enzyme gene expression have notably been observed in placental mammal salivary glands. For instance, in bat salivary glands, studies have shown that dietary adaptations can be associated with elevated expression levels in carbohydrase, lipase, and protease genes (Francischetti et al. 2013; Phillips et al. 2014; Vandewege et al. 2020).

In placental mammals, the salivary glands are composed of three major gland pairs (parotid, sublingual, and submandibular) and hundreds of minor salivary glands (Tucker 1958). In most myrmecophagous placental lineages, it has been shown that hypertrophied submandibular salivary glands are the primary source of salivary production. These enlarged horseshoe-shaped glands extend posteriorly along the side of the neck and ventrally over the chest. In the Malayan pangolin (*Manis javanica*), recent transcriptomic (Ma et al. 2017; Ma et al. 2019) and proteomic (Zhang et al. 2019) studies have shown that genes associated with digestive enzymes are highly expressed in salivary glands, which supports the hypothesis that the enlarged submandibular glands play an important functional role in social insect digestion. This result also found support in a study on the molecular evolution of the chitinase genes across 107 placental mammals that revealed the likely existence of a repertoire of five functional paralogous chitinase (*CHIA*, acidic mammalian chitinase) genes in the placental ancestor, which was subsequently shaped through multiple pseudogenization events associated with dietary adaptation during the placental radiation (Emerling et al. 2018). The widespread gene loss observed in carnivorous and herbivorous lineages resulted in a general positive correlation

between the number of functional *CHIA* paralogs and the percentage of invertebrates in the diet across placentals (Emerling et al. 2018). Indeed, mammals with a low proportion of insects in their diet present none or a few functional *CHIA* paralogs and those with a high proportion of insects in their diet generally have retained four or five functional *CHIA* paralogs (Emerling et al. 2018; Janiak et al. 2018; Wang et al. 2020). Among mammals, pangolins appear as an exception as the two investigated species (*M. javanica* and *Manis pentadactyla*) possess only one functional *CHIA* paralog (*CHIA5*) whereas other myrmecophagous species such as the southern tamandua (*Tamandua tetradactyla*) and the aardvark (*Orycteropus afer*) possess respectively four (*CHIA1-4*) and five (*CHIA1-5*) functional paralogs (Emerling et al. 2018). The presence of the sole *CHIA5* in pangolins was interpreted as the consequence of historical contingency with the probable loss of *CHIA1-4* functionality in the last common ancestor of Pholidota and Carnivora (Emerling et al. 2018). In Carnivora, it has recently been confirmed that a non insect-based diet has caused structural and functional changes in the *CHIA* gene repertoire resulting in multiple losses of function with only few species including insects in their diet retaining a fully functional *CHIA5* gene (Tabata et al. 2022). The fact that *CHIA5* was found to be highly expressed in the main digestive organs of the Malayan pangolin (Ma et al. 2017; Ma et al. 2019; Cheng et al. 2023) suggests that pangolins might compensate for their reduced chitinase repertoire by an increased ubiquitous expression of their only remaining functional paralog in multiple organs.

To test this hypothesis, we first reconstructed the detailed evolutionary history of the chitinase gene family in mammals. Then, we conducted a comparative transcriptomic analysis of chitinase gene expression in salivary glands of 33 placental mammal species including 16 newly generated transcriptomes from myrmecophagous placentals and other mammalian species. Finally, we compared the expression of chitinase paralogs in different organs between the nine-banded armadillo (*Dasypus novemcinctus*), the Malayan pangolin (*M. javanica*), and the southern tamandua (*T. tetradactyla*) for which we produced 13 new transcriptomes from nine additional organs. Our results shed light on the molecular underpinnings of convergent evolution in ant-eating mammals by revealing that divergent paths of chitinase molecular evolution underlie dietary convergence between anteaters and pangolins.

Results

Mammalian chitinase gene family evolution

The reconciled maximum likelihood tree of mammalian chitinase genes is presented in Figure 1A. The evolution of this gene family constituted by nine paralogs is characterized by the presence of numerous inferred gene losses with 384 speciation events followed by gene loss and 48 gene duplications as estimated by the gene tree/species tree reconciliation algorithm of GeneRax. At the base of the reconciled gene tree, we found the clade *CHIA1-2/OVGP1* (optimal root inferred by the reconciliation performed with TreeRecs) followed by a duplication separating the *CHIT1/CHI3L1-2* and *CHIA3-5* groups of paralogs. Within the *CHIT1/CHI3L* clade, two consecutive duplications gave rise to *CHIT1*, then *CHI3L1* and *CHI3L2*. In the *CHIA3-5* clade, a first duplication separated *CHIA3* from *CHIA4* and *CHIA5*, which were duplicated subsequently. Marsupial *CHIA4* sequences were located at the base of the *CHIA4-5* clade suggesting that this duplication might be specific to placentals. The *CHIA5* sequences of chiropterans were found at the base of the *CHIA5* clade. The duplication that gave rise to the *CHIA4* and *CHIA5* genes appears recent and specific to eutherians (marsupials and placentals) since no other taxon was found within these clades. This scenario of chitinase gene evolution is consistent with synteny analysis showing physical proximity of *CHIA1-2* and *OVGP1*, and *CHIA3-5* (Fig. 1B), which implies that chitinase genes evolved by successive tandem duplications. However, evidence of gene conversion between the two more recent duplicates (*CHIA4* and *CHIA5*) at least in some taxa suggests that further data are necessary to fully disentangle the origins of these two paralogs (Emerling et al. 2018). Within the *CHIA5* clade of Muroidea (Spalacidae, Cricetidae and Muridae), we found four subclades (named here *CHIA5a-d*) representing potential duplications specific to the muroid rodent species represented in our dataset. From the *CHIA5a* paralog, two consecutive duplications gave rise to the three *CHIA5b-d* paralogs represented by long branches, characterizing rapidly evolving sequences. The duplication giving rise to the *CHIA5c* and *CHIA5d* paralogs concerns only the Cricetidae and Muridae, *Nannospalax galili* (Spalacidae) being present only in the clade of the *CHIA5b* paralogous gene.

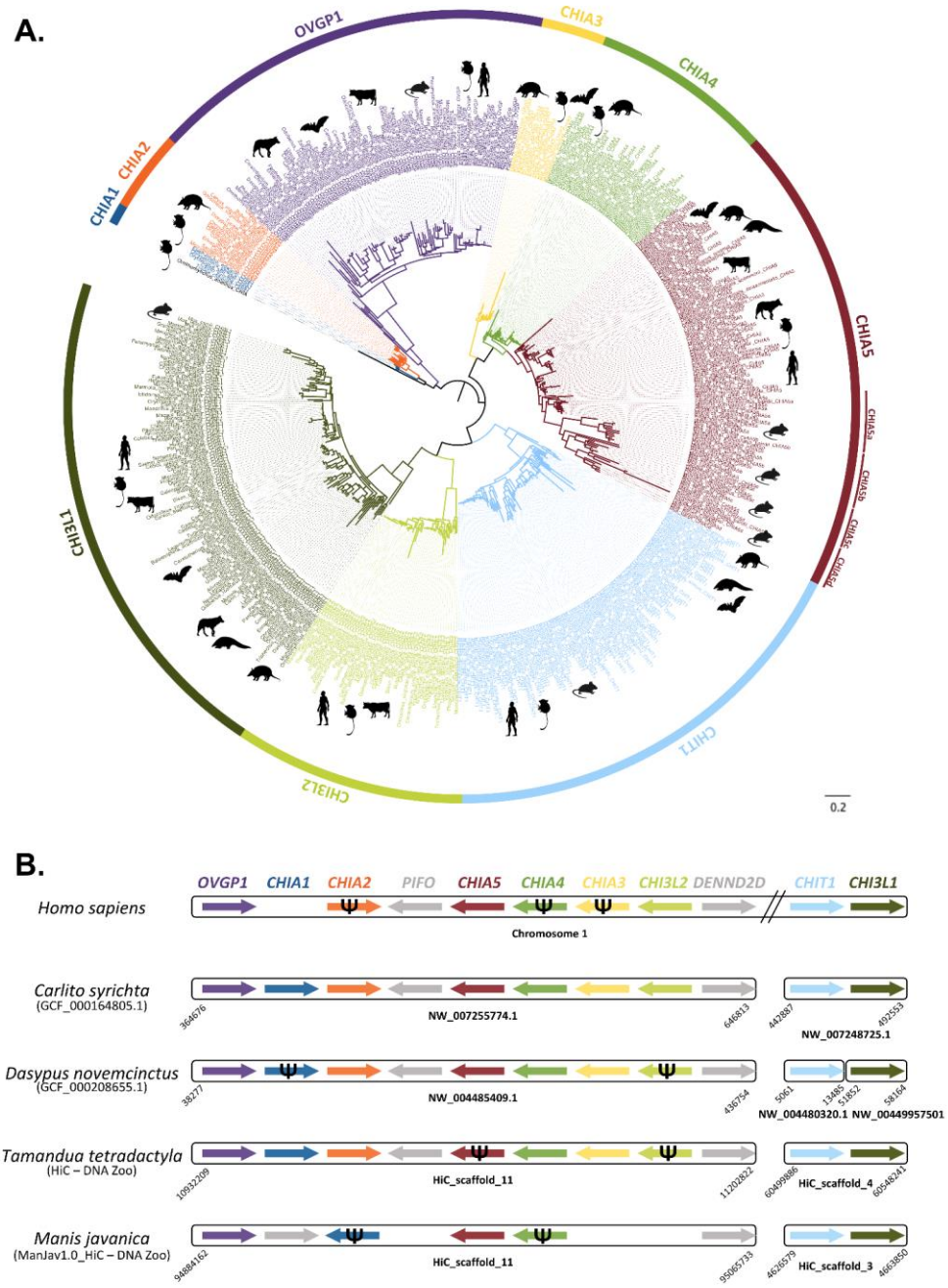


Figure 1: A. Mammalian chitinase gene family tree reconstructed using a maximum likelihood gene-tree/species-tree reconciliation approach on protein sequences. The nine chitinase paralogs are indicated on the outer circle. Scale bar represents the mean number of amino acid substitutions per site. B. Synteny of the nine chitinase paralogs in humans (*Homo sapiens*), tarsier (*Carlito syrichta*), nine-banded armadillo (*Dasypus novemcinctus*) and the two main focal convergent ant-eating species: the southern tamandua (*Tamandua tetradactyla*) and the Malayan pangolin (*Manis javanica*). Assembly names and accession numbers are indicated below species names. Arrows represent genes with scaffold/contig names and BLAST hit

positions indicated below. Arrow direction indicates gene transcription direction as inferred in Genomicus v100.01 (Nguyen et al. 2022) for genes located on short contigs. Ψ symbols indicate pseudogenes as determined in Emerling et al. (2018). Genes with negative BLAST results were not represented and are probably not functional or absent.

Ancestral sequences comparison

The ancestral amino acid sequences of the nine chitinase paralogs have been reconstructed from the reconciled mammalian gene tree and compared to gain further insight into the potential function of the enzymes they encode (Fig. 2). The alignment of predicted amino acid sequences locates the chitinolytic domain between positions 133 and 140 with the preserved pattern DXXDXDXE. The ancestral sequences of CHI3L1 and CHI3L2, as all contemporary protein sequences of these genes, have a mutated chitinolytic domain with absence of a glutamic acid at position 140 (Fig. 2A), which is the active proton-donor site necessary for chitin hydrolysis (Olland et al. 2009; Hamid et al. 2013). This indicates that the ability to degrade chitin has likely been lost before the duplication leading to CHI3L1 and CHI3L2 (Fig. 2B). It is also the case for the ancestral sequences of the muroid-specific CHIA5b-d, which thus cannot degrade chitin (data not shown). The ancestral sequence of OVGP1 also presents a mutated chitinolytic site although the glutamic acid in position 140 is present (Fig. 2A). The evolution of the different chitinases therefore seems to be related to changes in their active site. The six cysteine residues allowing the binding to chitin are found at positions 371, 418, 445, 455, 457 and 458 (Fig. 2C). The absence of one of these cysteines prevents binding to chitin (Tjoelker et al., 2000) as this is the case in the ancestral OVGP1 protein where the last four cysteine residues are changed (Fig. 2C). The other ancestral sequences present the six conserved cysteine residues and thus can bind to chitin (Fig. 2C).

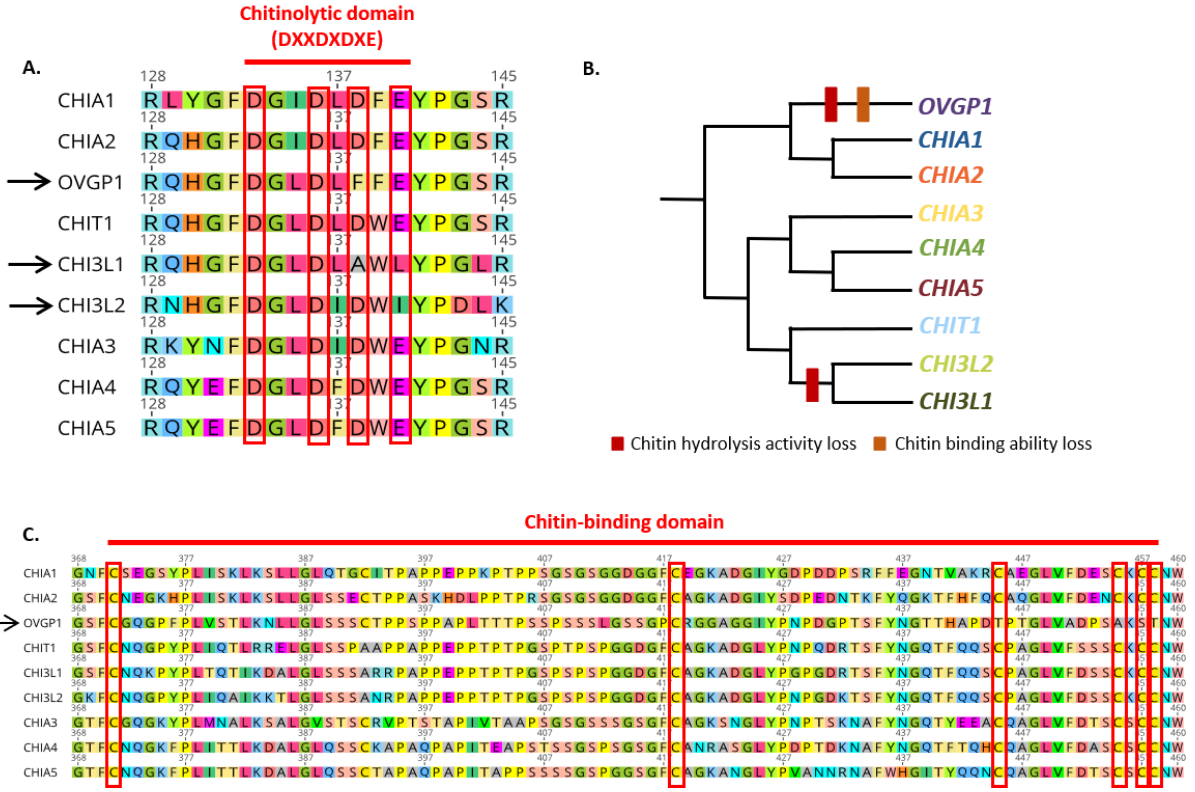


Figure 2: Comparison of predicted ancestral sequences of the nine mammalian chitinase paralogs. A. Conserved residues of the canonical chitinolytic domain active site (DXXDXDXE). Arrows indicate paralogs in which changes occurred in the active site. B. Summary of the evolution of chitinase paralogs functionality. C. Conserved cysteine residues of the chitin-binding domain. The arrow indicates OVGP1 in which the last four cysteines have been replaced.

Chitinase gene expression in mammalian salivary glands

To test the hypothesis that salivary glands play an important functional role in the digestion of ants and termites in ant-eating mammals, we analyzed the gene expression profiles of the nine chitinase paralogs revealed by the gene family tree reconstruction in 40 salivary gland transcriptomes (Fig. 3). *CHIA1* was expressed only in the elephant shrew (*Elephantulus myurus*; 23.22 normalized read counts [NC]). *CHIA2* was expressed only in the wild boar (*Sus scrofa*; 48.84 NC). *CHIA3* was expressed in the two insectivorous California leaf-nosed bats (*Macrotus californicus*; 367.70, and 35.03 NC) and in all three southern tamandua individuals (*T. tetradactyla*; 48.66, 41.52, and 15.14 NC). *CHIA4* was also highly expressed in all three southern tamandua individuals (565.61, 214.83, and 180.26 NC), in the giant anteater (*M. tridactyla*; 50.74 NC), and in the two California leaf-nosed bats (*M. californicus*; 17,224.06,

and 16,880.24 NC). Expression of *CHIA5* was much higher in the two Malayan pangolin individuals (*Manis javanica*; 196,778.69 and 729.18 NC) and Thomas's nectar bat (*Hsunycteris thomasi*; 7,301.82 NC) than in the three other species in which we detected expression of this gene: the domestic mouse (*Mus musculus*; 40.15 NC), common genet (*Genetta genetta*; 132.64 NC), and wild boar (*Sus scrofa*; 152.20 NC). *CHIT1* was expressed in many species (12 out of 40 samples) with NC values ranging from 46.76 NC in a single southern tamandua (*T. tetradactyla*) individual to 115,739.25 NC in the short-tailed shrew tenrec (*Microgale brevicaudata*). *CHI3L1* was expressed in most species (24 out of 40 samples) with values ranging from 61.68 NC in the giant anteater (*M. tridactyla*) to 1,297.01 NC in a Malayan pangolin (*M. javanica*) individual. *CHI3L2* was expressed in human (*H. sapiens*; 1334.07 NC), wild boar (*S. scrofa*; 246.41 NC), elephant shrew (*E. myurus*; 94.65 NC), and common tenrec (*Tenrec ecaudatus*; 68.62 NC). *OVGP1* was only found expressed at very low levels in domestic dog (*Canis lupus familiaris*; 6.80 NC), human (*H. sapiens*; 15.33 NC), one of the two Malayan pangolins (*M. javanica*; 4.99 NC) and wild boar (*S. scrofa*; 17.84 NC). Finally, the southern aardwolf (*P. cristatus*), Norway rat (*Rattus norvegicus*), Parnell's mustached bat (*Pteronotus parnellii*) and six phyllostomid bat species (*Carollia sowelli*, *Centurio senex*, *Glossophaga commissarisi*, *Sturnira hondurensis*, *Trachops cirrhosus*, and *Uroderma bilobatum*) did not appear to express any of the nine chitinase gene paralogs in any of our salivary gland samples.

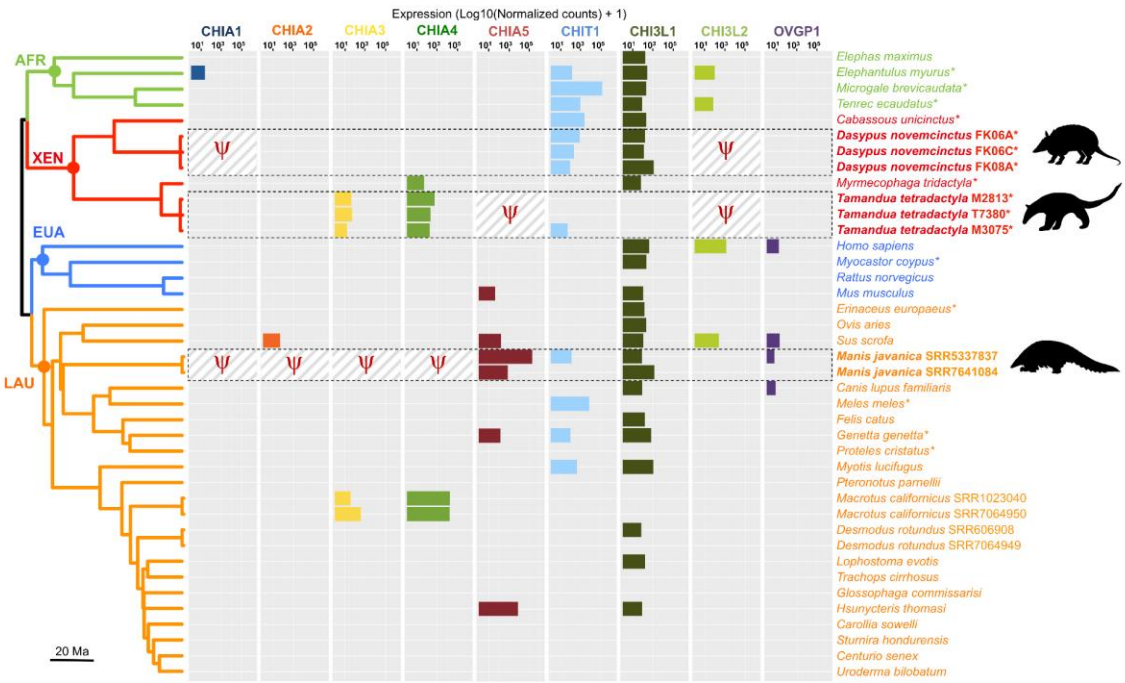


Figure 3: Comparative expression of the nine chitinase paralogs in 40 mammalian salivary gland transcriptomes. The 33 species are presented in the phylogenetic context covering the

four major placental clades: Afrotheria (AFR), Xenarthra (XEN), Euarchontoglires (EUA), and Laurasiatheria (LAU). The chronogram was extracted from www.timetree.org (Kumar et al. 2022). Non-functional pseudogenes of the three focal species (in bold) are represented by the Ψ symbol: nine-banded armadillo (*Dasypus novemcinctus*), southern tamandua (*Tamandua tetradactyla*) and Malayan pangolin (*Manis javanica*). Expression level is represented as log10 (Normalized Counts + 1). Asterisks indicate the 16 new transcriptomes produced in this study. Silhouettes were obtained from www.phylopic.org.

Chitinase gene expression in additional digestive and non-digestive organs

The expression level of the nine chitinase paralogs in several organs was compared among three species including an insectivorous xenarthran (the nine-banded armadillo; *D. novemcinctus*) and two of the main convergent myrmecophagous species (the southern anteater; *T. tetradactyla*, and the Malayan pangolin; *M. javanica*) (Fig. 4). This analysis revealed marked differences in expression level of these genes among the three species and among their digestive and non-digestive organs. *CHIT1* was expressed in all tissues in *M. javanica*, in the testes, tongue, salivary glands, and small intestine in *T. tetradactyla*, and in the cerebellum, lungs, salivary glands, and liver in *D. novemcinctus*. *CHI3L1* was found to be expressed in the majority of digestive and non-digestive tissues in all three species. *CHI3L2* is non-functional or even absent in the genome of these three species and was therefore not expressed. *OVGP1* was only weakly expressed in the lungs and salivary glands of *M. javanica* (2.22 and 4.99 NC, respectively).

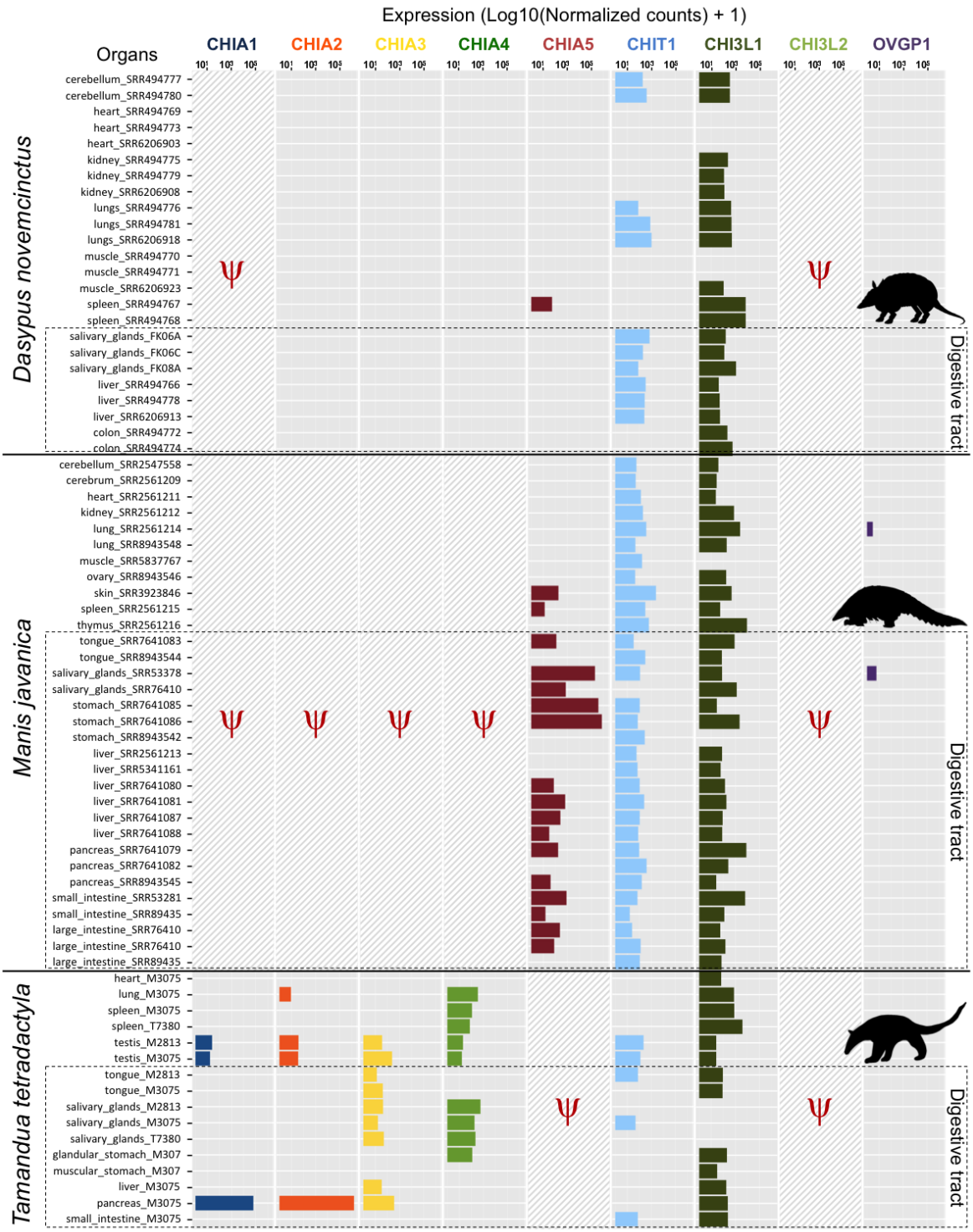


Figure 4: Comparative expression of the nine chitinase paralogs in 72 transcriptomes from different organs of the three focal species: the nine-banded armadillo (*Dasypus novemcinctus*), the Malayan pangolin (*Manis javanica*), and the southern tamandua (*Tamandua tetradactyla*). Non-functional pseudogenes are represented by the Ψ symbol and hatched background. Boxes indicate organs of the digestive tract. Expression level is represented as log10 (Normalized Counts + 1). Silhouettes were obtained from www.phylopic.org.

In the nine-banded armadillo (*D. novemcinctus*), although only *CHIA1* is pseudogenized and therefore logically not expressed, we did not detect any expression of *CHIA2*, *CHIA3*, and *CHIA4* in the tissues studied here, and *CHIA5* was only weakly expressed in one spleen sample (51.90 NC) (Fig. 4). In the Malayan pangolin (*M. javanica*), whereas *CHIA1-4* are non-functional and consequently not expressed, *CHIA5* was found expressed in all digestive organs with particularly high levels in the stomach (377,324.73 and 735,264.20 NC) and salivary glands (196,778.69 and 729.18 NC), and at milder levels in the tongue (121.24 NC), liver (254.79 NC on average when expressed), pancreas (168.64 and 39.33 NC), large intestine (238.45 and 79.32 NC), and small intestine (847.51 and 13.72 NC), but also in skin (178.95 NC) and spleen (12.06 NC) samples. Conversely, in the southern tamandua (*T. tetradactyla*), only *CHIA5* is pseudogenized and accordingly not expressed (Fig. 4). *CHIA1* was found highly expressed in the pancreas (64,443.05 NC) and weakly expressed in testes (22.74 and 14.73 NC), and *CHIA2* also had very high expression in the pancreas (1,589,834.39 NC), and low expression in testes (36.51 and 34.52 NC) and lungs (8.22 NC). *CHIA3* was also expressed in the pancreas (359.03 NC), testes (241.79 and 35.42 NC), tongue (39.53 and 12.44 NC), salivary glands (48.66, 41.52, and 15.14 NC), and liver (32.40 NC). Finally, *CHIA4* was expressed in the testes (19.48 and 14.59 NC), spleen (109.97 and 73.31 NC), lungs (340.84 NC), salivary glands (565.61, 214.83, and 180.26 NC), and glandular stomach (116.11 NC).

Discussion

Evolution of chitinase paralogs towards different functions

Chitinases have long been suggested to play an important role in mammalian insect digestion (Jeuniaux 1961; Jeuniaux 1966; Jeuniaux 1971; Jeuniaux and Cornelius 1997). Phylogenetic analyses of the Glycosyl Hydrolase gene family (GH18), which comprises genes encoding chitinase-like proteins, have revealed a dynamic evolutionary history despite a high degree of synteny among mammals (Bussink et al. 2007; Hussain and Wilson 2013). Our maximum likelihood phylogenetic analyses recovered nine functional paralogous chitinase gene sequences in mammalian genomes (Fig. 1A). In addition to the five previously characterized *CHIA* paralogs (Emerling et al. 2018; Janiak et al. 2018), we were able to identify an additional gene, *OVGP1*, which is most closely related to the previously characterized *CHIA1* and *CHIA2* genes. In mammals, OVGP1 plays a role in fertilization and embryonic development (Buhi 2002; Saint-Dizier et al. 2014; Algarra et al. 2016; Laheri et al. 2018). However, other aliases for OVGP1 include Mucin 9 and CHIT5 (www.genecards.org) suggesting a possible digestive

function. This result was further confirmed by synteny analyses suggesting a common origin by tandem duplication for *CHIA1-2* and *OVGP1* within the conserved chromosomal cluster that also includes *CHIA3-5* and *CHI3L2* (Fig. 1B). Comparison of the ancestral amino acid sequences of the nine chitinase paralogs revealed differences in their ability to bind and degrade chitin (Fig. 2), suggesting that these paralogs have evolved towards different functional specializations. The evolution of chitinase-like proteins was accompanied by a loss of enzymatic activity for chitin hydrolysis, which occurred several times independently (Bussink et al. 2007; Funkhouser and Aronson 2007; Hussain and Wilson 2013; Fig. 2B). CHI3L1 and CHI3L2, which are expressed in various cell types including macrophages and synovial cells, play roles in cell proliferation and immune response (Recklies et al. 2002; Areshkov et al. 2011; Lee et al. 2011). In contrast to these chitinase-like proteins, CHIT1 and the five CHIAs are able to degrade chitin. In humans, *CHIT1* is expressed in macrophages and neutrophils and is suspected to be involved in the defense against chitin-containing pathogens such as fungi (Gordon-Thomson et al. 2009; Lee et al. 2011). In addition to their role in chitin digestion (Boot et al. 2001), CHIAs are also suggested to play a role in the inflammatory response (Lee et al. 2011) and are expressed in non-digestive tissues, in agreement with our comparative transcriptomic results. Thus, it has been proposed that the expansion of the chitinase gene family is related to the emergence of the innate and adaptive immune systems in vertebrates (Funkhouser and Aronson 2007).

CHIA genes specific to muroid rodents and characterized by rapidly evolving sequences have also been described as chitinase-like rodent-specific (CHILrs) enzymes (Bussink et al. 2007; Hussain and Wilson 2013). These enzymes also appear to have evolved for functions in the immune response (Lee et al. 2011; Hussain and Wilson 2013). CHIA5b cannot bind to chitin, unlike CHIA5c and CHIA5d, suggesting different roles for these three paralogous proteins. The evolution of the different *CHIA1-5* genes has involved changes in their catalytic sites, which have consequences for the secondary structure of enzymes and potentially affect their optimal pH or function, as it has recently been shown for *CHIA5* in Carnivora (Tabata et al. 2022). Experimental testing of the chitin degrading activity on different substrates and at different pH of enzymes produced from the ancestral sequences reconstructed for each of the five *CHIA* paralogs would allow a better understanding of their enzymatic activity. Studying the potential binding of these enzymes to other substrates would shed more light on their functional roles. For example, changing a cysteine in the chitin-binding domain prevents binding to this substrate but not to tri-N-acetyl-chitotriose (Tjoelker et al. 2000), a compound derived from chitin with antioxidant properties (Chen et al. 2003; Salgaonkar et al. 2015). Such

functional assays, complemented by transcriptomic data to determine their expression profile in different tissues and organs (as previously done in the Malayan pangolin; Yusoff et al. 2016; Ma et al. 2017; Ma et al. 2019; Cheng et al. 2023), may help to decipher their respective roles in mammalian digestion (see below).

Impact of historical contingency and molecular tinkering on chitinase evolution and expression

In the specific case of adaptation to myrmecophagy, comparative genomic and transcriptomic analyses of these chitinase genes, particularly the chitin-degrading CHIAs, have led to a better understanding of how convergent adaptation to myrmecophagy in placentals occurs at the molecular level (Emerling et al. 2018; Cheng et al. 2023). On the one hand, anteaters (Pilosa; Vermilingua) likely inherited five *CHIA* genes from an insectivorous ancestor (Emerling et al. 2018), but then the *CHIA5* gene was lost. In the southern tamandua (*T. tetradactyla*), the inactivating mutations of *CHIA5* were identified and the estimated inactivation time of this gene was 6.8 Ma, subsequent to the origin of Vermilingua (34.2 Ma) and after the divergence with the giant anteater (*M. tridactyla*) at 11.3 Ma, suggesting a loss specific to lesser anteaters of the genus *Tamandua* (Emerling et al. 2018). In our study this gene was not found to be expressed in the salivary glands of the giant anteater. On the other hand, *CHIA5* is functional in insectivorous carnivores (Carnivora) and pangolins (Pholidota), whereas *CHIA1-4* are pseudogenized (Emerling et al. 2018; Tabata et al. 2022). Similar inactivating mutations have been observed in the *CHIA1* gene in carnivores and pangolins and dated to at least 67 Ma, well before the origin of carnivores (46.2 Ma) and pangolins (26.5 Ma) (Emerling et al. 2018). Thus, despite relying on a fully myrmecophagous diet, pangolins have only one functional *CHIA* gene, likely due to a historical contingency related to their common inheritance with carnivores. These analyses have thus revealed contrasting pseudogenization events between convergent myrmecophagous species, with lesser anteaters (genus *Tamandua*) retaining four out of the five functional chitin-degrading *CHIA* genes (*CHIA1-4*), while the Malayan pangolin (*M. javanica*) inherited only the fifth one (*CHIA5*). This peculiar evolutionary history raised the question whether the Malayan pangolin might compensate for the paucity of its functional chitinase gene repertoire by overexpressing *CHIA5* in different digestive organs.

Since the presence of enlarged salivary glands is a hallmark of convergent ant-eating mammals, ensuring massive production of saliva to help catch and potentially digest prey, we first investigated chitinase gene expression in mammalian salivary glands. Our comparative transcriptomic study spanning a diversity of species with different diets revealed that, among

ant-eating mammals, the Malayan pangolin (*M. javanica*), the southern tamandua (*T. tetradactyla*), and the giant anteater (*M. tridactyla*) all express one or more chitin-degrading genes in their salivary glands. More specifically, we found that *CHIA1* and *CHIA2* were almost never expressed in mammalian salivary glands. In contrast, *CHIA4* was found to be expressed in the giant anteater (M. tridactyla) and expression of both *CHIA3* and *CHIA4*was observed in the three southern tamandua (*T. tetradactyla*) individuals surveyed. Apart from anteaters, these two chitinase genes were found to be highly expressed only in the two individuals of the insectivorous California leaf-nosed bat (*M. californicus*), but not in any of the other 11 bat species including insectivorous species such as *M. myotis, P. parnellii, and L. evotis* (Fig. 3). . A possible explanation is that these genes have been pseudogenized in many of these bat species, which would be concordant with the findings of comparative genomic studies reporting widespread pseudogenizations of *CHIA* paralogs across multiple bat species (Emerling et al. 2018) with complete loss of *CHIA1-5* function in the vampire bat for instance (Wang et al. 2020). However, although *CHIA4* and *CHIA5* appear to be functional in the insectivorous little brown myotis (*M. lucifugus*; Emerling et al. 2018; Wang et al. 2020), we did not observe expression of these genes in the salivary gland transcriptome we analyzed. Also, *CHIA5* was found to be highly expressed in Thomas's nectar bat (*H. thomasi*). Although this bat species feeds mostly on nectar and fruits, its diet also includes a substantial part of insects suggesting that CHIA5 might play a role in chitin digestion in its salivary glands. Transcriptomic analyses of additional digestive tissues besides salivary glands in bats (Vandewege et al. 2020) may further clarify this pattern since chitinolytic activity has previously been reported in the stomachs of seven insectivorous bat species (Strobel et al. 2013). Finally, we were able to confirm the hypothesis implying an overexpression of the only functional *CHIA* gene possessed by the Malayan pangolin. Indeed, salivary gland expression profiles of *CHIA5* in *M. javanica* were much higher than in the four other species (Thomas's nectar bat, mouse, genet and wild boar) in which we detected expression of this gene, but also substantially higher than the expression of any other chitin-degrading *CHIA* in the 32 other mammalian species considered. Overall, our chitinase gene expression results therefore support a primary role for salivary glands in insect-eating placental mammal prey digestion through the use of distinct *CHIA* paralogs (*CHIA3*, *CHIA4*, and *CHIA5*) in different species.

Our differential expression comparison of the distinct chitinase paralogs across different organs further highlighted the importance of *CHIA5* for Malayan pangolin digestive physiology by confirming its ubiquitous expression in all major tissues of the digestive tract (tongue, salivary glands, stomach, pancreas, liver, and large and small intestines) 3; and Fig. 4). More

specifically, *CHIA5* was found to be expressed at particularly high levels in the stomach and salivary glands. These results are in line with previous proteomic studies that have also identified *CHIA5* as a digestive enzyme (Zhang et al. 2019), which has been confirmed to be highly expressed by RT-qPCR in the specialized oxyntic glands of the stomach (Ma et al. 2018a; Cheng et al. 2023), reflecting a key adaptation of the Malayan pangolin to its strictly myrmecophagous diet. By contrast, in the southern tamandua (*T. tetradactyla*) only *CHIA5* is pseudogenized (Emerling et al. 2018; Cheng et al. 2023) and all functional *CHIAs* were found expressed in its digestive tract but not in the same tissues. *CHIA1* and *CHIA2* were particularly highly expressed in the pancreas whereas *CHIA3* and *CHIA4* were expressed across several other organs of the digestive tract including tongue, salivary glands, stomach, and liver (Fig. 4). *CHIA1-4* were also expressed in other non-digestive organs (testes, lungs, and spleen), but their co-expression in the salivary glands of the three distinct southern tamandua individuals sampled here (Figs. 3, 4) strongly suggests that they play a crucial role in chitin digestion in this myrmecophagous species. Conversely, in the insectivorous nine-banded armadillo (*D. novemcinctus*), although only *CHIA1* is pseudogenized (Emerling et al. 2018) and therefore not expressed, we did not detect any expression of *CHIA2*, *CHIA3*, and *CHIA4* in the tissues of the individuals studied here, including salivary glands (Figs. 3, 4), and *CHIA5* was only weakly expressed in one spleen sample (Fig. 4). Yet, chitinases could still participate in prey digestion in the nine-banded armadillo as they have been isolated from gastric tissues (Smith et al. 1998); results we could not confirm here, the liver and colon being the only additional digestive organs besides salivary glands represented in our dataset for this species. However, the comparison with the two myrmecophagous species seems to fit well with its less specialized insectivorous diet and actually further underlines the contrasted specific use of distinct *CHIA* paralogs for chitin digestion in anteaters and pangolins.

Our results demonstrate that in the case of the southern tamandua (*T. tetradactyla*) and the Malayan pangolin (*M. javanica*), two myrmecophagous species that diverged about 100 Ma ago (Meredith et al. 2011), convergent adaptation to myrmecophagy has been achieved by using paralogs of different chitinase genes to digest chitin, probably due to phylogenetic constraints leading to the loss of *CHIA1*, *CHIA2*, *CHIA3*, and *CHIA4* in the ancestor of Ferae (Carnivora and Pholidota) as suggested by Emerling et al. (2018). Pangolins and anteaters present extreme morphological adaptations including the complete loss of dentition but a detailed study of their feeding apparatus has shown that convergent tooth loss resulted in divergent structures in the internal morphology of their mandible (Ferreira-Cardoso et al. 2019). Our results combined to this observation clearly show that the evolution of convergent phenotypes in myrmecophagous

mammals does not necessarily imply similar underlying mechanisms. Our study shows that historical contingency resulted in molecular tinkering (*sensu* Jacob 1977) of the chitinase gene family at both the genomic and transcriptomic levels. Working from different starting materials (*i.e.* different *CHIA* paralogs), natural selection led pangolins and anteaters to follow different paths in their adaptation to the myrmecophagous diet.

Material and Methods

Chitinase gene family tree reconstruction

Reconstruction of chitinase gene family evolution - The chitinase family in placental mammals appears to be composed of nine major paralogs (*CHIA1-5*, *CHIT1*, *CHI3L1*, *CHI3L2*, *OVGP1*). Mammalian sequences similar to the protein sequence of the human chitinase gene (NP_970615.2) were searched in the NCBI non-redundant protein database using BLASTP (E-value < 10). The protein sequences identified by BLASTP were then imported into Geneious Prime (Kearse et al. 2012) and aligned using MAFFT v7.450 (Katoh and Standley 2013) with the default parameters. Preliminary gene trees were then reconstructed with maximum likelihood using RAxML v8.2.11 (Stamatakis 2014) under the LG+G4 model (Le and Gascuel 2008) as implemented in Geneious Prime. From the reconstructed tree, the sequences were filtered according to the following criteria: (1) fast-evolving sequences with an E-value greater than zero and not belonging to the chitinase family were excluded; (2) in cases of multiple isoforms, only the longest was retained; (3) sequences whose length represented less than at least 50% of the total alignment length were removed; (4) in case of identical sequences from the same species the longest was kept; and (5) sequences labeled as "Hypothetical protein" and "Predicted: low quality protein" were discarded. This procedure resulted in a dataset containing 528 mammalian sequences that were realigned using MAFFT. This alignment was then cleaned up by removing sites not present in at least 50% of the sequences resulting in a total length of 460 amino acid sites. A maximum likelihood tree was then reconstructed with RAxML-NG v0.9.0 (Kozlov et al. 2019) using 10 tree searches starting from maximum parsimony trees under the LG+G8+F model. The species tree of the 143 mammal species represented in our dataset was reconstructed based on *COI* sequences extracted from the BOLD system database v4 (Ratnasingham and Hebert 2007) by searching for “Chordata” sequences in the “Taxonomy” section. Sequences were aligned using MAFFT, the phylogeny was inferred with RAxML and the topology was then adjusted manually based on the literature to correct ancient relationships.

To determine the optimal rooting scheme, a rapid reconciliation between the resulting gene tree and species tree was performed using the TreeRecs reconciliation algorithm based on maximum parsimony (Comte et al. 2020) as implemented in SeaView v5.0.2 (Gouy et al. 2010). The final chitinase gene family tree was produced using the maximum likelihood gene family tree reconciliation approach implemented in GeneRax v.1.1.0 (Morel et al. 2020) using the TreeRecs reconciled tree as input (source and result available from Zenodo). GeneRax can reconstruct duplications, losses, and horizontal gene transfer events but since the latter are negligible in mammals, only gene duplications and losses have been modeled here (--rec-model UndatedDL) and the LG+G model was used.

Ancestral sequence reconstructions - Ancestral sequences of the different paralogs were reconstructed from the reconciled tree using RAxML-NG (--ancestral function, --model LG+G8+F). The sequences were then aligned in Geneious Prime with MAFFT (source and result files available from Zenodo). Given that active chitinases are characterized by a catalytic site with a conserved amino acid motif (DXXDXDXE; Olland et al. 2009; Hamid et al. 2013), this motif was compared among all available species. Additionally, the six conserved cysteine residues responsible for chitin binding (Tjoelker et al. 2000; Olland et al. 2009) were also investigated.

Chitinase gene synteny comparisons - The synteny of the nine chitinase paralogs was compared between the two focal ant-eating species in our global transcriptomic analysis (*T. tetradactyla* and *M. javanica*), an insectivorous xenarthran species (*D. novemcinctus*), an insectivorous primate species with five functional *CHIA* genes (*Carlito syrichta*) and human (*Homo sapiens*). For *H. sapiens*, synteny information was added from Emerling et al. (2018) and completed by using Genomicus v100.01 (Nguyen et al. 2022). For *C. syrichta* and *D. novemcinctus*, genome assemblies have been downloaded from the National Center for Biotechnology Information (NCBI) and from the DNA Zoo (Choo et al. 2016; Dudchenko et al. 2017) for *M. javanica* and *T. tetradactyla*. Synteny information was retrieved by blasting (*megablast*) the different CDS sequences against these assemblies. Scaffold/contig names, positions and direction of BLAST hits were retrieved to compare their synteny (source and result files available from Zenodo). Genes with negative BLAST results were considered probably not functional or absent.

Transcriptome assemblies

Salivary gland transcriptomes - Biopsies of submandibular salivary glands (Gil et al. 2018) preserved in RNAlater were obtained from the Mammalian Tissue Collection of the Institut des Sciences de l'Evolution de Montpellier (ISEM) and the JAGUARS collection for 16 individuals representing 12 placental mammal species (Table S1). Total RNA was extracted from individual salivary gland tissue samples using the RNeasy extraction kit (Qiagen, Germany). Then, RNA-seq library construction and Illumina sequencing on a HiSeq 2500 system using paired-end 2x125bp reads were conducted by the Montpellier GenomiX platform (MGX) resulting in 16 newly produced salivary gland transcriptomes. This sampling was completed with the 26 mammalian salivary gland transcriptomes available as paired-end Illumina sequencing reads in the Short Read Archive (SRA) of the NCBI as of December 15th, 2022 representing an additional 21 species (Table S1). This taxon sampling includes representatives from all major mammal superorders Afrotheria (n = 4), Xenarthra (n = 4), Euarchontoglires (n = 4), and Laurasiatheria (n = 21) and covers six different diet categories: carnivory (n = 4), frugivory and herbivory (n = 8), insectivory (n = 9), myrmecophagy (n = 5), and omnivory (n = 7) (Table S1). Four of the five lineages in which myrmecophagous mammals evolved are represented: southern aardwolf (*P. cristatus*, Carnivora), Malayan pangolin (*M. javanica*, Pholidota), southern naked-tailed armadillo (*C. unicinctus*, Cingulata), giant anteater (*M. tridactyla*, Pilosa), and southern tamandua (*T. tetradactyla*, Pilosa). Species replicates in the form of different individuals were included for the southern tamandua (*T. tetradactyla*; n = 3), the nine-banded armadillo (*D. novemcinctus*; n = 3), the Malayan pangolin (*M. javanica*; n = 2), the vampire bat (*Desmodus rotundus*; n = 2), and the California leaf-nosed bat (*Macrotus californicus*; n = 2). We unfortunately were not able to obtain fresh salivary gland samples from the aardvark (*O. afer*, Tubulidentata), the only missing myrmecophagous lineage in our sampling.

Transcriptomes from additional organs - Tissue biopsies from nine additional organs (testis, lungs, heart, spleen, tongue, pancreas, stomach, liver, and small intestine) were sampled during dissections of three roadkill individuals of southern tamandua (*T. tetradactyla*; Table S1). Total RNA extractions from these RNAlater-preserved tissues, RNA-seq library construction, and sequencing were conducted as described above resulting in 13 newly generated transcriptomes. For comparative purposes, 21 additional transcriptomes of nine-banded armadillo (*D. novemcinctus*) representing eight organs and 30 transcriptomes of Malayan pangolin (*M. javanica*) representing 16 organs were downloaded from SRA (Table S1).

Comparative transcriptomics

Transcriptome assemblies and quality control - Adapters and low quality reads were removed from raw sequencing data using fastp v0.19.6 (Chen et al. 2018) using default parameters except for the PHRED score which was defined as "--qualified_quality_phred ≥ 15", as suggested by (MacManes 2014). Then, *de novo* assembly was performed on each individual transcriptome sample using Trinity v2.8.4 (Grabherr et al. 2011) using default parameters. For one individual vampire bat (*D. rotundus*), three salivary gland transcriptomes (SRR606902, SRR606908, and SRR606911) were combined to obtain a better assembly. For each of the 104 transcriptome assemblies, completeness was assessed by the presence of Benchmark Universal Single Copy Orthologs (BUSCO v5) based on a dataset of 9,226 single-copy orthologs conserved in over 90% of mammalian species (Manni et al. 2021). This pipeline was run through the gVolante web server (Nishimura et al. 2017) to evaluate the percentage of complete, duplicated, fragmented and missing single copy orthologs within each transcriptome (Table S2).

Transcriptome annotation and orthogroup inference - The 104 transcriptome assemblies were annotated following the pipeline implemented in assembly2ORF (https://github.com/ellefeg/assembly2orf). This pipeline combines evidence-based and gene-model-based predictions. First, potential transcripts of protein-coding genes are extracted based on similarity searches (BLAST) against the peptides of Metazoa found in Ensembl (Yates et al. 2020). Then, using both protein similarity and exonerate functions (Slater and Birney 2005), a frameshift correction is applied to candidate transcripts. Candidate open reading frames (ORFs) are predicted using TransDecoder (https://github.com/TransDecoder/TransDecoder) and annotated based on homology information inferred from both BLAST and Hmmscan searches. Finally, to be able to compare the transcriptomes obtained from all species, we relied on the inference of gene orthogroups. The orthogroup inference for the translated candidate ORFs was performed using OrthoFinder v2 (Emms and Kelly 2019) using FastTree (Price et al. 2010) for gene tree reconstructions. For expression analyses, orthogroups containing more than 20 copies for at least one species were discarded.

Gene expression analyses - Quantification of transcript expression was performed on Trinity assemblies with Kallisto v.0.46.1 (Bray et al. 2016) using the *align_and_estimate_abundance.pl* script provided in the Trinity suite (Grabherr et al. 2011). Kallisto relies on pseudo-alignments of the reads to search for the original transcript of a read without looking for a perfect alignment (as opposed to classical quantification by counting the

reads aligned on the assembled transcriptome; Wolf 2013). Counts (raw number of mapped reads) and the Transcripts Per kilobase Million are reported (result files available from Zenodo). Based on the previously inferred orthogroups, orthogroup-level abundance estimates were imported and summarized using tximport (Soneson et al. 2016). To minimize sequencing depth variation across samples and gene outlier effect (a few highly and differentially expressed genes may have strong and global influence on the total read count), orthogroup-level raw reads counts were normalized using the median of the ratios of observed counts using DESeq2 (Love et al. 2014) for orthogroups containing up to 20 gene copies by species.

Chitinase expression in salivary glands - The chitinase orthogroup was extracted from the orthogroups inferred by OrthoFinder2 using BLASTX with the reference chitinase database previously created. The 476 amino acid sequences composing this orthogroup were assigned to the nine chitinase orthologs (*CHIA1-5*, *CHIT1*, *CHI3L1*, *CHI3L2*, *OVGP1*) using the maximum likelihood Evolutionary Placement Algorithm implemented in RAxML-EPA (Berger et al. 2011) with the reference chitinase sequence alignment and reconciled phylogenetic tree previously inferred using GenRax (result files available from Zenodo). This allowed excluding three additional contaminant sequences and dividing the chitinase orthogroup into nine sub-orthogroups corresponding to each chitinase paralog. To take advantage of the transcriptome-wide expression information for the expression standardization, these new orthogroups were included in the previous orthogroup-level abundance matrix estimates and the same normalization approach using DESeq2 was conducted. Finally, gene-level abundance estimates for all chitinase paralogs were extracted and compared on a log10 scale.

Data and Resource Availability

Raw RNAseq Illumina reads have been submitted to the Short Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) and are available under BioProject number PRJNA909065. Transcriptome assemblies, phylogenetic datasets, corresponding trees, and other supplementary materials are available from zenodo.org (DOI: 10.5281/zenodo.7790047).

Acknowledgments

We would like to thank Hugues Parrinello (Montpellier GenomiX platform) for advice on RNAseq, Mariana Escobar Rodríguez and Gautier Debaecker for help with transcriptome assembly and annotation, and Marie Sémon for providing useful advice on RNAseq statistical analyses. We are also indebted to Frank Knight, Mark Scherz, Miguel Vences, Andolalao Rakotoarison, Nico Avenant, Pierre-Henri Fabre, Quentin Martinez, Nathalie Delsuc, Aude Caizergues, Roxanne Schaub, Lionel Hautier, Fabien Condamine, Sérgio Ferreira-Cardoso, and François Catzeflis for their help with tissue sampling. Computational analyses benefited from the Montpellier Bioinformatics Biodiversity (MBB) platform. The JAGUARS collection is supported through a FEDER/ERDF grant attributed to Kwata NGO, funded by the European Union, the Collectivité Territoriale de Guyane, and the DEAL Guyane. This work has been supported by grants from the European Research Council (ConvergeAnt project: ERC-2015-CoG-683257) and Investissements d'Avenir of the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01; CEMEB: ANR-10-LABX-0004). This is contribution ISEM 2023-XXX of the Institut des Sciences de l'Evolution de Montpellier.

References

Algarra B, Han L, Soriano-Úbeda C, Avilés M, Coy P, Jovine L, Jiménez-Movilla M. 2016. The C-terminal region of OVGP1 remodels the zona pellucida and modifies fertility parameters. *Sci. Rep.* 6:32556.

Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: What have we learned about the genetics of adaptation? *Trends Ecol. Evol.* 23:26–32.

Areshkov PO, Avdieiev SS, Balynska OV, LeRoith D, Kavsan VM. 2011. Two closely related human members of Chitinase-like family, CHI3L1 and CHI3L2, activate ERK1/2 in 293 and U373 cells but have the different influence on cell proliferation. *Int. J. Biol. Sci.* 8:39–48.

Berger SA, Krompass D, Stamatakis A. 2011. Performance, Accuracy, and Web Server for Evolutionary Placement of Short Sequence Reads under Maximum Likelihood. *Syst. Biol.* 60:291–302.

Blount ZD, Lenski RE, Losos JB. 2018. Contingency and determinism in evolution: Replaying life's tape. *Science* 362:eaam5979.

Boot RG, Blommaart EF, Swart E, Ghauharali-van der Vlugt K, Bijl N, Moe C, Place A, Aerts JM. 2001. Identification of a novel acidic mammalian chitinase distinct from chitotriosidase. *J. Biol. Chem.* 276:6770–6778.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34:525–527.

Buhi WC. 2002. Characterization and biological roles of oviduct-specific, oestrogen-dependent glycoprotein. *Reproduction* 123:355–362.

Bussink AP, Speijer D, Aerts JMFG, Boot RG. 2007. Evolution of mammalian Chitinase(-like) members of family 18 Glycosyl Hydrolases. *Genetics* 177:959–970.

Chen A-S, Taguchi T, Sakai K, Kikuchi K, Wang M-W, Miwa I. 2003. Antioxidant activities of Chitobiose and Chitotriose. *Biol. Pharm. Bull.* 26:1326–1330.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890.

Cheng S-C, Liu C-B, Yao X-Q, Hu J-Y, Yin T-T, Lim BK, Chen W, Wang G-D, Zhang C-L, Irwin DM, et al. 2023. Hologenomic insights into mammalian adaptations to myrmecophagy. *Natl. Sci. Rev.* 10:nwac174.

Choo SW, Rayko M, Tan TK, Hari R, Komissarov A, Wee WY, Yurchenko AA, Kliver S, Tamazian G, Antunes A. 2016. Pangolin genomes and the evolution of mammalian scales and immunity. *Genome Res.* 26:1312–1322.

Christin P-A, Weinreich DM, Besnard G. 2010. Causes and evolutionary significance of genetic convergence. *Trends Genet.* 26:400–405.

Comte N, Morel B, Hasić D, Guéguen L, Boussau B, Daubin V, Penel S, Scornavacca C, Gouy M, Stamatakis A, et al. 2020. Treerecs: an integrated phylogenetic tool, from sequences to reconciliations. *Bioinformatics* 36:4822–4824.

Conway Morris S. 1999. The crucible of creation: The Burgess Shale and the rise of animals. Oxford, New York: Oxford University Press

Delsuc F, Scally M, Madsen O, Stanhope MJ, de Jong WW, Catzeflis FM, Springer MS, Douzery EJP. 2002. Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. *Mol. Biol. Evol.* 19:1656–1671.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* 356:92–95.

Emerling CA, Delsuc F, Nachman MW. 2018. Chitinase genes (CHIAs) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. *Sci. Adv.* 4:eaar6478.

Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238.

Ferreira-Cardoso S, Delsuc F, Hautier L. 2019. Evolutionary tinkering of the mandibular canal

linked to convergent regression of teeth in placental mammals. *Curr. Biol.* 29:468–475.

Ferreira-Cardoso S, Fabre P-H, Thoisy B de, Delsuc F, Hautier L. 2020. Comparative masticatory myology in anteaters and its implications for interpreting morphological convergence in myrmecophagous placentals. *PeerJ* 8:e9690.

Francischetti IMB, Assumpção TCF, Ma D, Li Y, Vicente EC, Uieda W, Ribeiro JMC. 2013. The "Vampirome": Transcriptome and proteome analysis of the principal and accessory submaxillary glands of the vampire bat Desmodus rotundus, a vector of human rabies. *J. Proteomics* 82:288–319.

Funkhouser JD, Aronson NN. 2007. Chitinase family GH18: Evolutionary insights from the genomic history of a diverse protein family. *BMC Evol. Biol.* 7:96.

Gil F, Arencibia A, García V, Ramírez G, Vázquez JM. 2018. Anatomic and magnetic resonance imaging features of the salivary glands in the dog. *Anat. Histol. Embryol.* 47:551–559.

Gordon-Thomson C, Kumari A, Tomkins L, Holford P, Djordjevic JT, Wright LC, Sorrell TC, Moore GPM. 2009. Chitotriosidase and gene therapy for fungal infections. *Cell. Mol. Life Sci.* 66:1116–1125.

Gould SJ. 1990. Wonderful life: The Burgess Shale and the nature of history. WW Norton & Company

Gould SJ. 2002. The Structure of Evolutionary Theory. Harvard University Press

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27:221–224.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29:644–652.

Hamid R, Khan MA, Ahmad M, Ahmad MM, Abdin MZ, Musarrat J, Javed S. 2013. Chitinases: An update. *J. Pharm. Bioallied Sci.* 5:21–29.

Hussain M, Wilson JB. 2013. New paralogues and revised time line in the expansion of the vertebrate GH18 family. *J. Mol. Evol.* 76:240–260.

Jacob F. 1977. Evolution and tinkering. *Science* 196:1161–1166.

Janiak MC, Chaney ME, Tosi AJ. 2018. Evolution of acidic mammalian chitinase genes (CHIA) is related to body mass and insectivory in Primates. *Mol. Biol. Evol.* 35:607–622.

Jeuniaux C. 1961. Chitinase: An addition to the list of hydrolases in the digestive tract of vertebrates. *Nature* 192:135–136.

Jeuniaux C. 1966. [111] Chitinases. In: Methods in enzymology. Vol. 8. Elsevier. p. 644–650.

Jeuniaux C. 1971. On some biochemical aspects of regressive evolution in animals. In: Biochemical evolution and the origin of life. E. Schoffeniels. p. 304–313.

Jeuniaux C, Cornelius C. 1997. Distribution and activity of chitinolytic enzymes in the digestive tract of birds and mammals. In: First international conference on Chitin/Chitosan.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software Version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35:4453–4455.

Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022. TimeTree 5: An Expanded Resource for Species Divergence Times. *Mol. Biol. Evol.* 39:msac174.

Laheri S, Ashary N, Bhatt P, Modi D. 2018. Oviductal glycoprotein 1 (OVGP1) is expressed by endometrial epithelium that regulates receptivity and trophoblast adhesion. *J. Assist. Reprod. Genet.* 35:1419–1429.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25:1307–1320.

Lee CG, Da Silva CA, Dela Cruz CS, Ahangari F, Ma B, Kang M-J, He C-H, Takyar S, Elias JA. 2011. Role of chitin and Chitinase/Chitinase-like proteins in inflammation, tissue remodeling, and injury. *Annu. Rev. Physiol.* 73:479–501.

Losos JB. 2011. Convergence, adaptation, and constraint. *Evol. Int. J. Org. Evol.* 65:1827–1840.

Losos JB. 2018. Improbable destinies: Fate, chance, and the future of evolution. Penguin

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

Ma J-E, Jiang H-Y, Li L-M, Zhang X-J, Li H-M, Li G-Y, Mo D-Y, Chen J-P. 2019. SMRT sequencing of the full-length transcriptome of the Sunda pangolin (Manis javanica). *Gene* 692:208–216.

Ma J-E, Li L-M, Jiang H-Y, Zhang X-J, Li J, Li G-Y, Chen J-P. 2018. Acidic mammalian chitinase gene is highly expressed in the special oxyntic glands of Manis javanica. *FEBS Open*

Bio 8:1247–1255.

Ma J-E, Li L-M, Jiang H-Y, Zhang X-J, Li J, Li G-Y, Yuan L-H, Wu J, Chen J-P. 2017. Transcriptomic analysis identifies genes and pathways related to myrmecophagy in the Malayan pangolin (Manis javanica). *PeerJ* 5:e4140.

MacManes M. 2014. On the optimal trimming of high-throughput mRNA sequence data. *Front. Genet.* 5:13.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 38:4647–4654.

McGhee GR. 2011. Convergent evolution: Limited forms most beautiful. MIT Press

McGlothlin JW, Kobiela ME, Feldman CR, Castoe TA, Geffeney SL, Hanifin CT, Toledo G, Vonk FJ, Richardson MK, Brodie ED, et al. 2016. Historical contingency in a multigene family facilitates adaptive evolution of toxin resistance. *Curr. Biol.* 26:1616–1621.

McNab BK. 1984. Physiological convergence amongst ant-eating and termite-eating mammals. *J. Zool.* 203:485–510.

Meredith RW, Janečka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simão TLL, Stadler T, et al. 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* 334:521–524.

Morel B, Kozlov AM, Stamatakis A, Szöllősi GJ. 2020. GeneRax: A tool for species-tree-aware maximum likelihood-based gene family tree inference under gene duplication, transfer, and loss. *Mol. Biol. Evol.* 37:2763–2774.

Nguyen NTT, Vincens P, Dufayard JF, Roest Crollius H, Louis A. 2022. Genomicus in 2022: Comparative tools for thousands of genomes and reconstructed ancestors. *Nucleic Acids Res.* 50:D1025–D1031.

Nishimura O, Hara Y, Kuraku S. 2017. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinforma. Oxf. Engl.* 33:3635–3637.

Novacek MJ. 1992. Mammalian phylogeny: Shaking the tree. *Nature* 356:121–125.

O'Leary MA, Bloch JI, Flynn JJ, Gaudin TJ, Giallombardo A, Giannini NP, Goldberg SL, Kraatz BP, Luo Z-X, Meng J, et al. 2013. The Placental Mammal Ancestor and the Post–K-Pg Radiation of Placentals. *Science* 339:662–667.

Olland AM, Strand J, Presman E, Czerwinski R, Joseph-McCarthy D, Krykbaev R, Schlingmann G, Chopra R, Lin L, Fleming M, et al. 2009. Triad of polar residues implicated in pH specificity of acidic mammalian chitinase. *Protein Sci.* 18:569–578.

Phillips CJ, Phillips CD, Goecks J, Lessa EP, Sotero-Caio CG, Tandler B, Gannon MR, Baker

RJ. 2014. Dietary and flight energetic adaptations in a salivary gland transcriptome of an insectivorous bat. *PLOS ONE* 9:e83512.

Pillai AS, Chandler SA, Liu Y, Signore AV, Cortez-Romero CR, Benesch JLP, Laganowsky A, Storz JF, Hochberg GKA, Thornton JW. 2020. Origin of complexity in haemoglobin evolution. *Nature* 581:480–485.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS One* 5:e9490.

Ratnasingham S, Hebert PDN. 2007. bold: The barcode of life data system (http://www.barcodinglife.org). *Mol. Ecol. Notes* 7:355–364.

Recklies AD, White C, Ling H. 2002. The chitinase 3-like protein human cartilage glycoprotein 39 (HC-gp39) stimulates proliferation of human connective-tissue cells and activates both extracellular signal-regulated kinase- and protein kinase B-mediated signalling pathways. *Biochem. J.* 365:119–126.

Redford KH. 1987. Ants and termites as food. In: Genoways HH, editor. Current Mammalogy. Boston, MA: Springer US. p. 349–399.

Reiss KZ. 2001. Using phylogenies to study convergence: the case of the ant-eating mammals. *Am. Zool.* 41:507–525.

Saint-Dizier M, Marnier C, Tahir MZ, Grimard B, Thoumire S, Chastant-Maillard S, Reynaud K. 2014. OVGP1 is expressed in the canine oviduct at the time and place of oocyte maturation and fertilization. *Mol. Reprod. Dev.* 81:972–982.

Salgaonkar N, Prakash D, Nawani NN, Kapadnis BP. 2015. Comparative studies on ability of N-acetylated chitooligosaccharides to scavenge reactive oxygen species and protect DNA from oxidative damage. *Indian J. Biotechnol.* 14:186–192.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31.

Smith SA, Robbins LW, Steiert JG. 1998. Isolation and characterization of a chitinase from the nine-banded armadillo, Dasypus novemcinctus. *J. Mammal.* 79:486–491.

Soneson C, Love MI, Robinson MD. 2016. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000 Res.* 4:1521.

Springer MS, Meredith RW, Teeling EC, Murphy WJ. 2013. Technical Comment on "The Placental Mammal Ancestor and the Post–K-Pg Radiation of Placentals." *Science* 341:613–613.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

Strobel S, Roswag A, Becker NI, Trenczek TE, Encarnaçao JA. 2013. Insectivorous bats digest chitin in the stomach using acidic mammalian chitinase. *PloS One* 8:e72770.

Tabata E, Itoigawa A, Koinuma T, Tayama H, Kashimura A, Sakaguchi M, Matoska V, Bauer PO, Oyama F. 2022. Noninsect-based diet leads to structural and functional changes of Acidic Chitinase in Carnivora. *Mol. Biol. Evol.* 39:msab331.

Tjoelker LW, Gosting L, Frey S, Hunter CL, Le Trong H, Steiner B, Brammer H, Gray PW. 2000. Structural and functional definition of the human chitinase chitin-binding domain. *J. Biol. Chem.* 275:514–520.

Tucker R. 1958. Taxonomy of the salivary glands of vertebrates. *Syst. Biol.* 7:74–83.

Vandewege MW, Sotero-Caio CG, Phillips CD. 2020. Positive selection and gene expression analyses from salivary glands reveal discrete adaptations within the ecologically diverse bat family Phyllostomidae. *Genome Biol. Evol.* 12:1419–1428.

Wang K, Tian S, Galindo-González J, Dávalos LM, Zhang Y, Zhao H. 2020. Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. *Mol. Ecol.* 29:4366–4381.

Wolf JBW. 2013. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Mol. Ecol. Resour.* 13:559–572.

Xie VC, Pu J, Metzger BP, Thornton JW, Dickinson BC. 2021. Contingency and chance erase necessity in the experimental evolution of ancestral proteins. *eLife* 10:e67336.

Yates AD, Achuthan P, Akanni W, Allen James, Allen Jamie, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, et al. 2020. Ensembl 2020. *Nucleic Acids Res.* 48:D682–D688.

Yusoff AM, Tan TK, Hari R, Koepfli K-P, Wee WY, Antunes A, Sitam FT, Rovie-Ryan JJ, Karuppannan KV, Wong GJ. 2016. De novo sequencing, assembly and analysis of eight different transcriptomes from the Malayan pangolin. *Sci. Rep.* 6:1–11.

Zhang F, Xu N, Yu Y, Wu S, Li S, Wang W. 2019. Expression profile of the digestive enzymes of *Manis javanica* reveals its adaptation to diet specialization. *ACS Omega* 4:19925–19933.

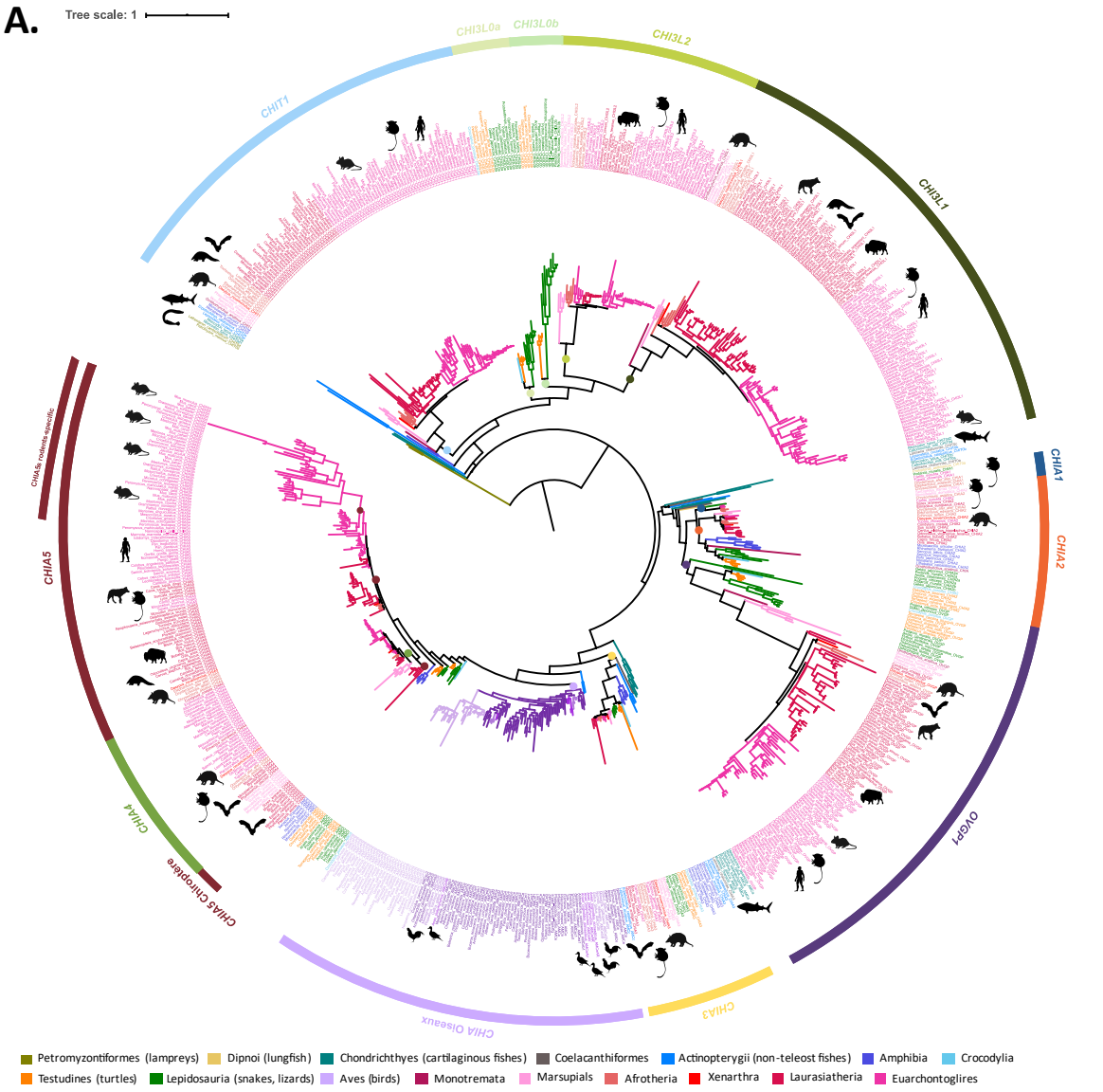
Annex: Evolutionary history of vertebrate chitinases

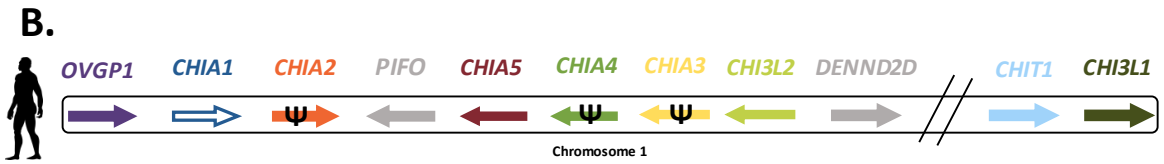
The following analysis was conducted during my Master degree to better understand the evolutionary history of chitinase genes in vertebrates beneficiating from the high amount of publicly available vertebrate genomes and using new phylogenetic approaches. In total, 939 chitinase sequences were retrieved from Genbank for 242 vertebrate species. Teleost fishes were not included as their specific whole genome duplication would have complexified the analyses. The overall evolutionary dynamics of this gene family was investigated through a reconciliation approach.

The evolution of the chitinase gene family (Fig Annex.1A) was characterized by the presence of numerous gene losses and few gene duplications giving birth to the main paralogues, and some duplications occurring more specifically in certain lineages consistent with what has been found previously (Bussink *et al*, 2007; Funkhouser and Aronson, 2007; Hussain and Wilson, 2013). The expansion of the family likely occurred after the divergence of lampreys with jawed vertebrates (Gnathostomata) as lamprey chitinases are placed as a sister-group to the rest of the tree. The first duplication gave birth to two clades: CHIT1-CHI3L and OVGP1-CHIA and occurred before the split between cartilaginous (Chondrichthyes) and bony (Osteichtyes) fishes. Within the CHIT1-CHI3L clade, CHIT1 was found only in mammals and another duplication led to the *CHI3L1* and *CHI3L2* paralogues restricted to mammals as well. The OVGP1/CHIA clade was characterized by an ancient duplication splitting OVGP1, CHIA1, and CHIA2 from CHIA3-5 confirming that CHIA1 and CHIA2 are more closely related than CHIA3-5, which was confirmed by synteny analyses (Hussain and Wilson, 2013; Emerling *et al*, 2018; Fig Annex.1B) suggesting that tandem duplications occurred during the evolution of the family. Several duplications led to the *CHIA1-5* paralogues in each of these two clades. CHIAs of birds formed a clade and no bird sequences were found in other clades, which suggests that numerous gene losses might have occurred in birds with only one *CHIA* gene retained that further expanded in certain lineages (*i.e*., Galliformes, Anseriformes, and Passeriformes) in agreement with previous results (Chen and Zhao, 2019). Finally, the duplication that led to the *CHIA4* and *CHIA5* paralogues seemed to be specific to mammals. *CHIA5* sequences of chiropterans were found closely related to *CHIA4* sequences which might be explained by

concerted evolution through gene conversion (Emerling *et al*, 2018). Finally, within the CHIA5 clade, duplications occurred within Muroidea where four paralogues can be identified (rodent-specific *CHIA5s*) consistent with previous results also highlighting rodent-specific duplications suggesting neo-functionalization potentially related to immune function (Hussain and Wilson, 2013).

During vertebrate evolutionary history, chitinase paralogues evolved toward different functions with some paralogues losing the ability to degrade and/or bind chitin (Fig Annex.1C). Chitinase-like proteins independently lost their ability to hydrolyze chitin due to substitutions in their chitinolytic site (Bussink *et al*, 2007; Funkhouser and Aronson, 2007; Hussain and Wilson, 2013). In mammals, *OVGP1* has been shown to have a role in reproduction (Buhi, 2002; Saint-Dizier *et al*, 2014; Algarra *et al*, 2016; Laheri *et al*, 2018), and *CHI3L1* and *CHI3L2* potentially participate in cell proliferation and immunity (Recklies *et al*, 2002; Lee *et al*, 2011; Areshkov *et al*, 2012). Rodent-specific chitinases have also been proposed to have immunity roles (Hussain and Wilson, 2013). Active chitinases can hydrolyze chitin and thus participate in digestion (*i.e.,* CHIAs), as well as defense against pathogens containing chitin and more generally be involved in immunity (Gordon-Thomson *et al*, 2009; Lee *et al*, 2011). It has been proposed that the expansion of the chitinase gene family has evolved with the development of the innate and adaptive immune systems (Funkhouser and Aronson, 2007). Along their evolution, chitinases and chitinase-like proteins likely evolved toward a specialized expression in certain tissues depending on their optimum pH (Bussink *et al*, 2007), which could be explained by amino acid changes in their chitinolytic sites (Olland *et al*, 2009). This was already proposed by Jeuniaux (1982) who identified three types of chitinases with different optimum pH corresponding to different digestive organs therefore suggesting diverse digestive functions for these chitinases.


A.
Tree scale: 1
CHI3L0a
CHI3L0b
CHI3L2
CHI3L1
CHIT1
CHIA1
CHIA2
OVGP1
CHIA3
CHIA Oligaux
CHIA4
CHIA5 Chiroptère
CHIA5 rodents-specific
CHIA5
Petromyzontiformes (lampreys)
Dipnoi (lungfish)
Chondrichthyes (cartilaginous fishes)
Coelacanthiformes
Actinopterygii (non-teleost fishes)
Amphibia
Crocodylia
Testudines (turtles)
Lepidosauria (snakes, lizards)
Aves (birds)
Monotremata
Marsupials
Afrotheria
Xenarthra
Laurasiatheria
Euarchontoglires


B.
OVGP1
CHIA1
CHIA2
PIFO
CHIA5
CHIA4
CHIA3
CHI3L2
DENND2D
CHIT1
CHI3L1
Ψ
Chromosome 1

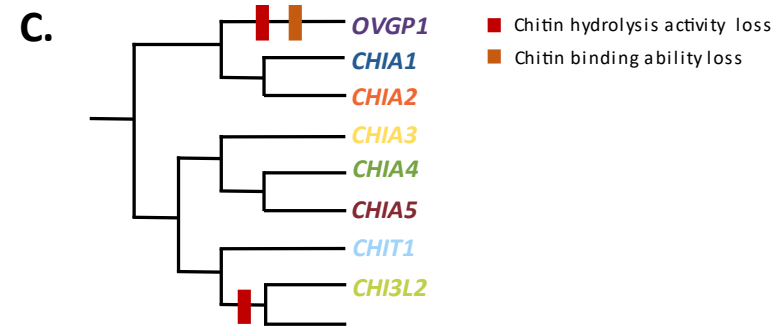

C.
OVGP1
CHIA1
CHIA2
CHIA3
CHIA4
CHIA5
CHIT1
CHI3L2
Chitin hydrolysis activity loss
Chitin binding ability loss

Figure Annex.1. Evolutionary history of vertebrate chitinases. A. Vertebrate chitinase gene tree (excluding teleost fishes) reconstructed using a maximum likelihood gene-species tree reconciliation approach. The outer circle indicates the nine paralogues. B. Chitinase gene synteny in human (*Homo sapiens*). Arrows represent genes with direction representing gene transcription direction. "Ψ" symbols indicate pseudogenes. Empty arrow (*i.e.*, *CHIA1*) indicates location of a gene that was not identified (*i.e.*, probably absent or pseudogenized, preventing its identification). C. Loss of chitin-degrading and chitin-binding abilities during the evolution of chitinase paralogues. Silhouettes were downloaded from phylopic.org.

Chapter references (excluding manuscripts' references)

Adler, E., Hoon, M.A., Mueller, K.L., Chandrashekar, J., Ryba, N.J., Zuker, C.S., 2000. A novel family of mammalian taste receptors. Cell 100, 693–702.

Albalat, R., Cañestro, C., 2016. Evolution by gene loss. Nature Reviews Genetics 17, 379–391.

Algarra, B., Han, L., Soriano-Úbeda, C., Avilés, M., Coy, P., Jovine, L., Jiménez-Movilla, M., 2016. The C-terminal region of OVGP1 remodels the zona pellucida and modifies fertility parameters. Sci Rep 6, 32556.

Allio, R., 2021. Phylogenomics and comparative genomics in ant-eating mammals. Agricultural sciences. Doctoral dissertation. University of Montpellier. English. NNT: 2021MONTG006. tel-03343948. https://theses.hal.science/tel-03343948

Allio, R., Tilak, M.-K., Scornavacca, C., Avenant, N.L., Kitchener, A.C., Corre, E., Nabholz, B., Delsuc, F., 2021. High-quality carnivoran genomes from roadkill samples enable comparative species delineation in aardwolf and bat-eared fox. eLife 10, e63167.

Allio, R., Teullet, S., Lutgen, D., Magdeleine, A., Koual, R., Tilak, M.-K., Thoisy, B. de, Emerling, C.A., Lefébure, T., Delsuc, F., 2023. Comparative transcriptomics reveals divergent paths of chitinase evolution underlying dietary convergence in ant-eating mammals. bioRxiv. https://doi.org/10.1101/2022.11.29.518312. In revision in Genome Biology and Evolution.

Álvarez-Carretero, S., Tamuri, A.U., Battini, M., Nascimento, F.F., Carlisle, E., Asher, R.J., Yang, Z., Donoghue, P.C.J., dos Reis, M., 2022. A Species-Level Timeline of Mammal Evolution Integrating Phylogenomic Data. Nature 602, 263–267.

Alves, L.Q., Ruivo, R., Fonseca, M.M., Lopes-Marques, M., Ribeiro, P., Castro, L.F.C., 2020. PseudoChecker: an integrated online platform for gene inactivation inference. Nucleic Acids Research 48, W321–W331.

Areshkov, P.O., Avdieiev, S.S., Balynska, O.V., Leroith, D., Kavsan, V.M., 2012. Two closely related human members of chitinase-like family, CHI3L1 and CHI3L2, activate ERK1/2 in 293 and U373 cells but have the different influence on cell proliferation. Int J Biol Sci 8, 39–48.

Bachmanov, A.A., Bosak, N.P., Lin, C., Matsumoto, I., Ohmoto, M., Reed, D.R., Nelson, T.M., 2014. Genetics of Taste Receptors. Curr Pharm Des 20, 2669–2683.

Balart-García, P., Cieslak, A., Escuer, P., Rozas, J., Ribera, I., Fernández, R., 2021. Smelling in the dark: Phylogenomic insights into the chemosensory system of a subterranean beetle. Molecular Ecology 30, 2573–2590.

Bardet, C., Delgado, S., Sire, J.-Y., 2010. *MEPE* evolution in mammals reveals regions and residues of prime functional importance. Cell. Mol. Life Sci. 67, 305–320.

Beier, S., Bertilsson, S., 2013. Bacterial chitin degradation—mechanisms and ecophysiological strategies. Front. Microbiol. 4, 149.

Bigiani, A., 2020. Does ENaC Work as Sodium Taste Receptor in Humans? Nutrients 12, 1195.

Boussau, B., and Scornavacca, C. (2020). Reconciling Gene trees with Species Trees. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, Phylogenetics in the Genomic Era, chapter No. 3.2, pp. 3.2:1–3.2:23. No commercial publisher | Authors open access book. The book is freely available at https://hal.inria.fr/PGE.

Buhi, W.C., 2002. Characterization and biological roles of oviduct-specific, oestrogen-dependent glycoprotein. Reproduction 123, 355–362.

Burki, F., Kaessmann, H., 2004. Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter flux. Nat Genet 36, 1061–1063.

Bussink, A.P., Speijer, D., Aerts, J.M.F.G., Boot, R.G., 2007. Evolution of Mammalian Chitinase(-Like) Members of Family 18 Glycosyl Hydrolases. Genetics 177, 959–970.

Carvalho, L.S., Pessoa, D.M.A., Mountford, J.K., Davies, W.I.L., Hunt, D.M., 2017. The Genetic and Evolutionary Drives behind Primate Color Vision. Frontiers in Ecology and Evolution 5, 34.

Casali, D.M., Martins-Santos, E., Santos, A.L.Q., Miranda, F.R., Mahecha, G.A.B., Perini, F.A., 2017. Morphology of the tongue of Vermilingua (Xenarthra: Pilosa) and evolutionary considerations. Journal of Morphology 278, 1380–1399.

Chandrashekar, J., Mueller, K.L., Hoon, M.A., Adler, E., Feng, L., Guo, W., Zuker, C.S., Ryba, N.J.P., 2000. T2Rs Function as Bitter Taste Receptors. Cell 100, 703–711.

Chandrashekar, J., Hoon, M.A., Ryba, N.J.P., Zuker, C.S., 2006. The receptors and cells for mammalian taste. Nature 444, 288–294.

Chandrashekar, J., Kuhn, C., Oka, Y., Yarmolinsky, D.A., Hummler, E., Ryba, N.J.P., Zuker, C.S., 2010. The cells and peripheral representation of sodium taste in mice. Nature 464, 297–301.

Chaudhari, N., Roper, S.D., 2010. The cell biology of taste. J Cell Biol 190, 285–296.

Chen, Y.-H., Zhao, H., 2019. Evolution of digestive enzymes and dietary diversification in birds. PeerJ 7, e6840.

Cheng, S.-C., Liu, C.-B., Yao, X.-Q., Hu, J.-Y., Yin, T.-T., Lim, B.K., Chen, W., Wang, G.-D., Zhang, C.-L., Irwin, D.M., Zhang, Z.-G., Zhang, Y.-P., Yu, L., 2023. Hologenomic insights into mammalian adaptations to myrmecophagy. National Science Review 10, nwac174.

Conant, G.C., Birchler, J.A., Pires, J.C., 2014. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. Current Opinion in Plant Biology, SI: Physiology and metabolism 19, 91–98.

Cortesi, F., Musilová, Z., Stieb, S.M., Hart, N.S., Siebeck, U.E., Malmstrøm, M., Tørresen, O.K., Jentoft, S., Cheney, K.L., Marshall, N.J., Carleton, K.L., Salzburger, W., 2015. Ancestral duplications and highly dynamic opsin gene evolution in percomorph fishes. Proceedings of the National Academy of Sciences 112, 1493–1498.

Dandrifosse, G., 1975. Purification of chitinases contained in pancreas or gastric mucosa of frog. Biochimie 57, 829–831.

De Oliveira Firmino, M., Pereira, H.C. da S., Carvalho, L.R.R.A., Guerra, R.R., 2020. External and digestive system morphology of the *Tamandua tetradactyla*. Anatomia, Histologia, Embryologia 49, 97–104.

Deeba, F., Shakir, H., Irfan, M., Qazi, J., 2016. Chitinase production in organisms: A review. Punjab University Journal of Zoology 31, 101–106.

Demuth, J.P., Hahn, M.W., 2009. The life and death of gene families. BioEssays 31, 29–39.

Dong, D., Jones, G., Zhang, S., 2009. Dynamic evolution of bitter taste receptor genes in vertebrates. BMC Evolutionary Biology 9, 12.

Doran, G.A., Allbrook, D.B., 1973. The tongue and associated structures in two species of African pangolins, *Manis gigantea* and *Manis tricuspis*. J Mammal 54, 887–899.

Doyon, J.-P., Ranwez, V., Daubin, V., and Berry, V. (2011). Models, algorithms and programs for phylogeny reconciliation. Brief Bioinform 12, 392–400.

Dube, N., Khan, S.H., Okafor, C.D., 2022. Ancestral sequence reconstruction for evolutionary characterization of proteins. Trends in Biochemical Sciences 47, 98–99.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., Aiden, E.L., 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356, 92–95

Dulai, K.S., Von Dornum, M., Mollon, J.D., Hunt, D.M., 1999. The Evolution of Trichromatic Color Vision by Opsin Gene Duplication in New World and Old World Primates. Genome Res. 9, 629–638.

Dyer, J., Salmon, K.S.H., Zibrik, L., Shirazi-Beechey, S.P., 2005. Expression of sweet taste receptors of the T1R family in the intestinal tract and enteroendocrine cells. Biochem Soc Trans 33, 302–305.

El Knidri, H., Belaabed, R., Addaou, A., Laajeb, A., Lahsini, A., 2018. Extraction, chemical modification and characterization of chitin and chitosan. International Journal of Biological Macromolecules 120, 1181–1189.

Emerling, C.A., 2017. Genomic regression of claw keratin, taste receptor and light-associated genes provides insights into biology and evolutionary origins of snakes. Mol Phylogenet Evol 115, 40–49.

Emerling, C.A., Delsuc, F., Nachman, M.W., 2018. Chitinase genes (*CHIAs*) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. Science Advances 4, eaar6478.

Emerling, C.A., Gibb, G.C., Tilak, M.-K., Hughes, J.J., Kuch, M., Duggan, A.T., Poinar, H.N., Nachman, M.W., Delsuc, F., 2023. Genomic data suggest parallel dental vestigialization within the xenarthran radiation. Peer Community Journal 3, e75.

Feng, P., Zheng, J., Rossiter, S.J., Wang, D., Zhao, H., 2014. Massive Losses of Taste Receptor Genes in Toothed and Baleen Whales. Genome Biology and Evolution 6, 1254–1265.

Ferreira-Cardoso, S., Delsuc, F., Hautier, L., 2019. Evolutionary Tinkering of the Mandibular Canal Linked to Convergent Regression of Teeth in Placental Mammals. Current Biology 29, 468-475.e3.

Ferreira-Cardoso, S., Fabre, P.-H., Thoisy, B. de, Delsuc, F., Hautier, L., 2020. Comparative masticatory myology in anteaters and its implications for interpreting morphological convergence in myrmecophagous placentals. PeerJ 8, e9690.

Ferreira-Cardoso, S., Claude, J., Goswami, A., Delsuc, F., Hautier, L., 2022. Flexible conservatism in the skull modularity of convergently evolved myrmecophagous placental mammals. BMC Ecology and Evolution 22, 87.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531–1545.

Fujimoto, W., Suzuki, M., Kimura, K., Iwanaga, T., 2002. Cellular Expression of the Gut Chitinase in the Stomach of Frogs *Xenopus laevis* and *Rana catesbeiana*. Biomedical Research 23, 91–99.

Funkhouser, J.D., Aronson, N.N., 2007. Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. BMC Evol Biol 7, 96.

Galindo, M.M., Voigt, N., Stein, J., van Lengerich, J., Raguse, J.-D., Hofmann, T., Meyerhof, W., Behrens, M., 2012. G Protein–Coupled Receptors in Human Fat Taste Perception. Chemical Senses 37, 123–139.

Garland, K. (2018). Evolution of bitter taste receptors in convergent myrmecophagous placentals. Master thesis. University of Montpellier.

Gordon-Thomson, C., Kumari, A., Tomkins, L., Holford, P., Djordjevic, J.T., Wright, L.C., Sorrell, T.C., Moore, G.P.M., 2009. Chitotriosidase and gene therapy for fungal infections. Cell. Mol. Life Sci. 66, 1116–1125.

Gooday, G.W., 1990. The Ecology of Chitin Degradation, in: Marshall, K.C. (Ed.), Advances in Microbial Ecology, Advances in Microbial Ecology. Springer US, Boston, MA, pp. 387–430.

Griffiths, M., 1968. Echidnas. Pergamon Press, Oxford; New York.

Hamid, R., Khan, M.A., Ahmad, M., Ahmad, M.M., Abdin, M.Z., Musarrat, J., Javed, S., 2013. Chitinases: An update. J Pharm Bioallied Sci 5, 21–29.

Han, Y.-Y., Ming, F., Wang, W., Wang, J.-W., Ye, M.-M., Shen, D.-L., 2006. Molecular evolution and functional specialization of chalcone synthase superfamily from Phalaenopsis Orchid. Genetica 128, 429–438.

Hayakawa, T., Suzuki-Hashido, N., Matsui, A., Go, Y., 2014. Frequent Expansions of the Bitter Taste Receptor Gene Repertoire during Evolution of Mammals in the Euarchontoglires Clade. Molecular Biology and Evolution 31, 2018–2031.

Hecker, N., Sharma, V., Hiller, M., 2019. Convergent gene losses illuminate metabolic and physiological changes in herbivores and carnivores. PNAS 116, 3036–3041.

Hiller, M., Schaar, B.T., Indjeian, V.B., Kingsley, D.M., Hagey, L.R., Bejerano, G., 2012. A "Forward Genomics" Approach Links Genotype to Phenotype using Independent Phenotypic Losses among Related Species. Cell Reports 2, 817–823.

Holt, C., Yandell, M., 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12, 491.

Hong, W., Zhao, H., 2014. Vampire bats exhibit evolutionary reduction of bitter taste receptor genes common to other bats. Proceedings of the Royal Society B: Biological Sciences 281, 20141079.

Hoon, M.A., Adler, E., Lindemeier, J., Battey, J.F., Ryba, N.J.P., Zuker, C.S., 1999. Putative Mammalian Taste Receptors: A Class of Taste-Specific GPCRs with Distinct Topographic Selectivity. Cell 96, 541–551.

Hu, L.-L., Shi, P., 2013. Smallest bitter taste receptor (T2Rs) gene repertoire in carnivores. Zoological Research 34, 13.

Huang, A.L., Chen, X., Hoon, M.A., Chandrashekar, J., Guo, W., Tränkner, D., Ryba, N.J.P., Zuker, C.S., 2006. The cells and logic for mammalian sour taste detection. Nature 442, 934–938.

Huang, Q.-S., Xie, X.-L., Liang, G., Gong, F., Wang, Y., Wei, X.-Q., Wang, Q., Ji, Z.-L., Chen, Q.-X., 2012. The GH18 family of chitinases: Their domain architectures, functions and evolutions. Glycobiology 22, 23–34.

Hughes, A.L., 1994. The evolution of functionally novel proteins after gene duplication. Proc. Biol. Sci. 256, 119–124.

Hughes, G.M., Boston, E.S.M., Finarelli, J.A., Murphy, W.J., Higgins, D.G., Teeling, E.C., 2018. The Birth and Death of Olfactory Receptor Gene Families in Mammalian Niche Adaptation. Mol Biol Evol 35, 1390–1406.

Hunt, D.M., Dulai, K.S., Cowing, J.A., Julliot, C., Mollon, J.D., Bowmaker, J.K., Li, W.-H., Hewett-Emmett, D., 1998. Molecular evolution of trichromacy in primates. Vision Research 38, 3299–3306.

Hussain, M., Wilson, J.B., 2013. New Paralogues and Revised Time Line in the Expansion of the Vertebrate GH18 Family. J Mol Evol 76, 240–260.

Innan, H., Kondrashov, F., 2010. The evolution of gene duplications: classifying and distinguishing between models. Nature Reviews Genetics 11, 97–108.

Ishimaru, Y., Inada, H., Kubota, M., Zhuang, H., Tominaga, M., Matsunami, H., 2006. Transient receptor potential family members PKD1L3 and PKD2L1 form a candidate sour taste receptor. Proceedings of the National Academy of Sciences 103, 12569–12574.

Jackson, S., Place, A.R., Seiderer, L.J., 1992. Chitin Digestion and Assimilation by Seabirds. The Auk 109, 758–770.

Janiak, M.C., Chaney, M.E., Tosi, A.J., 2018. Evolution of Acidic Mammalian Chitinase Genes (*CHIA*) Is Related to Body Mass and Insectivory in Primates. Molecular Biology and Evolution 35, 607–622.

Jeuniaux, C., 1959. Action consécutive de deux enzymes différents au cours de l'hydrolyse complète de la chitine. Archives of Physiology and Biochemistry 67, 115–116.

Jeuniaux, C., 1961a. Chitinase: An Addition to the List of Hydrolases in the Digestive Tract of Vertebrates. Nature 192, 135–136.

Jeuniaux, C., 1961b. Digestion de la chitine chez les oiseaux et les mammifères. Annales de la Société Royale Zoologique de Belgique 92, 27-45.

Jeuniaux, C., 1971. On some biochemical aspects of regressive evolution in animals. Biochemical Evolution and the Origin of Life. Ed E. Schoffeniels. North-Holland Publishing Company.

Jeuniaux, Ch., Dandrifosse, G., Micha, J.C., 1982. Caractères et évolution des enzymes chitinolytiques chez les vertébrés inférieurs. Biochemical Systematics and Ecology 10, 365–372.

Jiang, P., Josue, J., Li, X., Glaser, D., Li, W., Brand, J.G., Margolskee, R.F., Reed, D.R., Beauchamp, G.K., 2012. Major taste loss in carnivorous mammals. PNAS 109, 4956–4961.

Jiao, H., Zhang, L., Xie, H.-W., Simmons, N.B., Liu, H., Zhao, H., 2019. Trehalase Gene as a Molecular Signature of Dietary Diversification in Mammals. Mol Biol Evol 36, 2171–2183.

Jiao, H., Xie, H.-W., Zhang, L., Zhuoma, N., Jiang, P., Zhao, H., 2021. Loss of sweet taste despite the conservation of sweet receptor genes in insectivorous bats. PNAS 118.

Kataoka, S., Yang, R., Ishimaru, Y., Matsunami, H., Sévigny, J., Kinnamon, J.C., Finger, T.E., 2008. The Candidate Sour Taste Receptor, *PKD2L1*, Is Expressed by Type III Taste Cells in the Mouse. Chemical Senses 33, 243–254.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30, 3059–3066.

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30, 772–780.

Kinnamon, S.C., 2016. Chapter 15 - G Protein–Coupled Taste Transduction, in: Zufall, F., Munger, S.D. (Eds.), Chemosensory Transduction. Academic Press, pp. 271–285.

Kishida, T., Thewissen, J., Hayakawa, T., Imai, H., Agata, K., 2015. Aquatic adaptation and the evolution of smell and taste in whales. Zoological Letters 1, 9.

Kohl, S., Behrens, M., Dunkel, A., Hofmann, T., Meyerhof, W., 2013. Amino Acids and Peptides Activate at Least Five Members of the Human Bitter Taste Receptor Family. J. Agric. Food Chem. 61, 53–60.

Koonin, E.V., 2005. Orthologs, Paralogs, and Evolutionary Genomics. Annu. Rev. Genet. 39, 309–338.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35, 4453–4455.

Laheri, S., Modi, D., Bhatt, P., 2017. Extra-oviductal expression of oviductal glycoprotein 1 in mouse: Detection in testis, epididymis and ovary. J Biosci 42, 69–80.

Lee, C.G., Da Silva, C.A., Dela Cruz, C.S., Ahangari, F., Ma, B., Kang, M.-J., He, C.-H., Takyar, S., Elias, J.A., 2011. Role of chitin and chitinase/chitinase-like proteins in inflammation, tissue remodeling, and injury. Annu. Rev. Physiol. 73, 479–501.

Leurs, N., Martinand-Mari, C., Ventéo, S., Haitina, T., Debiais-Thibaud, M., 2021. Evolution of Matrix Gla and Bone Gla Protein Genes in Jawed Vertebrates. Frontiers in Genetics 12, 620659.

Li, X., Li, W., Wang, H., Cao, J., Maehashi, K., Huang, L., Bachmanov, A.A., Reed, D.R., Legrand-Defretin, V., Beauchamp, G.K., Brand, J.G., 2005. Pseudogenization of a Sweet-Receptor Gene Accounts for Cats' Indifference toward Sugar. PLOS Genetics 1, e3.

Li, X., Li, W., Wang, H., Bayley, D.L., Cao, J., Reed, D.R., Bachmanov, A.A., Huang, L., Legrand-Defretin, V., Beauchamp, G.K., Brand, J.G., 2006. Cats Lack a Sweet Taste Receptor. J Nutr 136, 1932S-1934S.

Li, X., Glaser, D., Li, W., Johnson, W.E., O'Brien, S.J., Beauchamp, G.K., Brand, J.G., 2009. Analyses of Sweet Receptor Gene (*Tas1r2*) and Preference for Sweet Stimuli in Species of Carnivora. Journal of Heredity 100, S90–S100.

Li, D., Zhang, J., 2014. Diet Shapes the Evolution of the Vertebrate Bitter Taste Receptor Gene Repertoire. Mol Biol Evol 31, 303–309.

Lindsay, G.J.H., Gooday, G.W., 1985. Chitinolytic enzymes and the bacterial microflora in the digestive tract of cod, *Gadus morhua*. Journal of Fish Biology 26, 255–265.

Liu, Z., Liu, G., Hailer, F., Orozco-terWengel, P., Tan, X., Tian, J., Yan, Z., Zhang, B., Li, M., 2016. Dietary specialization drives multiple independent losses and gains in the bitter taste gene repertoire of Laurasiatherian Mammals. Frontiers in Zoology 13, 28.

Liu, D., Costanzo, A., Evans, M.D.M., Archer, N.S., Nowson, C., Duesing, K., Keast, R., 2018. Expression of the candidate fat taste receptors in human fungiform papillae and the association with fat taste function. British Journal of Nutrition 120, 64–73.

Lopez, L.C., Morgan, E.D., 1997. Explanation of Bitter Taste of Venom of Ponerine Ant, Pachycondyla apicalis. J Chem Ecol 23, 705–712.

Lopez-Jimenez, N.D., Cavenagh, M.M., Sainz, E., Cruz-Ithier, M.A., Battey, J.F., Sullivan, S.L., 2006. Two members of the TRPP family of ion channels, *Pkd1l3* and *Pkd2l1*, are co-expressed in a subset of taste receptor cells. Journal of Neurochemistry 98, 68–77.

Luca, F., Perry, G.H., Di Rienzo, A., 2010. Evolutionary Adaptations to Dietary Changes. Annu Rev Nutr 30, 291–314.

Ma, J.-E., Li, L.-M., Jiang, H.-Y., Zhang, X.-J., Li, J., Li, G.-Y., Yuan, L.-H., Wu, J., Chen, J.-P., 2017. Transcriptomic analysis identifies genes and pathways related to myrmecophagy in the Malayan pangolin (*Manis javanica*). PeerJ 5, e4140.

Ma, J.-E., Jiang, H.-Y., Li, L.-M., Zhang, X.-J., Li, H.-M., Li, G.-Y., Mo, D.-Y., Chen, J.-P., 2019. SMRT sequencing of the full-length transcriptome of the Sunda pangolin (*Manis javanica*). Gene 692, 208–216.

Marsh, R.S., Moe, C., Lomneth, R.B., Fawcett, J.D., Place, A., 2001. Characterization of gastrointestinal chitinase in the lizard Sceloporus undulatus garmani (Reptilia: Phrynosomatidae). Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 128, 675–682.

Meredith, R.W., Gatesy, J., Murphy, W.J., Ryder, O.A., Springer, M.S., 2009. Molecular Decay of the Tooth Gene Enamelin (*ENAM*) Mirrors the Loss of Enamel in the Fossil Record of Placental Mammals. PLOS Genetics 5, e1000634.

Meyerhof, W., Batram, C., Kuhn, C., Brockhoff, A., Chudoba, E., Bufe, B., Appendino, G., Behrens, M., 2010. The molecular receptive ranges of human TAS2R bitter taste receptors. Chem Senses 35, 157–170.

Natarajan, C., Hoffmann, F.G., Weber, R.E., Fago, A., Witt, C.C., Storz, J.F., 2016. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. Science 354, 336–339.

Nei, M., Rooney, A.P., 2005. Concerted and Birth-and-Death Evolution of Multigene Families. Annu. Rev. Genet. 39, 121–152.

Nisa', C., Agungpriyono, S., Kitamura, N., Sasaki, M., Yamada, J., Sigit, K., 2010. Morphological Features of the Stomach of Malayan Pangolin, *Manis javanica*. Anatomia, Histologia, Embryologia 39, 432–439.

Ohno, S., 1970. Evolution by Gene Duplication. Springer, Berlin, Heidelberg.

Oka, Y., Butnaru, M., von Buchholtz, L., Ryba, N.J.P., Zuker, C.S., 2013. High salt recruits aversive taste pathways. Nature 494, 472–475.

Olland, A.M., Strand, J., Presman, E., Czerwinski, R., Joseph-McCarthy, D., Krykbaev, R., Schlingmann, G., Chopra, R., Lin, L., Fleming, M., Kriz, R., Stahl, M., Somers, W., Fitz, L., Mosyak, L., 2009. Triad of polar residues implicated in pH specificity of acidic mammalian chitinase. Protein Sci. 18, 569–578.

Pantages, E., Dulac, C., 2000. A Novel Family of Candidate Pheromone Receptors in Mammals. Neuron 28, 835–845.

Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528.

Patterson, B., 1975. The fossil aardvarks (Mammalia: Tubulidentata). Bull. Mus. Comp. Zool. 147, 185–237.

Peel, E., Silver, L., Brandies, P., Hayakawa, T., Belov, K., Hogg, C.J., Silver, L., Brandies, P., Hayakawa, T., Belov, K., Hogg, C.J., 2022. Genome assembly of the numbat (*Myrmecobius fasciatus*), the only termitivorous marsupial. Gigabyte 2022, 1–17.

Perry, G.H., Verrelli, B.C., Stone, A.C., 2005. Comparative Analyses Reveal a Complex History of Molecular Evolution for Human *MYH16*. Molecular Biology and Evolution 22, 379–382.

Perry, G.H., Kistler, L., Kelaita, M.A., Sams, A.J., 2015. Insights into hominin phenotypic and dietary evolution from ancient DNA sequence data. Journal of Human Evolution, Special Issue: Ancient DNA and Human Evolution 79, 55–63.

Prudent, X., Parra, G., Schwede, P., Roscito, J.G., Hiller, M., 2016. Controlling for Phylogenetic Relatedness and Evolutionary Rates Improves the Discovery of Associations Between Species' Phenotypic and Genomic Differences. Mol Biol Evol 33, 2135–2150.

Ramsey, I.S., DeSimone, J.A., 2018. Otopetrin-1: A sour-tasting proton channel. J Gen Physiol 150, 379–382.

Ranwez, V., Harispe, S., Delsuc, F., Douzery, E.J.P., 2011. MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. PLOS ONE 6, e22594.

Rathore, A.S., Gupta, R.D., 2015. Chitinases from Bacteria to Human: Properties, Applications, and Future Perspectives. Enzyme Research 2015, 791907.

Recklies, A.D., White, C., Ling, H., 2002. The chitinase 3-like protein human cartilage glycoprotein 39 (HC-gp39) stimulates proliferation of human connective-tissue cells and activates both extracellular signal-regulated kinase- and protein kinase B-mediated signalling pathways. Biochem J 365, 119–126.

Redford, K.H., 1987. Ants and Termites As Food, in: Genoways, H.H. (Ed.), Current Mammalogy. Springer US, Boston, MA, pp. 349–399.

Reiss, K.Z., 2001. Using Phylogenies to Study Convergence: The Case of the Ant-Eating Mammals. Integr Comp Biol 41, 507–525.

Romero, A., Leurs, N., Muñoz, D., Debiais-Thibaud, M., Marcellini, S., 2021. Divergent Expression of *SPARC*, *SPARC-L*, and *SCPP* Genes During Jawed Vertebrate Cartilage Mineralization. Frontiers in Genetics 12, 788346.

Roper, S.D., 2014. TRPs in Taste and Chemesthesis, in: Nilius, B., Flockerzi, V. (Eds.), Mammalian Transient Receptor Potential (TRP) Cation Channels: Volume II. Springer International Publishing, Cham, pp. 827–871.

Roper, S.D., Chaudhari, N., 2017. Taste buds: cells, signals and synapses. Nature Reviews Neuroscience 18, 485–497.

Saint-Dizier, M., Marnier, C., Tahir, M.Z., Grimard, B., Thoumire, S., Chastant-Maillard, S., Reynaud, K., 2014. *OVGP1* is expressed in the canine oviduct at the time and place of oocyte maturation and fertilization. Mol. Reprod. Dev. 81, 972–982.

Sato, J.J., Wolsan, M., 2012. Loss or major reduction of umami taste sensation in pinnipeds. Naturwissenschaften 99, 655–659.

Schmidt, J.O., 1986. Chemistry, Pharmacology, and Chemical Ecology of Ant Venoms. Elsevier, pp. 425–508.

Schrempf, D., and Szöllősi, G.J. (2020). The Sources of Phylogenetic Conflicts. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, Phylogenetics in the Genomic Era, chapter No. 3.1, pp. 3.1:1–3.1:23. No commercial publisher | Authors open access book. The book is freely available at https://hal.inria.fr/PGE.

Sclafani, A., Ackroff, K., Abumrad, N.A., 2007. CD36 gene deletion reduces fat preference and intake but not post-oral fat conditioning in mice. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 293, R1823–R1832.

Shan, L., Wu, Q., Wang, L., Zhang, L., Wei, F., 2018. Lineage-specific evolution of bitter taste receptor genes in the giant and red pandas implies dietary adaptation. Integrative Zoology 13, 152–159.

Sharma, V., Hecker, N., Roscito, J.G., Foerster, L., Langer, B.E., Hiller, M., 2018a. A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. Nat Commun 9, 1–9.

Sharma, V., Lehmann, T., Stuckas, H., Funke, L., Hiller, M., 2018b. Loss of *RXFP2* and *INSL3* genes in Afrotheria shows that testicular descent is the ancestral condition in placental mammals. PLOS Biology 16, e2005293.

Shigemura, N., Ohkuri, T., Sadamitsu, C., Yasumatsu, K., Yoshida, R., Beauchamp, G.K., Bachmanov, A.A., Ninomiya, Y., 2008. Amiloride-sensitive NaCl taste responses are associated with genetic variation of ENaC alpha-subunit in mice. Am J Physiol Regul Integr Comp Physiol 294, R66-75.

Silvent, J., Sire, J.-Y., Delgado, S., 2013. The dentin matrix acidic phosphoprotein 1 (*DMP1*) in the light of mammalian evolution. J Mol Evol 76, 59–70.

Siroski, P., Poletta, G., M.V, P., Ortega, H., Merchant, M., 2014. Presence of chitinase enzymes in crocodilians. Acta Herpetologica 9, 139–146.

Souza, C.P., Almeida, B.C., Colwell, R.R., Rivera, I.N.G., 2011. The Importance of Chitin in the Marine Environment. Mar Biotechnol 13, 823–830.

Springer, M.S., Emerling, C.A., Gatesy, J., Randall, J., Collin, M.A., Hecker, N., Hiller, M., Delsuc, F., 2019. Odontogenic ameloblast-associated (*ODAM*) is inactivated in toothless/enamelless placental mammals and toothed whales. BMC Evolutionary Biology 19, 31.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Stedman, H.H., Kozyak, B.W., Nelson, A., Thesier, D.M., Su, L.T., Low, D.W., Bridges, C.R., Shrager, J.B., Minugh-Purvis, N., Mitchell, M.A., 2004. Myosin gene mutation correlates with anatomical changes in the human lineage. Nature 428, 415–418.

Szöllősi, G.J., Tannier, E., Daubin, V., and Boussau, B. (2015). The Inference of Gene Trees with Species Trees. Syst Biol 64, e42–e62.

Tabata, E., Itoigawa, A., Koinuma, T., Tayama, H., Kashimura, A., Sakaguchi, M., Matoska, V., Bauer, P.O., Oyama, F., 2022. Noninsect-Based Diet Leads to Structural and Functional Changes of Acidic Chitinase in Carnivora. Mol Biol Evol 39.

Taniguchi, K., 2004. Expression of the sweet receptor protein, T1R3, in the human liver and pancreas. J Vet Med Sci 66, 1311–1314.

Teng, B., Wilson, C.E., Tu, Y.-H., Joshi, N.R., Kinnamon, S.C., Liman, E.R., 2019. Cellular and Neural Responses to Sour Stimuli Require the Proton Channel Otop1. Current Biology 29, 3647-3656.e5.

Thornton, J.W., 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. Nat Rev Genet 5, 366–375.

Tjoelker, L.W., Gosting, L., Frey, S., Hunter, C.L., Trong, H.L., Steiner, B., Brammer, H., Gray, P.W., 2000. Structural and Functional Definition of the Human Chitinase Chitin-binding Domain. J. Biol. Chem. 275, 514–520.

Torruella, G., de Mendoza, A., Grau-Bové, X., Antó, M., Chaplin, M.A., del Campo, J., Eme, L., Pérez-Cordón, G., Whipps, C.M., Nichols, K.M., Paley, R., Roger, A.J., Sitjà-Bobadilla, A., Donachie, S., Ruiz-Trillo, I., 2015. Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. Current Biology 25, 2404–2410.

Tu, Y.-H., Cooper, A.J., Teng, B., Chang, R.B., Artiga, D.J., Turner, H.N., Mulhall, E.M., Ye, W., Smith, A.D., Liman, E.R., 2018. An evolutionarily conserved gene family encodes proton-selective ion channels. Science 359, 1047–1050.

Wang, K., Tian, S., Galindo-González, J., Dávalos, L.M., Zhang, Y., Zhao, H., 2020. Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. Molecular Ecology 29, 4366–4381.

Wölfle, U., Elsholz, F.A., Kersten, A., Haarhaus, B., Müller, W.E., Schempp, C.M., 2015. Expression and functional activity of the bitter taste receptors TAS2R1 and TAS2R38 in human keratinocytes. Skin Pharmacol Physiol 28, 137–146.

Wu, V., Rozengurt, N., Yang, M., Young, S., Sinnett-Smith, J., Rozengurt, E., 2002. Expression of bitter taste receptors of the T2R family in gastrointestinal tract and enteroendocrine STC-1 cells. Proceedings of the National Academy of Sciences of the United States of America 99, 2392–7.

Yarmolinsky, D.A., Zuker, C.S., Ryba, N.J.P., 2009. Common Sense about Taste: From Mammals to Insects. Cell 139, 234–244.

Yokoyama, S., Yang, H., Starmer, W.T., 2008. Molecular Basis of Spectral Tuning in the Red- and Green-Sensitive (M/LWS) Pigments in Vertebrates. Genetics 179, 2037–2043.

Zhang, F., Xu, N., Yu, Y., Wu, S., Li, S., Wang, W., 2019. Expression Profile of the Digestive Enzymes of *Manis javanica* Reveals Its Adaptation to Diet Specialization. ACS Omega 4, 19925–19933.

Zhao, H., Yang, J.-R., Xu, H., Zhang, J., 2010a. Pseudogenization of the Umami Taste Receptor Gene *Tas1r1* in the Giant Panda Coincided with its Dietary Switch to Bamboo. Mol Biol Evol 27, 2669–2673.

Zhao, H., Zhou, Y., Pinto, C.M., Charles-Dominique, P., Galindo-González, J., Zhang, S., Zhang, J., 2010b. Evolution of the Sweet Taste Receptor Gene *Tas1r2* in Bats. Mol Biol Evol 27, 2642–2650.

Zhao, H., Xu, D., Zhang, S., Zhang, J., 2012. Genomic and genetic evidence for the loss of umami taste in bats. Genome Biol Evol 4, 73–79.

Zhao, H., Zhang, J., 2012. Mismatches between feeding ecology and taste receptor evolution: An inconvenient truth. PNAS 109, E1464–E1464.

Zhou, Y.H., Li, W.H., 1996. Gene conversion and natural selection in the evolution of X-linked color vision genes in higher primates. Mol Biol Evol 13, 780–783.

Zhou, Y., Shearwin-Whyatt, L., Li, J., Song, Z., Hayakawa, T., Stevens, D., Fenelon, J.C., Peel, E., Cheng, Y., Pajpach, F., Bradley, N., Suzuki, H., Nikaido, M., Damas, J., Daish, T., Perry, T., Zhu, Z., Geng, Y., Rhie, A., Sims, Y., Wood, J., Haase, B., Mountcastle, J., Fedrigo, O., Li, Q., Yang, H., Wang, J., Johnston, S.D., Phillippy, A.M., Howe, K., Jarvis, E.D., Ryder, O.A., Kaessmann, H., Donnelly, P., Korlach, J., Lewin, H.A., Graves, J., Belov, K., Renfree, M.B., Grutzner, F., Zhou, Q., Zhang, G., 2021. Platypus and echidna genomes reveal mammalian biology and evolution. Nature 592, 756–762.

Zimin, A.V., Puiu, D., Luo, M.-C., Zhu, T., Koren, S., Marçais, G., Yorke, J.A., Dvořák, J., Salzberg, S.L., 2017. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res. 27, 787–792.

CHAPTER II

ROLE OF THE GUT MICROBIOTA IN THE ADAPTATION TO MYRMECOPHAGY IN MAMMALS

II.1. Chapter introduction

Studying gut microbiota to understand convergent dietary adaptations in placental mammals

II.1.1. Host-microbiota coevolution

Microorganisms include bacteria, archaea, some unicellular eukaryotes (*i.e*., protists), some fungi (*i.e*., yeasts), and viruses. Microorganisms compose most of the diversity of living organisms, and represent the majority of the biomass on Earth (Whitman *et al*, 1998). They are present in all types of habitats including extreme environments as well as surfaces and internal compartments of other organisms such as plants and animals. The set of all microorganisms found in a defined environment constitutes a microbiota, while this environment and its microbiota including its genomic information defines a microbiome (Marchesi and Ravel, 2015; Berg *et al*, 2020). They play major roles in ecosystems, notably in global biogeochemical cycles but also through their interactions with other species with which they live in symbiosis (*i.e*., defined as a close and long-term interaction between two species). The symbionts can have both beneficial or deleterious effects on their host health. Based on these effects, different types of interactions have been described: i) mutualism happens when each partner beneficiates from the interaction (*e.g*., the term symbiosis is sometimes used to refer to mutualistic symbiosis), ii) parasitism occurs when one partner beneficiates from the interaction while having deleterious effects on the other, iii) commensalism corresponds to an interaction in which only one partner has benefits while the interaction has neither beneficial, nor deleterious effects on the other (*i.e*., neutral effects), and iv) amensalism occurs when the interaction is neutral for one partner and deleterious for the other. As the effects of these interactions are influenced by several biotic and abiotic factors, and therefore can shift in space and time, some authors have proposed to consider these interactions as a continuum from parasitism to mutualism (Hirsch, 2004; Leung and Poulin, 2008; Sachs *et al*, 2011).

Despite their huge diversity and important ecological roles, microbial taxonomic and functional diversities are still poorly known and many microbial species have not yet been described. Moreover, few microorganisms can be cultivated in laboratories making their description more difficult and, nowadays, mostly based on molecular information. Indeed, recent advances in next-generation sequencing have enabled the identification of new microbial species and the description of the functions they potentially fulfill by reconstructing their genomes which expanded our vision of the microbial diversity of many environments (*e.g.*, Parks *et al*, 2017; Pasolli *et al*, 2019; Vanni *et al*, 2022). Besides, new bacterial and archaeal phyla were described thanks to such analyses (*e.g.*, Brochier-Armanet *et al*, 2008; Rinke *et al*, 2013; Spang *et al*, 2015). One of these famous discoveries is the description of the Asgard archaeal superphylum (Zaremba-Niedzwiedzka *et al*, 2017) which completely changed our vision of the Tree of Life. Indeed, this superphylum comprises archaea carrying eukaryotic signature proteins which supports an archaeal origin of Eukaryotes (Zaremba-Niedzwiedzka *et al*, 2017). The discovery of new archaeal phyla and the identification of several proteins involved in cell composition and functioning that are shared between archaea and eukaryotes, now supports the two-domain Tree of Life classification with Eukaryotes emerging from Archaea instead of the three-domain classification (Cox *et al*, 2008; Guy and Ettema, 2011; Raymann *et al*, 2015; Spang *et al*, 2015; Zaremba-Niedzwiedzka *et al*, 2017). Furthermore, such advances in microbial molecular studies enabled us to answer new questions regarding the ecological and evolutionary role of microbes in diverse ecosystems. Initiatives like the Earth Microbiome Project (EMP; Gilbert *et al*, 2010) or the Human Microbiome Project (HMP; Gevers *et al*, 2012) were created to characterize global environmental and human microbial diversity and to understand the role microorganisms play in such ecosystems. More specifically, the HMP focuses on the interactions between the host and its microbiota in health and disease, and how external factors shape this interaction is of major public health interest. Such projects highlighted the numerous functions in which microbes participate within their host, and revealed their importance in host evolution.

In animals, the role of microbial symbionts in their evolution is increasingly recognized as they have coevolved for a long time (McFall-Ngai *et al*, 2013). Studying microbiome roles, especially from wild animals, is of upmost importance to understand the complex underlying mechanisms of such coevolution under diverse selective pressures (Amato, 2013; Hird, 2017). Captive and wild animals' microbiota are thus increasingly studied. More specifically, the gut microbiota has been the focus of several studies as it is involved in many host functions such as digestion (*e.g.*, energy uptake, detoxification), immunity (*e.g.*, pathogen defense), or behavior (*i.e.*, gut-brain axis) (Amato, 2013; Suzuki, 2017; Fig II.1). Several factors such as the host diet, physiology, and phylogeny but also environmental factors like anthropogenic factors, shape the gut microbiota taxonomic composition and functional structure (Alessandri *et al*, 2022; Fig II.1). The resulting changes affect many aspects of host fitness (Amato, 2013; Suzuki, 2017). Notably, the gut microbiota has expanded dietary niche possibilities for hosts, shaped host phenotypic plasticity to better adapt to changing environments, and contributed to the evolution of host systems to control microbiota assembly, such as the compartmentalization of the gut microbiota along the digestive tract or the evolution of the innate and adaptive immune systems (Moeller and Sanders, 2020; Mallot and Amato, 2021; Fig II.1). In return, the host affects the evolution of its symbiotic microbes, for instance, by controlling their dispersion via vertical and horizontal transmission, and its control systems can constraint microbial communities' diversification (Mallot and Amato, 2021; Fig II.1).

The host evolutionary history influences how microbial communities assemble and evolve within their host, sometimes resulting in closely related species sharing more similar microbiomes, a phenomenon called phylosymbiosis. Phylosymbiosis is particularly strong in non-flying mammals (Song *et al*, 2020) likely due to mammalian-specific traits, such as viviparity, parental care (*e.g.*, lactation), or their social behavior with numerous interactions between conspecifics, all facilitating the transmission of microbes between closely related individuals (Mallott and Amato, 2021). Flying mammals like bats make an exception as their gut microbiota is characterized by a weak phylosymbiosis and resemble those of birds probably due to physiological constraints related to flight (Song *et al*, 2020). Phylosymbiosis can be influenced by diverse processes such as host filtering (preventing some microbes to colonize a host), ecological drift (stochasticity might influence the acquisition or loss of microbes), host dispersal abilities (resulting in host facing different microbial communities

during their life and impacting microbial transfers between hosts), interactions between microbes, or bacterial transmission mode (Mazel *et al*, 2018, 2023; Moeller *et al*, 2017; Kohl, 2020). For instance, such processes can lead to higher specificity (*i.e*., phylosymbiosis) in internal compartments such as the gut (Mazel *et al*, 2018, 2023).

In mammals, the gut microbiota has played a major role in dietary diversification enabling transitions to novel carbon sources (Ley *et al*, 2008a; Moeller and Sanders, 2020). Consequently, many studies have focused on convergent dietary adaptations and the impact of host diet and investigated how host phylogeny shapes the gut microbiota (Ley *et al*, 2008b; Muegge *et al*, 2011; Groussin *et al*, 2017; Nishida and Ochman, 2018; Youngblut *et al*, 2019; Song *et al*, 2020; Thomas *et al*, 2022). Such studies have notably found effects of diet and phylogeny on the gut bacterial diversity, which increases from carnivory to herbivory (Ley *et al*, 2008b; Youngblut *et al*, 2019) and highlighted convergence in microbiota compositions and functions between species sharing similar diets (Muegge *et al*, 2011). Although less studied, gut archaea are key components of the gut microbiota playing major roles for the host, such as methanogenesis in ruminants. Recently, the gut archaeome has been the focus of several studies which expanded our knowledge of its taxonomic and functional diversity in humans (Chibani *et al*, 2021) and demonstrated that archaeal diversity and abundance are also shaped by the host phylogeny, diet, fiber content, and host physiology in vertebrates (Youngblut *et al*, 2021; Thomas *et al*, 2022).

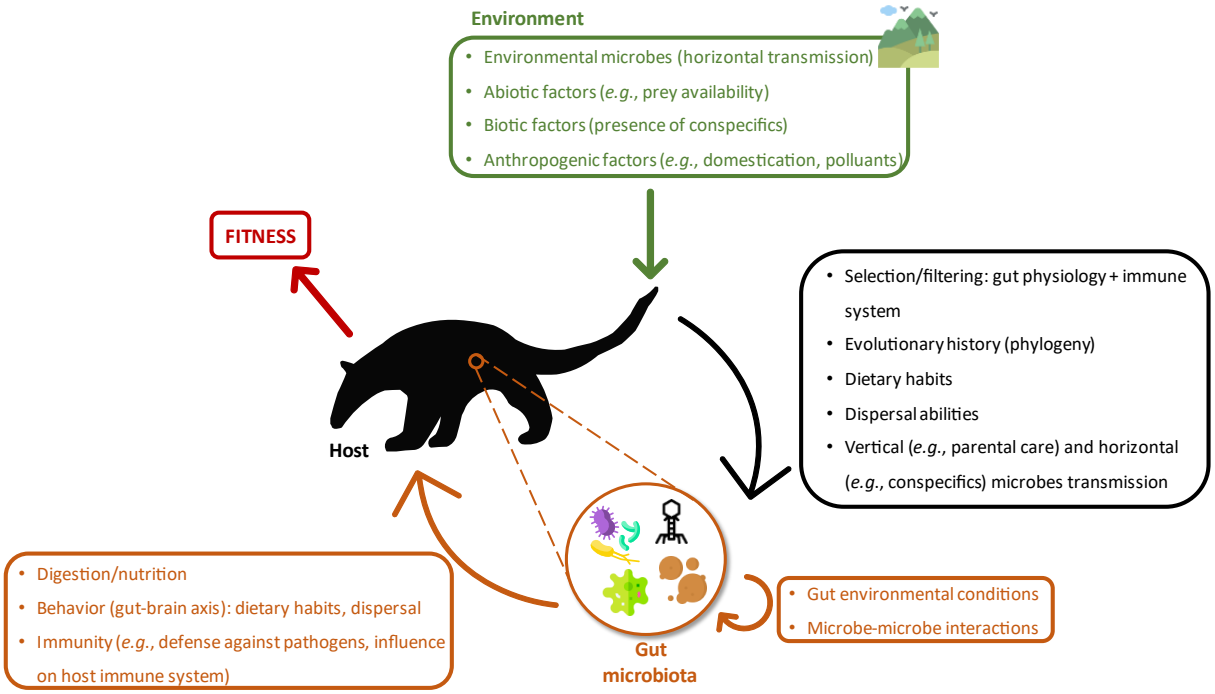


Figure II.1. Host-gut microbiota interactions and factors shaping their coevolution. The gut microbiota is the set of all the microorganisms present in the gut of an organism and includes prokaryotes (*i.e.*, bacteria, archaea), eukaryotes (*i.e.*, protists, fungi), and viruses. The gut microbiota (in brown) influences the host but also its own composition. The host (in black) influences and controls the gut microbiota composition through diverse mechanisms. External factors (in green) shape both the gut microbiota composition (thus its impact on the host) and the host lifestyle (that might result in changes in gut microbial communities). Overall, this coevolution between the host and its microbiota shaped by many internal and external factors influences the host fitness. Adapted from Amato (2013), Alessandri *et al* (2022), and Mallott and Amato (2021).

II.1.2. Molecular advances to study microbiomes and remaining challenges

Identification of microbes was first realized by classical microbiology methods based on isolation and culture of microbial cells, which prevent the study of the many non-cultivable microorganisms, a problem that was overcome through the amplification and sequencing of barcode gene sequences (*e.g.*, 16S rRNA for prokaryotes, 18S rRNA for eukaryotes, ITS rRNA for fungi). Interestingly, thanks to the PCR amplification this technique can help study rare organisms present at low abundances. These barcodes are then compared to reference databases containing sequences of taxonomically assigned microorganisms to identify microbial species (Marchesi and Ravel, 2015; Weissenbach and Sghir, 2016; Berg *et al*, 2020). Using barcodes relies on the assumption that interspecific variability of barcodes is higher than intraspecific variability (Valentini *et al*, 2009). Notably, taxonomic profiling of microbiota thanks to barcodes enables the identification of Operational Taxonomic Units (OTUs), which are clusters of sequences based on a similarity threshold (often 97%) (Sharpton, 2014; Breitwieser *et al*, 2019). However, OTUs, which were initially defined to deal with potential sequencing errors, reduce taxonomic resolution. For instance, closely related strains that might share high sequence similarity but play different roles cannot be studied. Besides sequencing errors can artificially increase the number of OTUs.

Therefore, new methods have been recently developed to deal with sequencing errors and focus on single-nucleotide resolution (*e.g.*, Eren *et al*, 2013; Callahan *et al*, 2016). Such new methods notably help to reconstruct Amplicon Sequence Variants (ASVs). ASVs better capture microbial diversity and make studies more comparable, and are now thought to replace OTUs (Callahan *et al*, 2017). Once defined, abundances of OTUs or ASVs can be assessed and statistical analyses conducted. For instance, alpha diversity (*i.e.*, microbial diversity of a specific microbial community) and beta diversity (*i.e.*, compositional dissimilarities between microbial communities) can be computed. Several microbial communities can then be compared and correlations with internal and external factors made to answer broad eco-evolutionary questions (Bharti and Grimm, 2021). For this purpose, pipelines like DADA2 (Callahan *et al*, 2016) or QIIME2 (Bolyen *et al*, 2019) have been designed to process and analyze amplicon sequencing data. Yet, the use of barcodes induces some

biases notably caused by horizontal gene transfers often occurring in prokaryotes, a lack of resolution sometimes preventing identification at low taxonomic scales, and the duplication of barcoding genes within some genomes biasing the estimation of relative abundances. Moreover, these barcodes represent a small part of microbial genomes preventing us to understand their functions, and prevent us to study plasmids or phages which play crucial roles in the functioning of microbiomes (Sharpton, 2014; Weissenbach and Sghir, 2016; Breitwieser *et al*, 2019). With the advance in high-throughput sequencing, the entire DNA present in a sample (*i.e*., metagenome) can now be sequenced to overcome some of these biases and allows us to characterize the microbiome in its entirety, both taxonomically and functionally (Sharpton, 2014; Breitwieser *et al*, 2019). Diverse methods and numerous bioinformatic tools have been developed to process and analyze such high amounts of complex metagenomic data: from the taxonomic assignment of sequences, to the assembly of microbial genomes, and the study of their functions (Sharpton, 2014; Breitwieser *et al*, 2019; Bharti and Grimm, 2021; Taş *et al*, 2021). Yet, many challenges remain and biases can be introduced at every step of a metagenomic workflow, from the sample preservation strategy (*e.g*., Blekhman *et al*, 2016; Song *et al*, 2016; Ma *et al*, 2020; Marotz *et al*, 2021), to the DNA extraction and library preparation methods (*e.g*., Costea *et al*, 2017; Yang *et al*, 2020; Poulsen *et al*, 2022; Trigodet *et al*, 2022), and the bioinformatic analyses (see below), which can all affect the resulting taxonomic and functional profiling of microbiomes. Such biases emphasize on the importance of controls at every step of a metagenomic workflow and especially during sample processing, standardized protocols and pipelines notably for comparative studies, and comparison of several tools to choose the best suited depending on the microbial communities studied.

Some of these main metagenomic analyses are illustrated in Figure II.2, which gives an example of a metagenomic workflow. This figure also presents the pipeline and programs (see the legend) used for most of the analyses conducted during this PhD and presented in the manuscript part II.2.

Metagenomic data allow the taxonomic profiling of microbial communities by comparing metagenomic sequences against multilocus reference databases or even whole microbial genomes whose taxonomy is known (Sharpton, 2014; Breitwieser *et al*, 2019; Fig II.2.b). Their abundances can be assessed based on the principle that the more reads map against the reference, the more abundant the taxa. However, identifying microbial taxa in samples from non-model organisms remains challenging because of the incompleteness of reference databases used for taxonomic profiling, which are often built on microbial taxa isolated from model organisms, resulting in only few taxa being identified. For instance, only 43% of reads on average mapped against reference databases in vertebrate gut metagenomes (Youngblut *et al*, 2020), and even in human gut metagenomes usually only about 60% of the sequences are classified (Pasolli *et al*, 2019). Studies comparing performances of different programs are useful to choose tools with the highest accuracy and that estimate few false positives (*i.e*., read assigned to the wrong taxa) and false negatives (*i.e*., reads not taxonomically assigned although present in the reference database), and emphasize on the importance of the reference database choice (*e.g*., Portik *et al*, 2022; Marić *et al*, 2023). To overcome problems encountered during taxonomic profiling, one solution can be to reconstruct microbial genomes from metagenomic data (*i.e*., metagenome-assembled genomes, MAGs) thanks to genome-resolved metagenomics (Sharpton, 2014; Breitwieser *et al*, 2019). Indeed, remapping reads against these MAGs can provide information about coverage and detection of each genome in the community. MAGs can also be placed in a phylogeny to assess their taxonomy and this is of particular interest when new microbial genomes not previously described are recovered. Moreover, genes and functions can be assigned to bacterial taxa and complete metabolic pathways can be identified to assess the potential functions of these microorganisms (Sharpton, 2014; Breitwieser *et al*, 2019; Berg *et al*, 2020).

To retrieve bacterial genomes, metagenomic reads are first assembled into contigs (Fig II.2 c). The assembly process can also be challenging notably because of sequencing errors, uneven coverage due to difference in taxa abundances, and the presence of intragenomic repeats as well as intergenomic repeats due to the presence of closely related strains in microbial communities (Ghurye *et al*, 2016; Bharti and Grimm, 2021). Combining different types of sequencing tools can help overcome these challenges. For instance, long-read

sequencing data obtained from Oxford Nanopore Technologies can be used to efficiently reconstruct more complete bacterial genomes, as their longer size (~5-10 kb as opposed to 100-300 bp for short Illumina reads) can facilitate the assembly process, in particular via helping to resolve repetitive regions (Ghurye *et al*, 2016; Bharti and Grimm, 2021). Long-read metagenomic sequencing has recently enabled the reconstruction of MAGs from diverse environmental and host-associated microbial communities (*e.g*., Lavrinienko *et al*, 2020; Arumugam *et al*, 2021; Singleton *et al*, 2021; Cuscó *et al*, 2022; Bickhart *et al*, 2022; Liu *et al*, 2022). Although improving, long reads still contain sequencing errors (~10-15% of errors at the beginning of my PhD as opposed to ~99% accuracy now; Oxford Nanopore Technologies), which can impact taxonomic assignation and gene content analyses (frameshifts can for instance bias open reading frame predictions). To improve MAGs quality, short reads are thus still useful to correct long reads through hybrid assembly or for polishing long-read assemblies (*e.g*., Arumugam *et al*, 2021; Chen *et al*, 2021; Singleton *et al*, 2021). Notably, polishing long read metagenomic assemblies with short reads can increase assembly quality by reducing the number of indels and improving gene predictions (Latorre-Pérez *et al*, 2020), and was the strategy used during my PhD (Fig II.2 c). As for taxonomic profiling, testing several assembly and polishing strategies is crucial to choose the best suited tools depending on the microbial communities studied and the computational resources available to reconstruct contiguous assemblies (*e.g*., few contigs, high N50) (Hu *et al*, 2020; Latorre-Pérez *et al*, 2020).

To recover bacterial genomes from the assembled metagenomes (*i.e*., MAGs), contigs belonging to the same genome are then clustered together, a step called genome binning (*i.e*., genome resolved-metagenomics; Fig II.2.c). In most programs, clustering of contigs is performed using the tetranucleotide frequency (assumed to be stable for a genome but different among genomes) and differential coverage (contigs of the same genome should have similar abundances and be consistent across samples) (*e.g*., CONCOCT, Alneberg, *et al*, 2014; MaxBin, Wu *et al*, 2016; metaBAT2, Kang *et al*, 2019; anvi'o, Eren *et al*, 2021). To assess genome bin quality, their redundancy and completion can be evaluated, for instance with programs such as anvi'o (Eren *et al*, 2021) or CheckM (Parks *et al*, 2015). Completion and redundancy are often based on the presence of single-copy core genes (SCGs). If all SCGs are found the completion of the bin is of 100%, and if they are found only once, there is 0% redundancy; else several copies of the same SCG can reflect potential contamination (*i.e*.,

contigs belonging to another genome). High-quality MAGs often have a completion higher than 90% and a redundancy lower than 5% but depending on the type of microbial communities studied and the research questions investigated, one can choose other thresholds. Because of the complexity of microbial communities and the close relatedness of certain genomes, genome binning can be challenging and can sometimes result in chimeric genome bins (*i.e*., with quite high level of contamination). Moreover, redundancy might not only reflect contamination but can be, for instance, the result of SCG duplications within a genome. anvi'o (Eren *et al*, 2021) presents the advantage that it offers the possibility to manually refine genome bins, and therefore identify cases of true contamination thanks to its graphical user-friendly interface. Such manual refinement helps to reconstruct high-quality genome bins.

To compare genome bins across samples, it can be useful to remove redundant ones as similar genomes can be obtained from different samples, and to keep only the ones with the highest quality. This step is called dereplication and is performed by computing the Average Nucleotide Identity (ANI) of genomes based on a threshold indicating the degree of similarity needed to consider two genomes identical (Fig II.2.d). Choosing an appropriate ANI depends on the analyses that will be conducted after. For instance, an ANI of 95% is often preferred for species-level dereplication whereas an ANI of 98% is recommended when mapping metagenomic reads against genome bins (Olm *et al*, 2020). Dereplication can also be useful to select the best genome bins generated from different tools. Once a set of high-quality microbial genomes (*i.e*., MAGs) have been generated, further analyses can be conducted, for example, to study and compare their gene content, taxonomy, and distribution across samples allowing the taxonomic and functional characterization of microbiomes (Fig II.2.e).

Genome resolved-metagenomics is a powerful approach to characterize diverse microbiomes and answer many eco-evolutionary questions. However, it does not capture the entirety of the microbial community as not all the genomes present in a sample can be assembled (*e.g*., not all reads are assembled, not all contigs are binned, some rare taxa need high sequencing depth, and complex and large genomes, for instance of unicellular eukaryotes, are difficult to assemble with these techniques). To fully characterize the taxonomic composition and functional potential of a microbiome, it is also important to consider the entire dataset, for instance, by using programs like MetaPhlAn (Segata *et al*,

2012) or HUMAnN (Abubucker *et al*, 2012) which respectively enable the taxonomic profiling and identification of microbial pathways in metagenomic datasets, or anvi'o (Eren *et al*, 2021), which for example enables the functional annotation of genes within contigs.

Finally, combined with other complementary *-omics* approaches (*e.g.*, metatranscriptomics, metaproteomics, metabolomics), metagenomics can reveal what microorganisms really do within a specific microbiome and increase our knowledge of the taxonomic and functional diversity of microbial communities (Breitwieser *et al*, 2019; Berg *et al*, 2020). Many studies are now integrating several approaches to understand the role of microbiomes in diverse environments (Nayfach *et al*, 2020; Shaffer *et al*, 2022) including the vertebrate gut microbiota (Youngblut *et al*, 2020; Levin *et al*, 2021). These studies have increased our knowledge of the taxonomic and functional diversity of such environments.

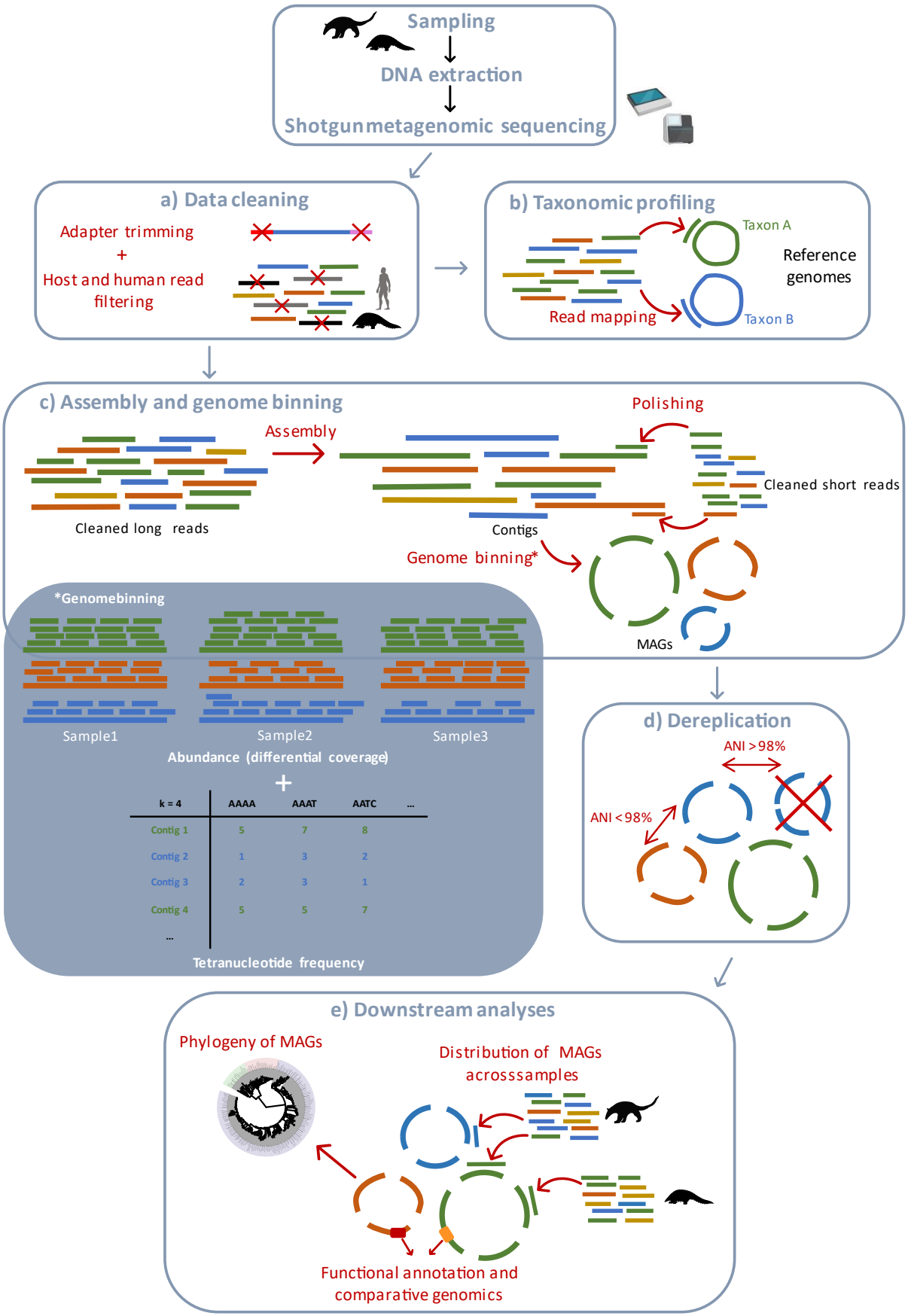

Sampling
DNA extraction
Shotgun metagenomic sequencing
a) Data cleaning
Adapter trimming
+
Host and human read filtering
b) Taxonomic profiling
Taxon A
Reference genomes
Read mapping
Taxon B
c) Assembly and genome binning
Assembly
Polishing
Cleaned long reads
Contigs
Cleaned short reads
Genome binning*
MAGs
*Genomebinning
Sample1
Sample2
Sample3
Abundance (differential coverage)
+
k = 4
AAAA
AAAT
AATC
...
Contig 1
5
7
8
Contig 2
1
3
2
Contig 3
2
3
1
Contig 4
5
5
7
...
Tetranucleotide frequency
d) Dereplication
ANI > 98%
ANI < 98%
e) Downstream analyses
Phylogeny of MAGs
Distribution of MAGs acrosssamples
Functional annotation and comparative genomics

Figure II.2. From metagenomic reads to metagenome-assembled genomes and the study of functions. After sampling, microbial DNA is extracted and metagenomic sequencing is performed; metagenomic data are then processed as follow: **a) data cleaning**: long reads are basecalled to convert the electric signal generated during sequencing into read sequences using Guppy (Oxford Nanopore Technologies, https://nanoporetech.com/). Both short and long reads are then cleaned to remove sequencing adapters, here respectively with FASTP (Chen *et al*, 2018) and Porechop (Wick *et al*, 2017), and host and human reads are filtered by mapping metagenomic reads against host and human reference genomes with bowtie2 (Langmead and Salzberg, 2012). **b) taxonomic profiling**: it can be done by mapping metagenomic reads against reference databases of clade-specific markers or whole genomes whose taxonomy is known, for instance with programs like MetaPhlAn (Segata *et al*, 2012) or MetaMaps (Dilthey *et al*, 2019) respectively for short and long reads; although this has not been done for the study presented in part II.2 for which taxonomy was assessed at the MAG level. **c) assembly and genome-resolved metagenomics:** long reads can be *de novo* assembled into contigs based on the overlap between reads; performed here with metaFlye (Kolmogorov *et al*, 2020) as it has been shown to perform better in assembling contiguous long-read metagenomes compared to other tools (Hu *et al*, 2020; Kolmogorov *et al*, 2020; Latorre-Pérez *et al*, 2020). Contigs are then polished by mapping short reads to correct potential long-read sequencing errors which was conducted here with Pilon (Walker *et al*, 2014). Genome binning: contigs belonging to the same genome are clustered together based on their tetranucleotide frequency and differential coverage; performed here with MetaBAT2 (Kang *et al*, 2019). Genome statistics, redundancy, and completion were assessed here with anvi'o (Eren *et al*, 2021) to evaluate genome bin quality. Good-quality MAGs can then be selected and further refined to remove contaminant contigs, for instance with programs like anvi'o. **d) dereplication**: Average Nucleotide Identity (ANI) between genomes can be computed and used to remove redundant MAGs; here with dRep (Olm *et al*, 2017). **e) downstream analyses**: to assess MAG taxonomy, clade-specific markers can be used, for example within anvi'o. MAGs can also be placed in a phylogeny based on universal markers which was performed here using PhyloPhlAn (Segata *et al*, 2013; Asnicar *et al*, 2020). Their distribution across samples can be studied by mapping metagenomic reads on MAGs within anvi'o. Functional annotation can also be done by first predicting Open Reading Frames (ORFs) (here performed with Prodigal; Hyatt *et al*, 2010), and comparing them to reference gene databases, enabling for instance to reconstruct metabolic pathways. Here, dbCAN2 (Zhang *et al*, 2018) was used to annotate carbohydrate-active enzymes (CAZymes) within MAGs. Analyzing gene contents between different genomes (*i.e.*, pangenomics) can for instance allow the identification of accessory genes, which might reflect potential specific adaptations in certain MAGs; this can be done, for instance, within anvi'o.

To study the potential of the gut microbiota of myrmecophagous mammals in prey digestion (see part II.2), we used genome-resolved metagenomics to reconstruct high-quality bacterial genomes with the aim of directly identifying bacteria carrying chitin-degrading enzymes. We combined long- (Oxford Nanopore Technologies) and short- (Illumina) read metagenomic sequencing and used short reads to polish our long-read assemblies that were then used for genome binning (Fig II.2). In order to use long-read sequencing, we needed to work with high molecular weight DNA that would allow us to sequence long DNA fragments. As we worked with field-collected fecal samples obtained either from dissections of roadkill specimens or by collecting samples directly in the field, our samples were not as fresh as those used in classic microbial studies on model/captive animals, implying more degraded DNA. Therefore, we optimized a DNA extraction protocol, presented in the following section, to extract high molecular weight bacterial DNA from these field-collected fecal samples preserved in ethanol (96%) and stored at -20°C suitable for long-read sequencing.

Protocol. High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing

The following protocol was optimized for the extraction of bacterial high-molecular weight DNA from field-collected fecal samples of ant- and termite-eating mammals suitable for Nanopore long-read shotgun metagenomic sequencing.

Samples used for the optimization of this protocol were obtained from myrmecophagous species that ingest large amounts of soil while foraging, resulting in feces having a peculiar texture. Before my PhD, several DNA extraction protocols were thus tested to extract bacterial DNA from these fecal samples. The “Genomic DNA from soil” kit, designed for DNA extraction of soil microbial communities, was the one yielding the most DNA and therefore chosen to work on myrmecophagous fecal samples. This kit first lyses samples by combining mechanistic (*i.e.*, with beads) and chemical (*i.e.*, with a lysis buffer) processes. Two

columns are then used during DNA extraction: the first aims at removing contaminants and the second, with silica, is used to bind DNA that is then eluted.

This DNA extraction protocol was optimized in three ways to ensure high molecular weight DNA suitable for Nanopore long-read sequencing:

- The choice of the lysis buffer to ensure high yields of DNA without degrading it too much. Two lysis buffers are included in this kit with an enhancer buffer to increase DNA yield. These two buffers are suitable for different types of samples. Both were tested with and without the enhancer buffer before choosing the combination used in the protocol.
- Two successive extractions were done, with the same starting material that was lysed again before the second extraction. This allowed us to obtain high-molecular weight DNA after the second extraction.
- The addition of a DNA purification step after the extraction with magnetic beads to remove short DNA fragments and to further ensure DNA quality and purity needed for long-read sequencing.

Using the second extraction, long-read sequencing on a MinION instrument yielded around 5 million reads per samples (between 10 and 15 Gb) with read N50 varying between 2 and 10 Kb. Illumina short-read sequencing was also done to compare metagenomic assemblies obtained from long- and short-read sequencing and confirmed that long-read assemblies were more contiguous. The two extractions did not differ in terms of microbial content extracted but rather in their quality (*i.e.*, short fragments after the first one and longer fragments after the second). We hypothesized that the first one yielded extracellular degraded DNA whereas the second extracted better-preserved intracellular bacterial DNA that can be accessed thanks to the second lysis. The first extraction can still be used for short-read sequencing, which combined with long-read data, can facilitate the recovery of high quality contiguous metagenomic assemblies useful to reconstruct bacterial genomes.

Personal contribution

I contributed, as a third author, as follow:

- Analyses: metagenome assemblies from first and second DNA extractions and comparison of their assembly statistics.
- Figure editing.
- Most of the writing.

This protocol can be found here:

Magdeleine, A., Tilak, M.-K., **Teullet, S**., Delsuc, F., 2021. High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing. protocol.io. dx.doi.org/10.17504/protocols.io.bvyin7ue.

Jun 28, 2021

High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing

Amandine Magdeleine[1], Marie-Ka Tilak[1], Sophie Teullet[1], Frédéric Delsuc[1]

[1]Institut des Sciences de l'Evolution de Montpellier (ISEM), CNRS, IRD, Université de Montpellier, Montpellier, France

dx.doi.org/10.17504/protocols.io.bvyin7ue

High molecular weight DNA extraction from all kingdoms
Tech. support email: See@each.protocol

Sophie Teullet
Institut des Sciences de l'Evolution de Montpellier

ABSTRACT

Long-read metagenomic sequencing enables the reconstruction of better-quality assemblies compared to short-read assemblies. Long-read assemblies are typically more contiguous and this is of interest to recover full bacterial genomes. Here, we propose a protocol to extract high molecular weight bacterial DNA from mammalian fecal samples for long-read sequencing on the portable MinION instrument. DNA was extracted from feces collected in the field or from roadkill specimens and preserved in ethanol (96%) and stored at -20°C. The DNA extraction was done according to the "Genomic DNA from soil" kit (NucleoSpin Soil, Macherey-Nagel) which was optimized in three ways: i) SL1 + SX lysis buffers were used because this combination gave a high yield and degraded DNA the less, ii) two successive extractions were performed to get long fragments of DNA and iii) an additional DNA sizing step was added after the extraction to remove the remaining short DNA fragments and ensure DNA purity. The comparison of the two successive extractions showed that the obtained reads did not differ in terms of taxonomy but only the second yielded high molecular weight DNA suitable for long-read sequencing. Resulting long-read assemblies were more contiguous than Illumina short-read ones with fewer contigs (~4 000 and ~50 000 respectively) and a higher N50 (~90 000 kb and ~10 000 kb respectively) illustrating the advantage of using long reads to reconstruct bacterial genomes from ethanol-preserved fecal samples.

DOI

dx.doi.org/10.17504/protocols.io.bvyin7ue

PROTOCOL CITATION

Amandine Magdeleine, Marie-Ka Tilak, Sophie Teullet, Frédéric Delsuc 2021. High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing. **protocols.io** https://dx.doi.org/10.17504/protocols.io.bvyin7ue

KEYWORDS

high molecular weight DNA extraction, long-read sequencing, metagenomics, field-collected fecal samples

CREATED

Jun 20, 2021

LAST MODIFIED

Jun 28, 2021

PROTOCOL INTEGER ID

50922

MATERIALS TEXT

Reagents

- DNA extraction kit: Genomic DNA from soil, NucleoSpin Soil, Macherey-Nagel; reference: 740780.50
- Library preparation kit: 1D Genomic DNA by Ligation (SQK-LSK109), Oxford Nanopore Technologies
- Ethanol 96-100%
- Agencourt AMPure XP beads

Consumables

- DNA LoBind tubes 1.5 and 2 mL
- Tips P10, P20, P200 and P1000

Equipment

- Precision scale
- Vortex
- Centrifuge mini spin
- Magnetic racks
- NanoDrop spectrophotometer
- Qubit

1- DNA extraction

1 DNA extraction is done using the "**Genomic DNA from soil**" kit (**NucleoSpin Soil, Macherey-Nagel**) from feces preserved in ethanol (96%) at -20°C. **Two successive extractions** are done to retrieve high molecular weight DNA. The protocol was optimized by using **Buffer SL1 + Enhancer SX**.

Before the extraction:

- Incubate Buffer SL1 for 5 min at 50°C before use.
- Add 100 mL of ethanol (96%) to Buffer SW2 before use.

1.1 Sample preparation

1.1.1- Weigh between 500 and 700 mg of feces in a 2 mL LoBind tube.
1.1.2- Centrifuge for 5 min at full speed and remove the supernatant (ethanol). Repeat 5 times.
1.1.3- Weigh the feces without ethanol. NB: between 300 and 500 mg of fecal material remains.
1.1.4- Transfer the feces into the NucleoSpin MN Bead Tube Type A.
1.1.5- Add 700 µL of Buffer SL1.

1.2 Sample lysis

1.2.1- Add 150 µL of Enhancer SX.
1.2.2- Vortex horizontally at full speed for 5 min.
1.2.3- Centrifuge for 2 min at 12.8 rpm.
1.2.4- Transfer the supernatant (~600 µL) into a new 1.5 mL LoBind tube. **Keep the pellet in the fridge for the second extraction**.
1.2.5- Add 150 µL of Buffer SL3 and vortex for 5 sec.

Citation: Amandine Magdeleine, Marie-Ka Tilak, Sophie Teullet, FrÃÂ©dÃÂ©ric Delsuc (06/28/2021). High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing. https://dx.doi.org/10.17504/protocols.io.bvyin7ue

1.2.6- Incubate for 10 min at 0-4°C.
1.2.7- Centrifuge for 1 min at 12.8 rpm.

Removal of inhibitors
1.2.8- Transfer 500 µL of supernatant onto a NucleoSpin Inhibitor Removal Column (red ring) placed in a collection tube (2 mL).
1.2.9- Centrifuge for 1 min at 12.8 rpm. Keep the supernatant in a new tube and repeat steps 1.2.8 and 1.2.9 with the remaining supernatant.
1.2.10- Add 250 µL of Buffer SB and vortex for 5 sec.

DNA binding on silica membrane
1.2.11- Transfer 550 µL of supernatant onto a NucleoSpin Soil Column (green ring) placed in a collection tube (2 mL).
1.2.12- Centrifuge for 1 min at 12.8 rpm.
1.2.13- Discard the liquid and replace the collection tube below the column.
1.2.14- Repeat steps 1.2.12 and 1.2.13 with the remaining supernatant.

1.3 Whashing step

1.3.1- Add 500 µL of Buffer SB to the NucleoSpin Soil Column.
1.3.2- Centrifuge for 30 sec at 12.8 rpm.
1.3.3- Discard the liquid and replace the collection tube below the column.

1.3.4- Add 550 µL of Buffer SW1 to the NucleoSpin Soil Column.
1.3.5- Centrifuge for 30 sec at 12.8 rpm.
1.3.6- Discard the liquid and replace the collection tube below the column.

1.3.7- Add 700 µL of Buffer SW2 to the NucleoSpin Soil Column.
1.3.8- Vortex for 2 sec.
1.3.9- Centrifuge for 30 sec at 12.8 rpm.
1.3.10- Discard the liquid and replace the collection tube below the column.

1.3.11- Repeat steps 1.3.7 to 1.3.10.

Dry silica membrane
1.3.12- Centrifuge for 2 min at 12.8 rpm.

1.4 DNA elution

1.4.1- Place a NucleoSpin Soil Column in a 1.5 mL LoBind tube.
1.4.2- Add 60 µL of Buffer SE to the column.
1.4.3- Incubate for 1 min at room temperature (RT) with the lid opened.
1.4.4- Centrifuge for 30 sec at 12.8 rpm.

For the second extraction:
- Use the pellet from step 1.2.4 kept in the fridge.
- Add 700 µL of Buffer SL1 and 150 µL of Buffer SX.
- Resuspend the pellet by vortexing.
- Follow the same steps as described above from 1.2.3 to 1.3.12.

For DNA elution:
- Place a NucleoSpin Soil Column in a 1.5 mL LoBind tube.
- Add **35 µL of Buffer SE** to the column.
- Incubate for 1 min at RT with the lid opened.
- Centrifuge for 30 sec at 12.8 rpm.

2- DNA quality and purity control

Citation: Amandine Magdeleine, Marie-Ka Tilak, Sophie Teullet, FrÃÂ©dÃÂ©ric Delsuc (06/28/2021). High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing. https://dx.doi.org/10.17504/protocols.io.bvyin7ue

2 Quantify the DNA with Qubit fluorometer (Broad Range) and NanoDrop spectrophotometer.
Qubit/Nanodrop ratio must be between 0.6 and 1 (below 0.6 DNA is not clean enough).

Check DNA quality by migration on a 1% agarose gel.

DNA from the first extraction is more degraded than DNA from the second extraction (Figure 1).

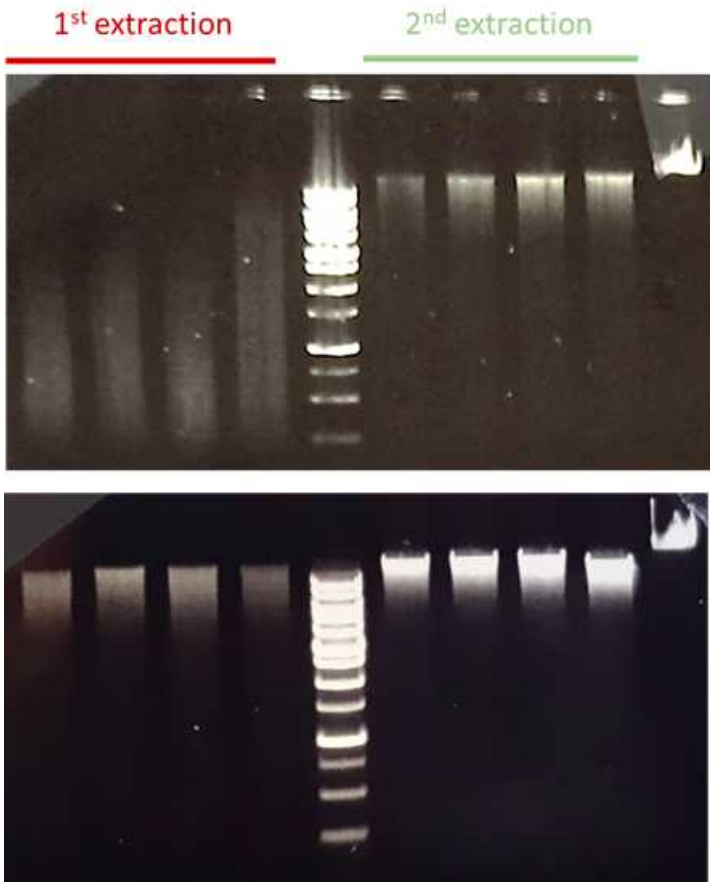


Figure 1. Agarose gel 1% in TAE 1x showing first and second DNA extractions from ethanol-preserved fecal samples of one ground pangolin (*Smutsia temminckii*, TS525, top) and one southern aardwolf (*Proteles cristatus*, TDR22, bottom). Wells 1-4: 1st extraction; well 5: 200 bp - 10 Kb ladder; wells 6-9: 2nd extraction and well 10: Lambda ladder (48.5 Kb).

3- DNA sizing with AMPure XP beads

3 This step enables to remove small DNA fragments and/or residual inhibitors.

The volume of beads (DNA volume*ratio) is adjusted according to sample degradation between 0.4 up to 1x.

3.1- Resuspend the pellet by gently shaking the tube with your finger.
3.2- Centrifuge briefly.
3.3- Incubate for 10 min at RT. In the meantime, prepare a fresh solution of ethanol 70%.
3.4- Place the tube on a magnetic rack for 4 min.
3.5- Remove the supernatant.
3.6- Add 200 µL of ethanol (70%) without disturbing the pellet.
3.7- Wait for 30 sec and remove the ethanol.
3.8- Repeat steps 3.6 and 3.7.
3.9- Centrifuge briefly, put the tube back on the magnetic rack and remove residual ethanol with P10.
3.10- Remove the tube from the magnetic rack.

Citation: Amandine Magdeleine, Marie-Ka Tilak, Sophie Teullet, FrÃÂ©dÃÂ©ric Delsuc (06/28/2021). High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing. https://dx.doi.org/10.17504/protocols.io.bvyin7ue

3.11- Air-dry for 1 min at RT.
3.12- Add 55 µL of H2O on the beads.
3.13- Resuspend the pellet by gently shaking the tube with your finger.
3.14- Incubate for 10 min at RT.
3.15- Place the tube on a magnetic rack.
3.16- Transfer the supernatant into a new LoBind tube.

Quantify DNA with Qubit fluorometer (Broad Range) and NanoDrop spectrophotometer. Check DNA quality by migration on a 1% agarose gel.

4- Library preparation

4 Library preparation is done according to the "**1D Genomic DNA by Ligation (SQK-LSK109)**" kit (**Oxford Nanopore Technologies**) instructions.

Quantify DNA with Qubit fluorometer (Broad Range).

5- MinION sequencing

5 Sequencing is done on MinION and Mk1C portable devices using R9 flowcells.

Comparison of first and second extractions

6 We performed short-read Illumina sequencing (150 PE) on both extractions to compare with long-read sequencing (done only on the second extraction).

Short reads (SR) were assembled for each extraction and compared to long-read assemblies (LR). Long-read assemblies were more contiguous (less contigs, higher N50, higher length of the longest contig) compared to short-read assemblies (Table 1). Short-read assemblies from first and second extractions were similar in terms of assembly statistics (Table 1).

Table 1. Comparison of short-read and long-read assemblies of 3 samples (two aardwolves, *Proteles cristatus* and one aardvark, *Orycteropus afer*). Short-read assemblies were done with *metaSPAdes* v3.15.2 (Nurk *et al*, 2017). Long-read assemblies were done with *metaFlye* v2.8.3 in strain mode (Kolmogorov *et al*, 2020). Assemblies were performed after removing adapters, low-quality reads and host and human reads. Contig statistics were computed for contigs >1 000 bp with *anvi'o* v7 (Eren *et al*, 2021).

Samples	Number of contigs			Contig N50 (bp)			Longest contig (bp)		
	SR 1	SR 2	LR	SR 1	SR 2	LR	SR 1	SR 2	LR
PRO TDR22	34 376	52 809	3 415	12 598	9 579	90 053	642 707	248 478	913 039
PRO TDR49	45 422	35 347	5 229	9 151	14 317	95 389	211 799	191 444	2 328 571
ORY TS513	82 582	62 637	4 655	8 091	9 055	82 470	953 702	959 158	3 962 646

Mapping short reads from the first extraction on assembled contigs of the second extraction from the same sample confirmed that the two extractions were similar (high percentage of mapped reads, Table 2). They were also similar to what was obtained from long-read sequencing and assembly (Table 2).

Table 2. Mapping of short reads (SR) from both extractions against contigs (C) of short-read (SR) and long-read (LR) assemblies. Short reads were mapped using *bowtie2* v2.3.4.1 (Langmead and Salzberg, 2012) with default parameters.

Citation: Amandine Magdeleine, Marie-Ka Tilak, Sophie Teullet, FrÃÂ©dÃÂ©ric Delsuc (06/28/2021). High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing. https://dx.doi.org/10.17504/protocols.io.bvyin7ue

Samples	% mapped reads (overall alignment rate)			
	SR against SR assemblies		SR against LR assemblies	
	SR_1 / C_{SR2}	SR_2 / C_{SR1}	SR_1 / C_{LR}	SR_2 / C_{LR}
PRO TDR22	96.36	92.99	93.56	91.65
PRO TDR49	94.45	96.04	92.31	94.7
ORY TS513	86.68	92.31	78.56	86.78

Conclusion

7 Taxonomic profiling of microbial communities associated with non-model organisms is challenging because of the incompleteness of reference databases (only few taxa could be identified). Extracting and reconstructing genomes from those unknown microbial taxa is thus of interest because they can then be placed in a phylogeny to assess their taxonomy. To do so, long reads present the advantage that they enable the reconstruction of more contiguous assemblies than short reads; this can then facilitate the recovery of contiguous bacterial genomes. With this optimized protocol, we showed that from mammalian field-collected fecal samples preserved in 96% ethanol, two successive extractions are needed to get high molecular weight DNA suitable for long-read sequencing. The two extractions only differ in the fact that the first extraction more likely yields extracellular degraded DNA whereas the second extraction allows accessing better preserved intracellular bacterial DNA. One remaining challenge with long reads is their higher error rate compared to short reads. Here, the high percentage of mapped reads on long-read assemblies suggests that short reads from each extraction could be used for polishing long-read assemblies. This will ensure that high quality contiguous assemblies are obtained and can then be used to reconstruct complete bacterial genomes.

References

8 Eren, A.M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S.E., Schechter, M.S., Fink, I., Pan, J.N., Yousef, M., Fogarty, E.C., *et al.* (2021). Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 6, 3–6.

Kolmogorov, M., Bickhart, D.M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S.B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T.P.L., *et al.* (2020). metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17, 1103–1110.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Res. 27, 824–834.

Citation: Amandine Magdeleine, Marie-Ka Tilak, Sophie Teullet, FrÃÂ©dÃÂ©ric Delsuc (06/28/2021). High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing. https://dx.doi.org/10.17504/protocols.io.bvyin7ue

II.1.3. Chitin-degrading bacteria in the gut microbiota of mammals

Chitin is a major structural component of many organisms (*e.g.*, arthropods, fungi) and is present in diverse environments (*i.e.*, soil, sediments, aquatic environments) (Swiontek Brzezinska *et al*, 2014; Beier and Bertilsson, 2013; Raimundo *et al*, 2021). To process these high ubiquitous amounts of chitin, chitinases are produced by a broad range of chitinolytic microorganisms including archaea, bacteria, as well as eukaryotes such as some fungi and planktonic communities, that have been isolated from diverse habitats (Gooday, 1990; Nawani and Kapadnis, 2003; Štrojsová and Vrba, 2005; Štrojsová and Dyhrman, 2008; Beier and Bertilsson, 2013; Swiontek Brzezinska *et al*, 2014). Chitin processing plays major roles in ecosystem functioning, notably in carbon and nitrogen cycles where it is mainly performed by chitinolytic bacteria (Swiontek Brzezinska *et al*, 2014; Beier and Bertilsson, 2013; Souza *et al*, 2011).

Fungal and bacterial chitinase genes are mainly found in the Glycoside Hydrolase (GH) 18, 19, 20, and 23 enzymes families (Adrangi and Faramarzi, 2013; Yan and Fong, 2015; Deeba *et al*, 2016), which are all carbohydrate-active enzymes (CAZymes) (Drula *et al*, 2022). Bacterial chitinases mainly belong to the GH18 enzyme family in which they are classified in three subfamilies (A, B, C) depending on their amino acid sequences defining their structure, and domain composition (*i.e.*, catalytic, binding domains), and with optimum pH varying from 5 to 8 (Adrangi and Faramarzi, 2013; Yan and Fong, 2015; Deeba *et al*, 2016). Bacterial chitinases are diverse and present in several bacterial clades. Genera in which chitinases have been mostly identified and studied include: *Streptomyces*, *Serratia*, *Clostridium*, *Vibrio*, *Arthrobacter*, *Pseudomonas*, *Bacillus*, *Aeromonas*, and *Klebsiella* (Swiontek Brzezinska *et al*, 2014; Yan and Fong, 2015; Deeba *et al*, 2016). The evolutionary history of bacterial chitinases is complex. Time-calibrated phylogenetic analyses have revealed a probable origin of bacterial chitinases from a Horizontal Gene Transfer (HGT) from fungi to marine bacteria around 600 Mya (Gruen *et al*, 2019). The evolution of chitinases within bacteria is characterized by multiple HGT events that dispersed chitinase genes in several bacterial lineages notably to terrestrial ones (Gruen *et al*, 2019). The propagation of these genes in terrestrial bacteria has been estimated to have occurred in the early Paleozoic, coinciding and probably linked to arthropod terrestrialization resulting in an increase in chitin availability (Gruen *et al*, 2019),

which is consistent with their crucial role in the processing of this polymer in diverse environments.

As mentioned previously, the gut microbiota of mammals has played a major role in their dietary diversification (see part II.1.1). Many mammals ingest prey containing high amount of chitin such as insects with their chitinous exoskeleton or crustaceans whose shell is composed of chitin (Gooday, 1990; Hamid *et al*, 2013; Rathore and Gupta, 2015; El Knidri *et al*, 2018). These hard chitinous structures need to be digested to access nutrients. As discussed in Chapter I, endogenous chitinases produced by the host organism itself can participate in chitin digestion but its associated microbiota can also play a role. The human gut microbiota contains a diversity of carbohydrate-active enzymes, which play a major role in the digestion of complex polysaccharides (El Kaoutari *et al*, 2013; Wardman *et al*, 2022). Notably, microbial GH18 are involved in glycan metabolism and have been identified in *Bacteroidetes* (Wardman *et al*, 2022). Jeuniaux (1950) was one of the first to investigate the presence of microbial chitin-degrading enzymes in the digestive tract of diverse animals and identified a microbial chitinase in snails. In other animals, chitinolytic microbes have been identified in the digestive tract of several species confirming their potential role in digestion (Stevens and Hume, 1998). For instance, they have been identified in the digestive tract of fishes (Goodrich and Morita, 1977; Itoi *et al*, 2006), lizards (Pardosi *et al*, 2018), krill-eating seabirds (Stemmier *et al*, 1984 cited in Olsen and Mathiesen, 1997), and several mammalian species having diverse diets (Šimůnek *et al*, 2001), as well as mammals with chitin-rich diets such as insectivorous bats (Whitaker *et al*, 2004; Irulan *et al*, 2011), insectivorous monkeys (Macdonald *et al*, 2014), crustacean-eating whales (Olsen *et al*, 1996; Sanders *et al*, 2015), or rodents fed with a chitin-enriched diet (Kohl *et al*, 2022).

Case study: gut chitinolytic bacteria in myrmecophagous mammals

The aforementioned studies therefore suggest that other species having chitin-rich diets might also beneficiate from their gut microbial symbionts to digest their prey. Myrmecophagous mammals have a diet composed almost exclusively of ants and/or termites (Redford, 1987) and need to digest important quantities of chitin. Their gut microbiota might thus have played an important role in their convergent adaptation to this diet. Taxonomic analyses based on 16S rRNA barcode sequences (Delsuc *et al*, 2014) have highlighted convergences in the gut microbiota composition of myrmecophagous species with some bacterial taxa, such as *Blautia* (Lachnospiraceae), *Streptococcus* (Streptococcaceae), or *Lactobacillus* (Lactobacillaceae) being significantly more abundant in myrmecophagous species compared to their non-myrmecophagous sister species. Yet, questions remained regarding the role these symbionts play in the digestion of the chitinous exoskeleton of prey. This was partially answered by metagenomic studies focusing on the gut microbiota of the Malayan pangolin (*Manis javanica*), and giant anteater (*Myrmecophaga tridactyla*), which identified microbial chitinases associated to specific chitinolytic bacteria (Ma *et al*, 2018; Cheng *et al*, 2023). Ma *et al*, (2018) identified GH18 and GH19 genes in 37 gut bacterial species of the Malayan pangolin, notably *Enterobacter cloacae*, *Lactococcus lactis*, *Chitimonas koreensis*, *Enterococcus faecalis*, and *Chitinophaga pinensis* that could be involved in chitin degradation, highlighting the potential role of the gut microbiota in prey digestion in this species. Cheng *et al* (2023) revealed enrichment in chitinolytic bacteria carrying chitin-degrading enzymes in three myrmecophagous species compared to non-myrmecophagous ones confirming previous results on 16S rRNA data (Delsuc *et al*, 2014), and suggesting potential patterns of convergence.

These results therefore exposed the potential role of chitinolytic symbionts in prey digestion in two highly specialized myrmecophagous species. However, the functions that gut symbionts play in other myrmecophagous species have not been studied. Moreover, comparative approaches are lacking to fully understand how these species convergently adapted to this highly specialized diet and to decipher whether similar adaptive mechanisms were involved between the different species. The following subparts of this chapter will focus on the role of the gut microbiota in the adaptation to myrmecophagy in placental mammals, which was the main subject of my PhD.

Part II.2 focuses on ant- and termite-eating species to i) determine whether chitinolytic bacteria are present in their gut microbiota and, if yes, ii) compare their distribution among myrmecophagous species to understand whether the same bacterial species carrying the same functions (*i.e.*, ability to degrade chitin) have been independently recruited in the different host species. Part II.3 presents a dataset that was assembled at the end of my PhD, and some of the analyses that could be performed, to understand the role of the gut microbiota in chitin digestion in placental mammals by including myrmecophagous species as well as other less specialized insectivorous species and other species with chitin-rich diets such as marine mammals. Moreover, including non-insectivorous species closely related to myrmecophagous ones, such as folivorous sloths (sister-group of anteaters) and carnivorous hyenas (sister-group of aardwolves), should help reveal potential patterns of convergence in the gut microbiota of myrmecophagous species compared to their non-myrmecophagous sister species.

Overall, these studies should help deciphering the role of the gut microbiota of myrmecophagous mammals in adapting to this highly specialized diet, and more generally contributing to our understanding of the complex mechanisms leading to convergent phenotypes.

II.2. Potential role of the gut microbiota in prey digestion in myrmecophagous mammals

The following article focuses on identifying gut chitinolytic symbionts in ant- and termite-eating placental mammals to unravel the potential role of their gut microbiota in prey digestion and more generally in their adaptation toward this highly specialized diet. To further understand the underlying adaptive mechanisms of this convergent evolution, this study focuses on comparing the distribution of gut chitinolytic bacteria among myrmecophagous species to assess whether shared (*i.e.*, potentially convergently recruited) or specific (*i.e.*, probably constrained by phylogenetic inheritance) bacteria are involved in prey digestion. To answer these questions 29 gut metagenomes of myrmecophagous species were sequenced by combining long- and short-read shotgun metagenomic sequencing and used to reconstruct more than 300 high-quality bacterial genomes using genome-resolved metagenomic approaches. Almost 400 chitin-degrading genes were identified in 132 genomes suggesting chitinolytic bacteria are present in the gut microbiota of these mammalian species and have the potential to digest their insect prey. Comparison of the distribution of these chitinolytic bacteria among myrmecophagous species revealed patterns of shared and specific bacteria.

Manuscript. Metagenomics uncovers dietary adaptations for chitin-digestion in the gut microbiota of myrmecophagous mammals

Personal contribution

I contributed, as the first author, as follow:

- Conceptualization of the study.
- Dissections and feces sampling of the specimens from French Guiana.
- DNA extractions, long-read library preparation, and long-read sequencing.
- Analyses.
- Figure editing.
- Writing of the first draft.
- Reading and editing.

This manuscript has been **published in mSystems and can be found here:**

Teullet, S., Tilak, M.-K., Magdeleine, A., Schaub, R., Weyer, N.M., Panaino, W., Fuller, A., Loughry, W.J., Avenant, N.L., Thoisy, B. de, Borrel, G., Delsuc, F., 2023. Metagenomics uncovers dietary adaptations for chitin digestion in the gut microbiota of convergent myrmecophagous mammals. mSystems 8, e00388-23. https://doi.org/10.1128/msystems.00388-23.

AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems

| Open Peer Review | Evolution | Research Article

Metagenomics uncovers dietary adaptations for chitin digestion in the gut microbiota of convergent myrmecophagous mammals

Sophie Teullet,[1] Marie-Ka Tilak,[1] Amandine Magdeleine,[1] Roxane Schaub,[2,3] Nora M. Weyer,[4] Wendy Panaino,[4,5] Andrea Fuller,[4] W. J. Loughry,[6] Nico L. Avenant,[7] Benoit de Thoisy,[8,9] Guillaume Borrel,[10] Frédéric Delsuc[1]

AUTHOR AFFILIATIONS See affiliation list on p. 19.

ABSTRACT In mammals, myrmecophagy (ant and termite consumption) represents a striking example of dietary convergence. This trait evolved independently at least five times in placentals with myrmecophagous species comprising aardvarks, anteaters, some armadillos, pangolins, and aardwolves. The gut microbiome plays an important role in dietary adaptation, and previous analyses of 16S rRNA metabarcoding data have revealed convergence in the composition of the gut microbiota among some myrmecophagous species. However, the functions performed by these gut bacterial symbionts and their potential role in the digestion of prey chitinous exoskeletons remain open questions. Using long- and short-read sequencing of fecal samples, we generated 29 gut metagenomes from nine myrmecophagous and closely related insectivorous species sampled in French Guiana, South Africa, and the United States. From these, we reconstructed 314 high-quality bacterial genome bins of which 132 carried chitinase genes, highlighting their potential role in insect prey digestion. These chitinolytic bacteria belonged mainly to the family Lachnospiraceae, and some were likely convergently recruited in the different myrmecophagous species as they were detected in several host orders (i.e., *Enterococcus faecalis*, *Blautia* sp.), suggesting that they could be directly involved in the adaptation to myrmecophagy. Others were found to be more host-specific, possibly reflecting phylogenetic constraints and environmental influences. Overall, our results highlight the potential role of the gut microbiome in chitin digestion in myrmecophagous mammals and provide the basis for future comparative studies performed at the mammalian scale to further unravel the mechanisms underlying the convergent adaptation to myrmecophagy.

IMPORTANCE Myrmecophagous mammals are specialized in the consumption of ants and/or termites. They do not share a direct common ancestor and evolved convergently in five distinct placental orders raising questions about the underlying adaptive mechanisms involved and the relative contribution of natural selection and phylogenetic constraints. Understanding how these species digest their prey can help answer these questions. More specifically, the role of their gut microbial symbionts in the digestion of the insect chitinous exoskeleton has not been investigated in all myrmecophagous orders. We generated 29 new gut metagenomes from nine myrmecophagous species to reconstruct more than 300 bacterial genomes in which we identified chitin-degrading enzymes. Studying the distribution of these chitinolytic bacteria among hosts revealed both shared and specific bacteria between ant-eating species. Overall, our results highlight the potential role of gut symbionts in the convergent dietary adaptation of myrmecophagous mammals and the evolutionary mechanisms shaping their gut microbiota.

KEYWORDS convergent evolution, myrmecophagy, mammals, gut microbiota, chitinases, metagenomics, genome assembly

Editor Sarah M. Hird, University of Connecticut, Storrs, Connecticut, USA

Ad Hoc Peer Reviewer Connie Rojas, University of California, Davis, Davis, California, USA

Address correspondence to Sophie Teullet, sophie.teullet@umontpellier.fr, or Frédéric Delsuc, frederic.delsuc@umontpellier.fr.

The authors declare no conflict of interest.

See the funding table on p. 20.

Received 20 April 2023
Accepted 19 June 2023
Published 31 August 2023

In mammals, the gut microbiota has played a major role in dietary diversification, enabling transitions to novel carbon sources (1, 2). Several factors such as host diet, physiology, genetics, and phylogeny, but also environmental factors, shape the taxonomic composition and functional structure of the gut microbiota (3). Many studies have focused on convergent dietary adaptations and the effects of host diet, investigating how host phylogeny shapes the gut microbiota (4–10). A striking example of convergent dietary specialization is the adaptation to myrmecophagy in placental mammals. Myrmecophagous species are characterized by a diet composed of at least 90% of ants and/or termites (11). A total of 22 placental species evolved toward this diet and are found in five independent orders: Tubulidentata (aardvarks), Pilosa (anteaters), Cingulata (tolypeutine armadillos), Pholidota (pangolins), and Carnivora (aardwolves) (12–14). Their divergent evolutionary histories raise the question of how these species convergently adapted to the myrmecophagous diet and whether similar adaptive mechanisms were involved between the different species. Myrmecophagous species need to digest the chitinous exoskeleton of their prey to extract nutrients. Mammals carry chitinase genes in their genomes, which participate in chitin digestion (15–18), but their gut microbiota might also have played an important role in facilitating prey digestion (19–21).

Indeed, chitinolytic bacteria are present in a diversity of environments (i.e., soils, sediments, and aquatic environments) where they ensure chitin degradation and play an important role in the carbon cycle (22–25). They also have been identified in the digestive tract of mammals where they could participate in prey digestion (26–30). Taxonomic analyses based on 16S rRNA sequences have highlighted similarities in the gut microbiota composition of myrmecophagous species (19) compared to their non-myrmecophagous sister species. For instance, genera such as *Blautia* (Lachnospiraceae), *Streptococcus* (Streptococcaceae), *Peptococcus* (Peptococcaceae), or *Eubacterium* (Lachnospiraceae) were found to be significantly more abundant in the gut microbiota of myrmecophagous species than in their sister species. Focusing on the gut archaeome of placentals, *Methanobrevibacter* has been found almost only in myrmecophagous species (10). Yet, little is known regarding the role these symbionts play in the digestion of the chitinous exoskeleton of their prey.

Microbial chitinases and *N*-acetylglucosaminidases are carbohydrate-active enzymes (CAZymes) (31), which ensure the degradation of chitin into chitosan and are mostly found in the glycosyl hydrolases (GH) families 18, 19, and 20 (23, 32). Since bacterial chitinases mainly belong to the GH18 family (32–35), they represent relevant candidates to assess the potential of the gut microbiota to digest chitin in our focal species. The GH18 enzyme family comprises chitinases and chitin-binding proteins and within this family, bacterial chitinases are classified into three subfamilies (A, B, and C) based on sequence homology (32–34). Symbiotic chitin-degrading bacteria have been identified, as well as their chitinase genes, in the Malayan pangolin and the giant anteater (20, 21) and in other mammals having a chitin-rich diet (36, 37). This suggests that other myrmecophagous species might also carry chitinolytic gut bacteria that could participate in prey digestion. In the specific example of dietary convergence, one question resides in understanding whether these potential microbial mechanisms of chitin exoskeleton digestion have converged among the five different placental orders of myrmecophagous mammals. The different myrmecophagous mammal lineages diverged millions of years ago and evolved in diverse environments. Chitinolytic microbes might thus have been independently recruited to ensure chitin digestion, but whether similar bacteria carrying similar functions are involved still needs further investigation. Assessing whether chitinolytic bacteria present in myrmecophagous mammals are taxonomically and functionally similar will shed light on their origin.

To investigate the chitin-degrading potential of the gut microbiota of myrmecophagous mammals, we took advantage of recent advances in metagenomics to reconstruct high-quality bacterial genomes and identify GH18 as this enzyme family comprises most bacterial chitinases. By combining Nanopore long-read and Illumina short-read shotgun

metagenomic sequencing, we generated 29 new gut metagenomes from field-collected fecal samples of nine myrmecophagous and insectivorous species representatives of the five myrmecophagous placental orders. From these, we reconstructed 314 high-quality bacterial genomes harboring a diversity of GH18 chitinases. Identification of GH18 sequences in the same bacterial genomes revealed a potential role of gut symbionts in prey digestion through putative complex metabolic pathways. Both generalist and host-specific bacteria were identified in the different myrmecophagous species, potentially reflecting the divergent evolutionary histories of their hosts, and raising questions about the evolutionary forces (historical contingency and determinism) at play in shaping the gut microbiome of convergently evolved myrmecophagous mammals.

RESULTS

Potential role of the gut microbiota in chitin digestion

Combining long-read and short-read assemblies, we were able to reconstruct 314 dereplicated high-quality genome bins (156 and 158 from each dataset, respectively; see Table S1F) and highlighted the usefulness of using the two methods (see supplementary results part 1 available via Zenodo). The following analyses were done on this set of high-quality selected bins to assess the presence of chitinolytic bacteria in the gut microbiota of myrmecophagous mammals.

All selected bins were taxonomically assigned to bacterial genomes and their phylogeny is presented in Fig. S1. Overall, 58.3% of the bins (n = 183) were not assigned at the species level, 29.3% (n = 92) at the genus level, and 6.7% (n = 21) at the family level. Taxonomically assigned bins belonged mainly to the Lachnospiraceae (n = 63), Burkholderiaceae (n = 24), Acutalibacteraceae (n = 18), Ruminococcaceae (n = 12), and Bacteroidaceae (n = 11) families (Table S1F). The 314 selected bins were integrated into a phylogeny of prokaryote reference genomes to confirm these results (Fig. 1A). All selected bins were well placed in this phylogeny, suggesting they are similar to known bacterial genomes and not distantly related, even for genome bins for which taxonomic assignment failed. Some of the reconstructed bacterial genome bins clustered together within the Firmicutes (Fig. 1B), Bacteroidetes, and Proteobacteria (Fig. S2A and B), in clades not including reference genomes, suggesting that these genome bins could be specific to myrmecophagous species (here called myrmecophagous-specific clades). Within the Firmicutes, six clades containing more than two bins reconstructed from myrmecophagous samples can be defined within the Lachnospiraceae family with two clades including, respectively, 12 and 11 genome bins (Fig. 1B). In the first clade, six bins had no taxonomic assignment below the family level and could represent bacterial genomes not yet described, two belonged to the *Acetatifactor* genus, and four to the *CAG-510* genus. In the other clade, four bins were not taxonomically assigned below the family level, three bins belonged to the *CAG-590* genus, one was assigned to the *CAG-127* genus, and two were assigned at the species level (*Frisingiococcus caecimuris* and *Acetivibrio ethanolgignens*) (Fig. 1B). These bins were all reconstructed from xenarthran gut metagenomes (Fig. 1B). Within the Acutalibacteraceae family, one clade contained 18 genome bins reconstructed mostly from xenarthran gut metagenomes and the three aardvark samples (Fig. 1B). Eleven had no taxonomic assignment below the family level, four belonged to the *Eubacterium* genus, and three to the *UBA1227*, *UBA1691*, and *UBA6857* genera (Fig. 1B). Other clades containing only genome bins from myrmecophagous species were found within the Ruminococcaceae, Oscillospiraceae, Anaerovoracaceae, and Erysipelotrichaceae families.

To understand whether these bacteria could potentially degrade chitin, these high-quality genome bins were then used to search for GH18 enzymes. This resulted in 132 bins containing at least one GH18 sequence (between 0 and 17 bins per sample, 4.93 on average). Having complete genome bins enabled us to identify several GH18 genes in the same bin (between 0 and 17 per bin, 1.26 on average; Table S1F). In total, 394 GH18 sequences were identified (Fig. 2) with 237 sequences presenting an active chitinolytic site (DXXDXDXE) (Fig. 2) and distributed among 82 bins (here called chitinolytic bins).

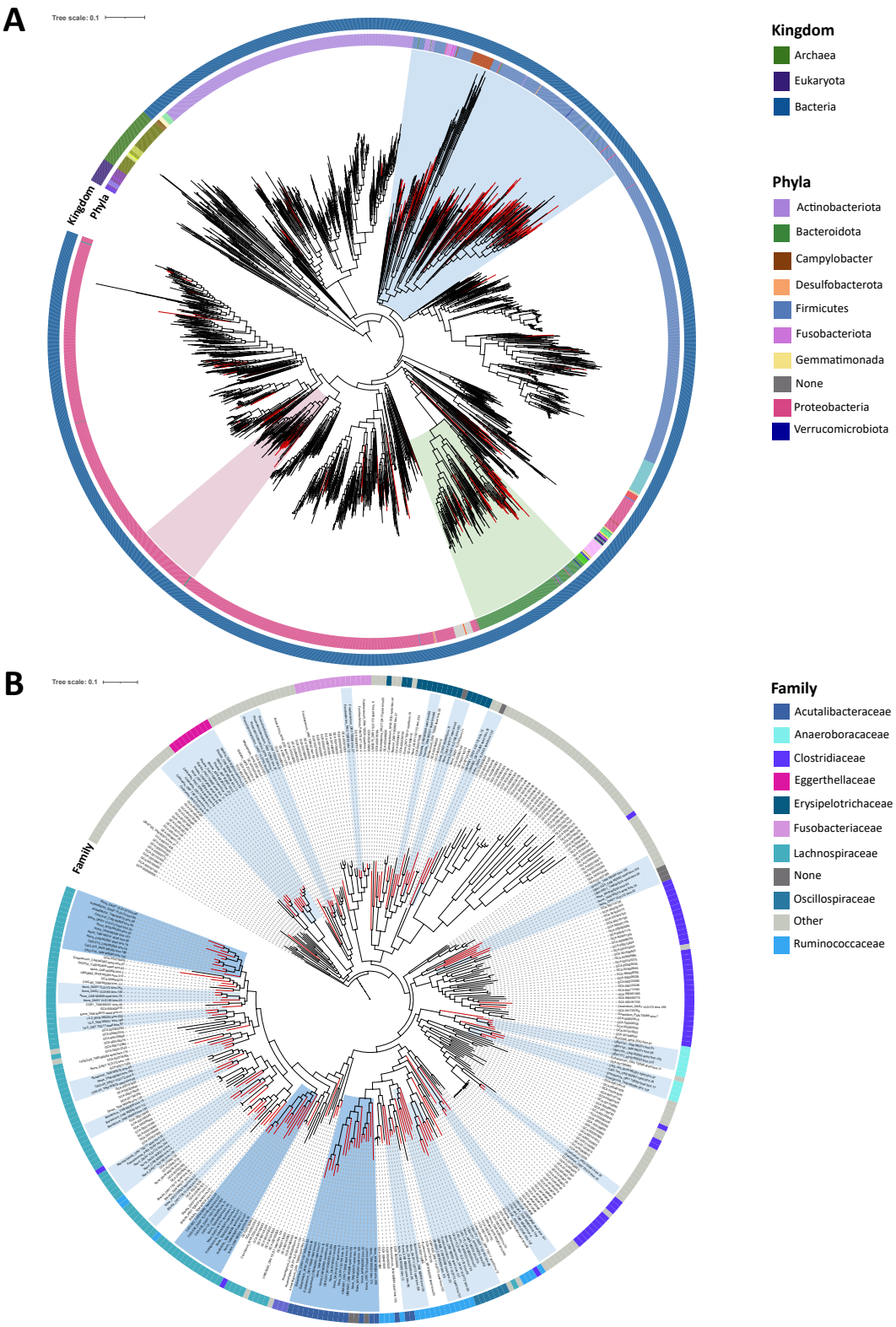


FIG 1 Phylogenetic position of the 314 high-quality selected bins reconstructed from 29 gut metagenomes of the nine focal myrmecophagous species within a reference prokaryotic phylogeny. (A) Phylogeny of the 314 selected bins (red branches) with 2496 prokaryote reference genomes. Circles, respectively, indicate (from inner to outer circles): the bacterial phyla and kingdom to which these genome bins were assigned based on the Genome Taxonomy Database release 7

(Continued on next page)

FIG 1 (Continued)

(38). Clades, where a subtree was defined, are highlighted in blue for the Firmicutes (Fig. 1B), green for the Bacteroidetes, and pink for the Proteobacteria (Fig. S2A and B, respectively). (B) Subtree within Fimircutes showing myrmecophagous-specific clades (blue highlights; dark blue corresponds to the three clades mentioned in the results, light blue to the other clades). The outer circle indicates the bacterial family to which these genome bins were assigned based on the Genome Taxonomy Database. Bins' names of the myrmecophagous-specific clades are indicated at leaves of the phylogenetic tree together with the genus to which they were assigned to.

These chitinolytic sequences are found in genome bins belonging mainly to the Lachnospiraceae (*n* = 183 sequences; e.g., *Blautia*, *Acetatifactor*, *Roseburia*, *Clostridium* genera), Acutalibacteraceae (*n* = 76; e.g., *Eubacterium* genus), and Ruminoccocaceae (*n* = 23; e.g., *Ruminococcus*, *Acetanaerobacterium* genera) bacterial families (Fig. 2). Fifty-three sequences not presenting an active site were found in a clade with sequences similar to lysin motif (LysM) domain-containing proteins (Fig. 2), which is a 40 amino acid domain involved in peptidoglycan and chitin-binding (39). Sixty-two were placed in a clade with sequences similar to src Homology-3 (SH3) domain-containing proteins (Fig. 2), which is a 50 amino acid domain found in intracellular and membrane proteins involved in the binding of ligands (40). These two clades were used to root the tree. Finally, 278 GH18 sequences formed a clade with sequences similar to known bacterial chitinases (Fig. 2). The majority had an active chitinolytic site. These sequences were identified in bins reconstructed from different host species, representative of the five myrmecophagous placental orders. They represent a diversity of chitinase genes, as they are distributed in distinct clades (Fig. 2).

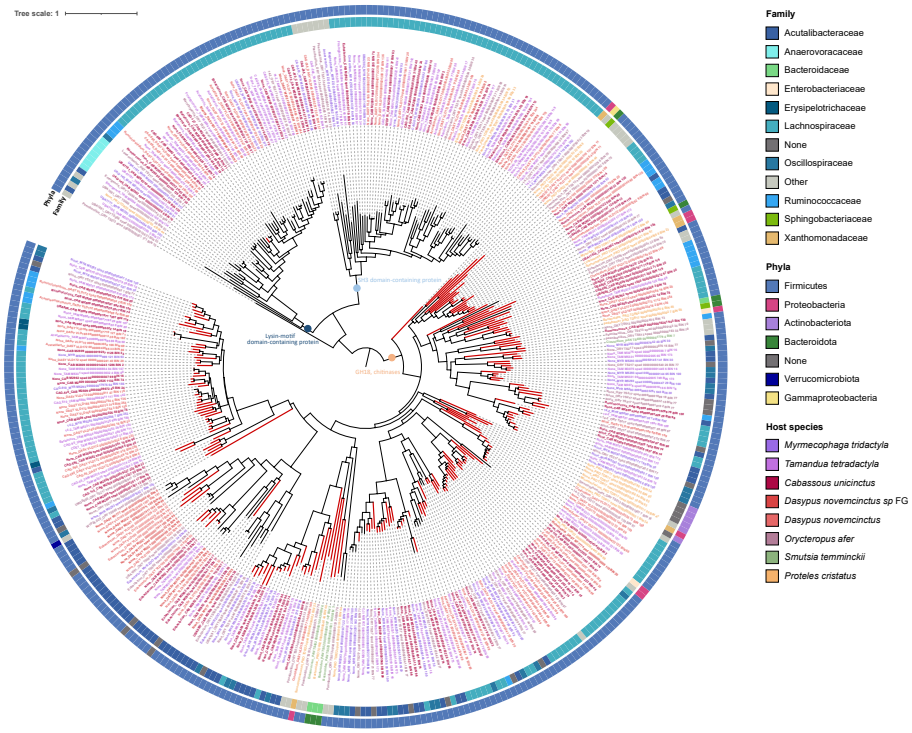


FIG 2 Phylogeny of the 394 GH18 sequences identified in 132 high-quality selected bins reconstructed from 29 gut metagenomes of the nine focal myrmecophagous species and relatives. Red branches indicate the 237 sequences having an active chitinolytic site (DXXDXDXE). Circles, respectively, indicate (from inner to outer circles): the bacterial family and phyla of the bin the sequence was retrieved from. Colored sequence names indicate the host species. Colored circles at certain nodes indicate enzymes to which sequences are similar when blasting them against the NCBI nonredundant protein database. Sequence names are indicated at leaves of the tree and begin with the genus to which the bin they were identified in was assigned to.

Distribution of chitinolytic bacteria among myrmecophagous mammals

Numbers of shared and specific selected bins in the nine host species were computed based on a detection threshold of 0.25 (the percentage of the reference covered by at least one read) of the selected bins across samples (Fig. 3; Table S2 , and detection table available via Zenodo). According to this threshold, five selected bins were not considered to be detected in any sample. Between 6 and 124 selected bins were detected in each sample (52.62 on average) and 252 selected genomes were shared and detected in samples other than the ones they were reconstructed from (Fig. 3). Selected bins were detected in 4.86 different samples on average (between 0 and 18). Selected bins were shared between individuals of the same species but with some intra-specific variability (Fig. 3). Conversely, 57 bacterial genomes were detected in only one sample (Fig. 3; Table S2) and 194 bins were detected in more than one host species (between 0 and 7 host species; 2.73 on average), between closely related species or distantly related species (Fig. 3; Table S2). On the other hand, 115 selected bins were specific to a particular host species, meaning they were found in only one of the nine species studied here (Fig. 3;

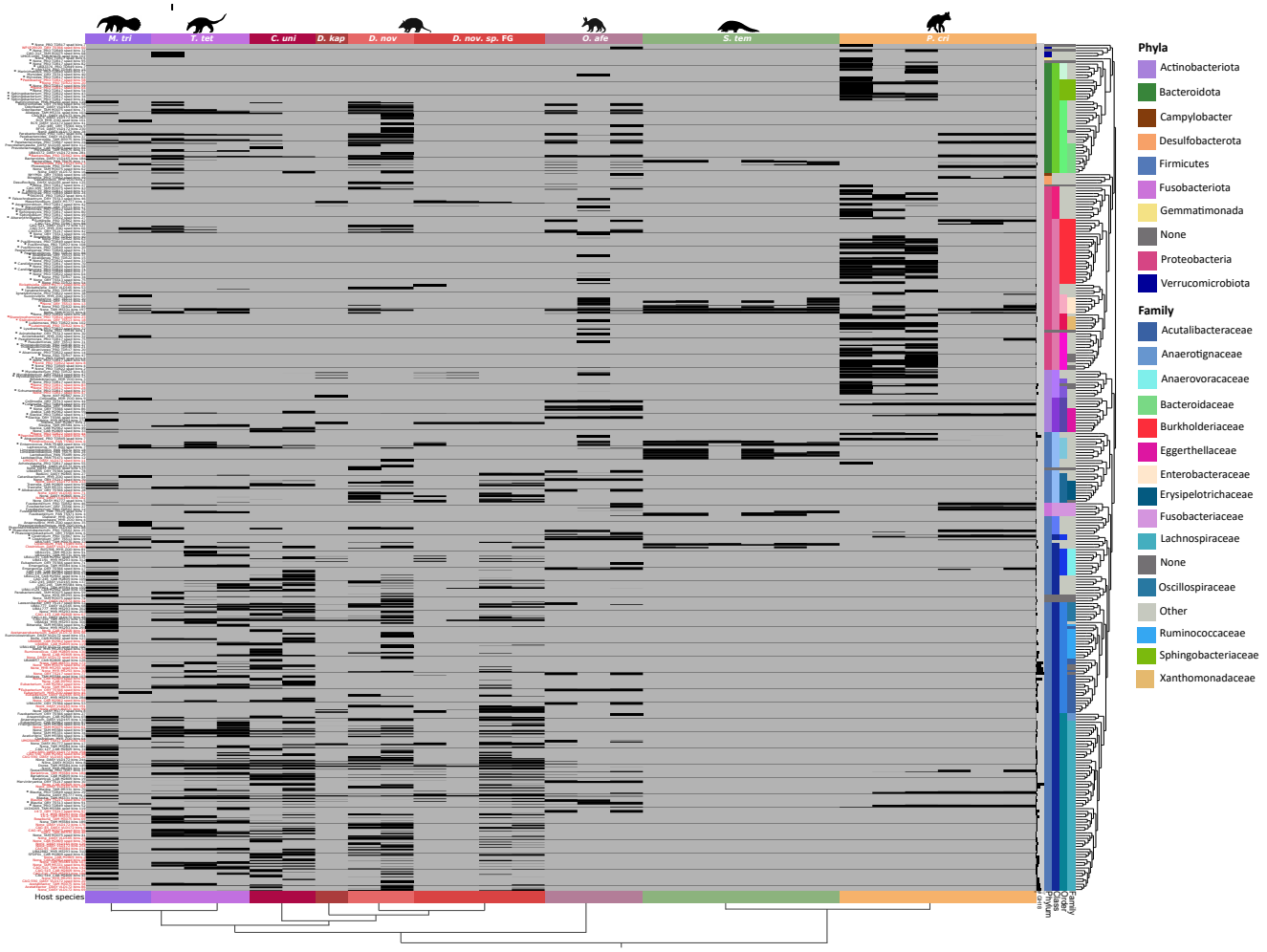


FIG 3 Detection of the 314 high-quality bacterial genomes (lines) in the 29 gut metagenomes (columns) of the nine focal species. Each square indicates the detection of a genome bin in a sample as estimated by anvi'o v7 (41). Names of bins are indicated on the left with red indicating chitinolytic bins (Table S2). The names begin with the genus to which the bin was assigned to. Asterisks (*) indicate bins detected in at least one soil sample (detection >0.25) (Fig. S4; Table S2, and detection table available via Zenodo). Phylogenetic relationships of host species distinguished by different color strips are represented at the bottom of the graph. Columns on the right indicate (from left to right) the number of GH18 sequences identified in each bin (from 0 to 17), the bin's taxonomic phylum, class, order, and family. The phylogeny of the 314 selected bins inferred with PhyloPhlAn v3.0.58 (42) is also represented on the right of the graph (see Fig. S1). Silhouettes were downloaded from phylopic.org.

Table S2). Selected bins were detected in 1.80 different host orders on average (between 0 and 5) for a total of 179 bins detected in more than one host order. Finally, 130 bins were detected in only one host order.

There were more shared selected bins between host species carrying GH18 genes than shared bins with no GH18 gene (Fig. S3). Among the 194 bins shared between host species, 110 had at least one GH18 gene. These bins belonged to 15 different bacterial families, mainly the Lachnospiraceae (n = 60; e.g., *Blautia*, *Acetatifactor*, *Roseburia* genera) and Acutalibacteraceae (n = 10; e.g., *Eubacterium* genus) (Fig. S3). The 84 shared selected bins without chitinases belonged to a more diverse range of bacterial families (n = 33), mainly Burkholderiaceae (n = 8; e.g., *CAG-521*, *Sutterella*, *Bordetella* genera) and Bacteroidaceae (n = 7; e.g., *Bacteroidetes*, *Prevotella* genera) (Fig. S3). On the contrary, among bins found in only one host species, there were fewer chitinase-carrying bins. Among the 115 host-species-specific bins, 21 had at least one GH18 gene. They were found in 12 different bacterial families (Fig. S3). The 94 specific bins with no chitinase genes belonged to 37 different bacterial families, mainly Burkholderiaceae (n = 16) (Fig. S3). A similar pattern was observed at the host order level (see supplementary results part 2 available via Zenodo).

Three high-quality selected bins carrying GH18 were shared between species belonging to the five different myrmecophagous orders (Table S1F): one bin of *Bacteroides fragilis* (mean absolute abundance across samples 2.39, see abundance table available via Zenodo) and two bins of Enterobacteriaceae (mean absolute abundances across samples 0.58 and 3.66). Two of these three bins shared across orders were a bit more abundant, on average, than the mean absolute abundance of selected bins across samples of 0.89 (the minimal abundance of a bin across samples was 2.82e-6 and the maximal absolute abundance was 198). Seven selected bins carrying GH18 were shared among four myrmecophagous orders (Table S1F). Two bins belonged to *Enteroccocus faecalis* and were reconstructed from pangolin samples, along with bins belonging to Lachnospiraceae bacteria (notably one from *Blautia* sp., one from *Faecalimonas* sp., and one from an unknown genus), one from *Bacteroides* sp. (Bacteroidaceae), and one from *Emergencia timonensis* (Anaerovoracaceae) with mean absolute abundances across samples ranging from 0.21 (*E. timonensis*) to 4.74 (*Bacteroides* sp.), the latter being found to be abundant in several samples [e.g., one *D. novemcinctus* (12.2), one *T. tetradactyla* (21.2), one *O. afer* (41.2)]. One bin of *Enteroccus faecalis* also had a mean absolute abundance (3.43) above the mean absolute abundance of selected bins and was found to be abundant in several samples as well [one *D. sp. nov* FG (33.8), three pangolin samples (10.8, 13.3 and 21.8), and one *T. tetradactyla* (15.3)].

At least one chitinolytic bin (a high-quality selected bin carrying at least one GH18 sequence with an active chitinolytic site) was detected in the gut microbiota of each host species (Fig. 4; Table S2). Consistent with our previous results (Fig. 2) these bins mainly belonged to the Firmicutes (Fig. 4). More chitinolytic bins were detected in gut metagenomes of xenarthran species compared to the aardvark, the southern aardwolf, and the ground pangolin, except for *Dasypus kappleri* for which we only had one sample (Fig. 4; Table S3). When compared to the total number of genome bins detected per species, xenarthran species had higher proportions of chitinolytic genome bins (Table S3).

DISCUSSION

Potential role of the gut microbiota in prey digestion in myrmecophagous mammals

The gut microbiota plays an important role in host digestion, which has led to several cases of gut microbiome convergence in distantly related species that share similar diets (5, 19, 44). In animals with a chitin-rich diet, symbiotic chitinolytic bacteria could be involved in prey digestion, as they have been identified in species as diverse as nine-banded armadillos (45), insectivorous bats (28), insectivorous monkeys (30), or crustacean-eating whales (46). Here, we focused on reconstructing high-quality bacterial

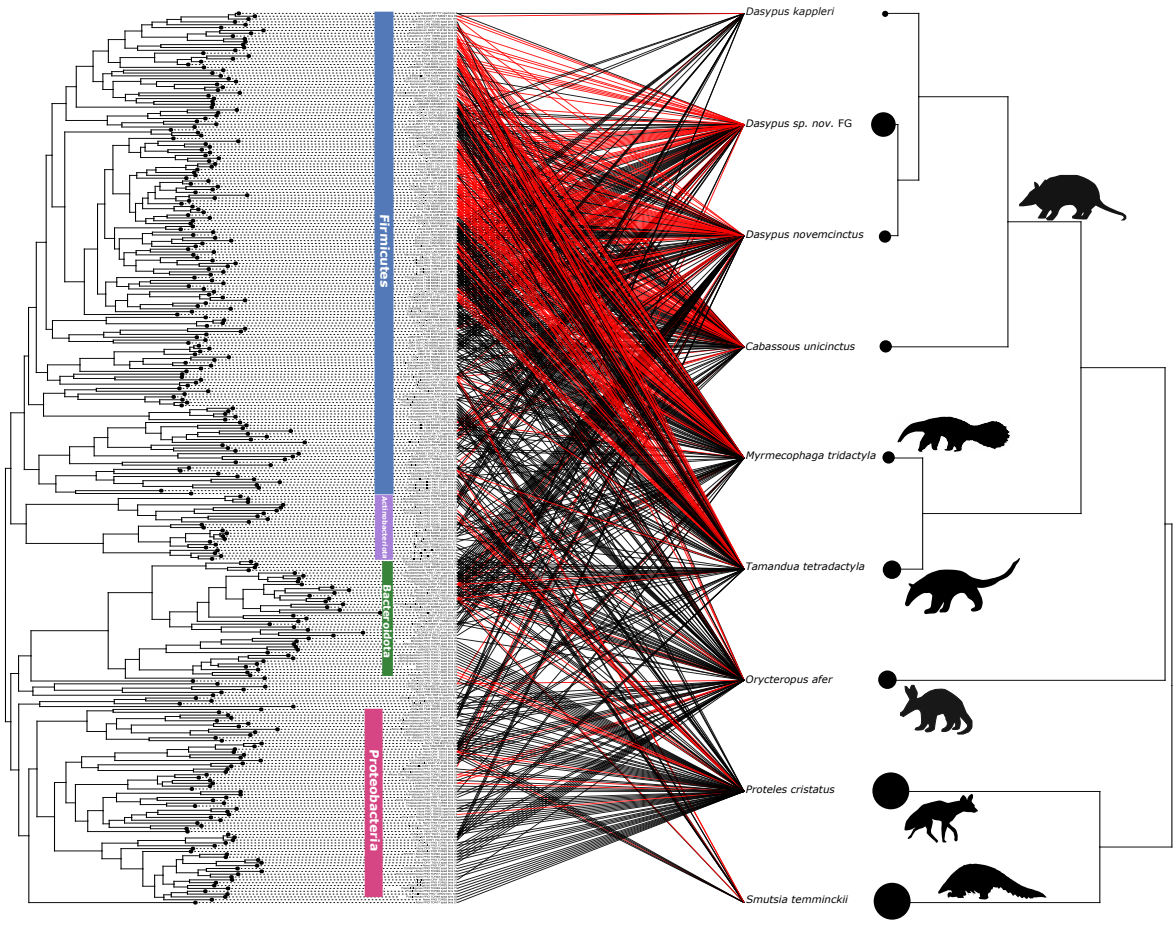


FIG 4 Distribution of chitinolytic selected bins (red links) among the nine focal myrmecophagous species and relatives. Phylogenies of the 314 high-quality selected bins (Fig. S1) and of the nine host species (downloaded from timetree.org) are represented, respectively, on the left and the right of the graph. Links illustrate, for each bin, in which host species the bin was detected (detection threshold >0.25). Red links indicate bins in which at least one GH18 sequence with an active chitinolytic site (DXXDXDXE) was found (chitinolytic bins). The size of the circles at the tips of the host phylogeny is proportional to the number of samples (n = 1 for *D. kap*; n = 2 for *D. nov*, *C. uni* and *M. tri*; n = 3 for *T. tet* and *O. afe*; n = 4 for *D. sp. nov* FG; n = 6 for *P. cri* and *S. tem*). Bins' names are indicated at the tip of the bins' phylogeny and main bacterial phyla are indicated by colored vertical bars. This graph was done with the cophylo R package within the phytools suite (43). Silhouettes were downloaded from phylopic.org.

genome bins from gut metagenomes of species representative of the five myrmecophagous orders and identified sequences belonging to the main bacterial chitinases (i.e., GH18) (32, 33, 47). This allowed us to assess the potential of chitin degradation of gut symbionts in myrmecophagous placentals and better understand their convergent adaptation to this highly specialized diet.

Among the high-quality bacterial genome bins reconstructed, several clustered together in myrmecophagous-specific clades. These clades were found mainly within the Lachnospiraceae and Actualibacteraceae families within Firmicutes but also within the Oscillospiraceae and Ruminococcaceae. Several of these genomes had no taxonomic assignment below the family level, suggesting they could represent bacterial taxa not described yet and could be linked with the adaptation to myrmecophagy. Most of these bacterial taxa (i.e., Lachnospiraceae such as *Blautia* or *Roseburia* genera or Ruminococcaceae) were previously found to be significantly more abundant in the gut microbiota of the aardwolf, giant anteater, and southern tamandua when compared to their non-myrmecophagous sister species based on 16S rRNA metabarcoding data (19),

which suggests that their presence may be associated with dietary adaptations. Besides, chitin-degrading metabolic pathways were identified in the gut metagenome of the Malayan pangolin (*Manis javanica*) as well as genes encoding chitinases and chitin-binding proteins belonging to the GH18 and GH19 enzyme families (20). This allowed the identification of chitinolytic symbionts such as *Enterococcus faecalis* (Enterococcaceae), *Clostridium paraputrificum* (Clostridiaceae), and *Bacteroides fragilis* (Bacteroidaceae) (20), whose genomes were also reconstructed here from Temminck's pangolin (*Smutsia temminckii*) gut metagenomes, confirming their probable role in prey digestion in pangolins.

Numerous GH18 genes were identified within the high-quality bacterial genomes we reconstructed here. We were able to identify 132 genome bins carrying at least one GH18 gene among which 83 had at least one GH18 sequence with an active chitinolytic site (DXXDXDXE) (48, 49), here called chitinolytic bins, and representing bacteria that might thus play a role in insect prey digestion. Among genome bins found in the three myrmecophagous-specific clades mentioned previously, almost all had at least one GH18 sequence. This further suggests that these genome bins could represent bacterial taxa playing a significant role in chitin digestion in myrmecophagous species. Studying the distribution of these bacteria in other insectivorous mammals will allow further understanding of their role in the adaptation toward these specialized diets.

In addition, some of the bacterial taxa we identified here as being potentially involved in chitin digestion in myrmecophagous species belonged to or were closely related to bacterial taxa known for their chitinolytic properties. For example, *Clostridium* species are known for their chitinolytic activity (50–52). Chitinases have also been identified and studied in *Enterococcus faecalis* (53, 54). More generally, Lachnospiraceae, for which we reconstructed the most genomes, are known to degrade complex polysaccharides (55) and our results suggest they could be involved in chitin hydrolysis. Moreover, several genomes of Ruminococcaceae bacteria reconstructed here encode a GH18. Within this family, *Ruminococcus* species have been suggested to play a role in chitin digestion in the gut microbiota of insectivorous mammals (56). Several GH18 sequences were found in the same bacterial genomes, suggesting that complex chitin-degrading pathways are present in the gut microbiota of the myrmecophagous species, consistent with previous analyses of the chitinolytic properties of certain bacteria. Indeed, several enzymes are involved in the different steps of chitin hydrolysis (23). For example, two types of chitin-binding and chitinase enzymes have been described in *E. faecalis* (54). This bacteria also carries endo-β-*N*-acetylglucosaminidases (57) which are part of the chitin hydrolysis pathway and degrade chitin into chitosan (23). This further highlights the complex chitinolytic machinery of these bacteria.

Finally, some sequences do not have the ability to hydrolyze chitin because no active site was identified, but could still bind chitin. For example, the lysin motif (LysM) domain of certain proteins can bind peptidoglycan (39). It is present in certain eukaryotes' chitinases (i.e., algae, nematodes) and participates in the recognition of symbiotic rhizobial bacteria by leguminous plants and is thus thought to bind chitin (39, 58). The SH3 domain is also known to have binding properties and is involved in protein interactions (40). Therefore, sequences similar to LysM and SH3 domain-containing proteins could still be part of the chitin hydrolysis process by influencing molecular interactions. In addition, enzymes carrying the carbohydrate-binding module 37 (CBM37), which is known to have binding properties notably to chitin, could also be involved in the chitin-degrading process. The presence of other enzymes potentially involved in insect digestion could be investigated, such as trehalases, enzymes that break down trehalose, a sugar found in insect blood. The gut microbiota of the Malayan pangolin and giant anteater were found to be significantly enriched in these enzymes when compared to non-myrmecophagous species (21). The gut microbiota also plays a role in the detoxification of ingested compounds (26, 44), such as insect toxins. For example, metabolic pathway analyses have revealed the potential of microbial symbionts to detoxify formic acid in some myrmecophagous species (21).

Shared and species-specific gut bacteria among myrmecophagous mammals

Morphological and genomic comparative studies have shown that the convergent adaptation to myrmecophagy in mammals involved different mechanisms between the different species and that phylogenetic constraints played a central role (15, 18, 59–61). As diet is one of the main factors shaping the evolution of the gut microbiota (5), this raises the question of whether the same bacterial symbionts with similar functions were convergently recruited between the different ant-eating species, or whether it was different bacteria with similar functions (functional convergence).

Analysis of the distribution of selected genome bins across samples revealed both shared and host-specific bacteria among myrmecophagous species. Selected genomes were mainly shared between closely related species and mainly among *Xenarthra* but less so between distantly related orders such as pangolins (Pholidota) and anteaters (Pilosa), highlighting the influence of host evolutionary history. In mammals, phylosymbiosis is particularly strong (9, 62) in part because of mammalian-specific traits such as viviparity or parental care (63) but also limited microbial dispersal abilities (64) facilitating the vertical transmission of microbes, which may explain these results. Some bacterial genomes were shared between distantly related host species and those shared across more than one order mainly belong to the family Lachnospiraceae. One selected genome of *Bacteroides fragilis* and two of Enterobacteriaceae were shared between species belonging to the five different myrmecophagous orders and carry GH18. Similarly, seven selected genome bins were shared among four host orders and belong to *Enteroccocus faecalis*, the Lachnospiraceae (i.e., *Blautia* sp., *Faecalimonas* sp.), *Bacteroides* sp. (Bacteroidaceae), and *Emergencia timonensis* (Anaerovoracaceae). These shared high-quality selected genome bins carrying GH18 may thus participate in the adaptation to myrmecophagy and highlight the influence of the host diet, consistent with previous studies showing that it is an important factor in shaping the mammalian gut microbiota (1, 4, 5, 8). Each host species appears to carry chitinolytic bacteria in its gut microbiota, confirming their potential role in adapting to an insect-based diet. However, differences in the distribution of these chitinolytic selected bins could highlight divergent microbiota adaptations, especially in xenarthran species, which carry more chitinolytic bacteria than aardvark, ground pangolin, and southern aardwolf. Some bins were found to be abundant in some samples raising questions on how the gut microbiota is involved in prey digestion. Having few highly abundant chitinolytic bacterial species producing large amounts of chitinases might be sufficient for a specific host to digest its prey, whereas having diverse chitinolytic bacterial species might have been selected for in other host species.

The distribution of high-quality selected genome bins across samples and species revealed how shared and host-specific bacteria may reflect the divergent evolutionary histories of host species and the effect of adapting to a similar diet despite phylogenetic constraints. This relates to the evolution of the mammalian gut microbiota. From a mammalian ancestor that probably had an insectivore-like gut microbiota (62), its composition might have changed as placentals diversified and occupied new niches, constraining its composition in certain species. Chitinolytic bacteria shared by all myrmecophagous orders may represent ancient bacterial lineages inherited from a common ancestor, while bacteria found in only certain host species may reflect more recent adaptations with bacteria acquired in specific lineages.

The environment and biogeography also influence host-associated microbial communities, which could explain the distribution pattern of selected bins across samples, with large differences observed between South Africa and South America. Anteaters and armadillos (*Xernarthra*) diverged anciently from aardvarks, pangolins, and aardwolves (~80 Ma) (14), and these lineages have evolved in very distinct environments and biogeographic contexts for most of their evolutionary history. Soil samples from the fecal sampling sites have been collected for the South African samples. Some selected bins reconstructed from gut metagenomes of the ground pangolin, southern aardwolf, and aardvark belonged to bacterial taxa that were indeed not expected to be found in

the gut but rather in the environment (e.g., Sphingobacteriaceae, Xanthomonadaceae, and Burkholdericaeae). The distribution of selected genome bins reconstructed from gut metagenomes of these species was examined in soil samples and showed that 91 selected bins were also detected in at least one soil sample (detection >0.25; Fig. S4; Table S2). For example, among the Proteobacteria selected genome bins found in myrmecophagous-specific clades (Fig. S2B), some belonged to Burkholderiaceae bacteria and were detected in at least one soil sample. This may not necessarily reflect environmental contamination, as the center of the feces was specifically sampled to minimize contamination. Rather, it may reflect the fact that these species ingest soil and thus environmental microbes while foraging (11). This could be beneficial to the host (65), as it may compensate for the lack of mastication and/or help to deal with toxins that might be present in prey (11). Environmental acquisition of microbes has also been suggested in lemurs (66) and wild echidnas (67), species that ingest soil during foraging and in which soil bacteria have been identified in their gut microbiota. These environmental bacteria could therefore represent transient bacteria in the gut, reflecting what is ingested by the host. They could also represent resident bacteria that have been recruited from the environment into the gut microbiota of myrmecophagous species. Indeed, some beneficial microorganisms could be acquired from soil microbiomes (68), which could allow the host to better adapt to its environment through horizontal gene transfer of beneficial genes (69). Moreover, it has also been suggested that environmental bacteria could be a source of chitinase genes to digest prey (22). Thus, their recruitment into the gut microbiota may have been selected to participate in the host digestion and further contribute to the adaptation to its environment. This would highlight the potential influence of the environment on the gut microbiota of myrmecophagous mammals.

Role of the holobiont in prey digestion in myrmecophagous mammals

The influence of the microbiota on host evolution and their coevolution is increasingly recognized as studies of the diverse microbiota of captive and wild animals proliferate. Many scientists now recognize the term holobiont to describe the host and all its associated symbionts, which would be the unit of natural selection as defined by the hologenome theory of evolution (70). Integrative studies (21, 71) and initiatives such as the "Earth Hologenome Initiative" (http://www.earthhologenome.org/) to study hologenomic adaptations are now becoming more common. In the specific case of myrmecophagous mammals, a hologenomic approach would help to better understand the adaptive mechanisms involved. Indeed, if chitinolytic bacteria can digest chitin, prey digestion can also be ensured by host-produced chitinases. In mammals, from a placental ancestor that probably carried five functional chitinase paralogs (*CHIA*), some of these paralogs were subsequently lost during placental diversification in non-insectivorous species, leading to a positive correlation between the number of functional chitinase paralogs and the proportion of invertebrates in the diet (15); a correlation also observed in primates (72). Among myrmecophagous species, this chitinase gene repertoire has evolved differently to ensure chitin digestion, reflecting phylogenetic constraints (15, 18). Thus, both endogenous and microbial chitinases may be involved in prey digestion, raising the question of the relative contribution of the host and its symbionts in providing the same function. For example, myrmecophagous mammals with only one functional *CHIA* paralog (*CHIA5*), such as aardwolves and pangolins, may compensate for this by overexpressing their only functional paralog or by relying more on their chitinolytic symbionts to digest their prey. In pangolins, *CHIA5* has been found to be overexpressed in all digestive organs (18). In the aardwolf, which recently diverged from the other Hyaenidae species (~10 Ma) (73), and does not present strong morphological adaptations to myrmecophagy, no host chitinases have been found to be expressed in its salivary glands (18) and expression of these enzymes in other digestive organs needs further investigation. Therefore, this species could rely more on its gut microbiota to ensure prey digestion. Based on our results, fewer chitinolytic bacteria were found in the gut microbiota of the southern aardwolf compared to other

species but they could still ensure chitin digestion, for instance, by overexpressing their chitinases. The combination of the host genomic and transcriptomic data with metagenomic and metatranscriptomic data of its microbiota would allow answering this type of question by comparing host gene repertoires in light of the chitinolytic abilities of the gut microbiota. This will shed light on the adaptation to chitin digestion in placentals and, more generally, on the role of the holobiont in the adaptation to a specific function.

MATERIALS AND METHODS

Sampling

Thirty-three fecal samples were collected from nine species representative of the five myrmecophagous placental orders (Table 1). For armadillos and anteaters provided by the JAGUARS collection (Cayenne, French Guiana), fecal samples from roadkill and deceased zoo animals were obtained after unfreezing the specimens and dissecting the lower part of the digestive tract in the lab facilities provided by Institut Pasteur de la Guyane (Cayenne, French Guiana). Roadkill armadillos collected in the United States were also dissected to sample feces. For the aardvark, ground pangolin, and southern aardwolf, fresh fecal samples were collected directly in the field during fieldwork sessions conducted in Tswalu Kalahari and Tussen-die-Riviere reserves (South Africa). The inner part of the feces was sampled with a sterile scalpel blade to avoid soil contamination. In the South African reserves, eight soil samples were also collected near feces sampling sites to serve as a control for potential environmental contamination (Table S4). All fecal and soil samples were stored at −20°C in 96% ethanol before DNA extraction.

DNA extraction

Whole DNA was extracted from fecal samples following the optimized protocol of (79) using the enomic DNA from soil kit (NucleoSpin, Macherey-Nagel). Two successive extractions were done and a purification step was added to retrieve high-molecular-weight DNA suitable for long-read sequencing (79). The same kit was used to extract DNA from soil samples. Before the extraction, samples were incubated with 700 µL of the lysis buffer (SL1) (79) and 30 µL of proteinase K at 56°C for 30 min.

Library preparation and DNA sequencing

Long-read sequencing

Long-read libraries were constructed using the SKQ-LSK109 and SKQ-LSK110 library preparation kits (ONT, Oxford Nanopore Technologies, https://nanoporetech.com/). Shotgun metagenomic sequencing was done using the MinION, MK1C, and GridION devices using one R9 flowcell per sample. Between 150 and 988 ng of DNA were loaded per flowcell, which were run for 48 to 72 h (Table S1A). For samples sequenced on the GridION, super-accurate basecalling was performed with Guppy v5+ (Qscore = 10) whereas on the MinION and MK1C fast basecalling was used (Qscore = 7). Sequencing output statistics were checked using PycoQC v2.5.2 (80) with a minimum quality score set to seven or 10 depending on the sequencing device used (Table S1A). Rebasecalling of samples sequenced on the MinION and MK1C was done on a GPU machine with Guppy v5.0.16 (super accurate mode, Qscore = 10, config_file = dna_r9.4.1_450bps_sup.cfg) (ONT).

Short-read sequencing

Short-read shotgun metagenomic Illumina sequencing was done for all samples to generate data for long-read assemblies polishing. Either the first or second extraction was used for sequencing as they have previously been shown to be both suitable for short-read sequencing (79). Library preparation and Illumina sequencing on a

TABLE 1 Detailed sample information for the 33 fecal samples collected[a,b]

Sample name	Species name	Common name	Order	Diet	Sex	Age	Wild/ Captive	Location		Accession number	
								Country of origin		Illumina	ONT
CAB M2809	*Cabassous unicinctus*	Southern naked-tailed armadillo	Cingulata	Myrmecophagous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925023	SRR23925022
CAB M2962	*Cabassous unicinctus*	Southern naked-tailed armadillo	Cingulata	Myrmecophagous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925011	SRR23925000
CAB M3141	*Cabassous unicinctus*	Southern naked-tailed armadillo	Cingulata	Myrmecophagous	Female	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924989	SRR23924978
DASY M1746	*Dasypus novemcinctus* sp FG	Guianan long-nosed armadillo	Cingulata	Omnivorous	Male	Adult	Wild	French Guiana	coll. JAGUARS	NA	SRR23924967
DASY M1777	*Dasypus novemcinctus* sp FG	Guianan long-nosed armadillo	Cingulata	Omnivorous	Male	Juvenile	Wild	French Guiana	coll. JAGUARS	SRR23924962	SRR23924961
DASY M2255	*Dasypus novemcinctus* sp FG	Guianan long-nosed armadillo	Cingulata	Omnivorous	NA	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924960	SRR23925021
DASY M2865	*Dasypus novemcinctus* sp FG	Guianan long-nosed armadillo	Cingulata	Omnivorous	Female	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925020	SRR23925019
DASY M3021	*Dasypus novemcinctus* sp FG	Guianan long-nosed armadillo	Cingulata	Omnivorous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925018	SRR23925017
DASY VLD165	*Dasypus novemcinctus*	Nine-banded armadillo	Cingulata	Omnivorous	Female	Adult	Wild	USA	Valdosta (GA)	SRR23925016	SRR23925015
DASY VLD168	*Dasypus novemcinctus*	Nine-banded armadillo	Cingulata	Omnivorous	Male	Adult	Wild	USA	Valdosta (GA)	NA	SRR23925014
DASY VLD172	*Dasypus novemcinctus*	Nine-banded armadillo	Cingulata	Omnivorous	Female	Adult	Wild	USA	Valdosta (GA)	SRR23925013	SRR23925012
KAP M2867	*Dasypus kappleri*	Greater long-nosed armadillo	Cingulata	Insectivorous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925010	SRR23925009
MYR M5293	*Myrmecophaga tridactyla*	Giant anteater	Pilosa	Myrmecophagous	Female	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925008	SRR23925007
MYR M5295	*Myrmecophaga tridactyla*	Giant anteater	Pilosa	Myrmecophagous	NA	Juvenile	Wild	French Guiana	coll. JAGUARS	SRR23925006	SRR23925005
MYR ZOO	*Myrmecophaga tridactyla*	Giant anteater	Pilosa	Myrmecophagous	Female	Adulte	Captive	France	Montpellier zoo	SRR23925004	SRR23925003

(*Continued on next page*)

TABLE 1 Detailed sample information for the 33 fecal samples collected, (*Continued*)

Sample name	Species name	Common name	Order	Diet	Sex	Age	Wild/ Captive	Location		Accession number	
								Country of origin			
ORY TS217	*Orycteropus afer*	Aardvark	Tubulidentata	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23925002	SRR23925001
ORY TS513	*Orycteropus afer*	Aardvark	Tubulidentata	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924999	SRR23924998
ORY TS566	*Orycteropus afer*	Aardvark	Tubulidentata	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924997	SRR23924996
PAN TS471	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924995	SRR23924994
PAN TS475	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924993	SRR23924992
PAN TS482	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924991	SRR23924990
PAN TS488	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924988	SRR23924987
PAN TS489	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924986	SRR23924985
PAN TS525	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924984	SRR23924983
PRO TDR7	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924982	SRR23924981
PRO TDR17	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924980	SRR23924979
PRO TDR22	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924977	SRR23924976
PRO TDR49	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924975	SRR23924974
PRO TDR62	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924973	SRR23924972
PRO TDR67	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924971	SRR23924970
TAM M3075	*Tamandua tetradactyla*	Southern tamandua	Pilosa	Myrmecophagous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924969	SRR23924968
TAM M5331	*Tamandua tetradactyla*	Southern tamandua	Pilosa	Myrmecophagous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924966	SRR23924965

(*Continued on next page*)

TABLE 1 Detailed sample information for the 33 fecal samples collected, (*Continued*)

Sample name	Species name	Common name	Order	Diet	Sex	Age	Wild/ Captive	Location		Accession number	
								Country of origin			
TAM M5584	*Tamandua tetradactyla*	Southern tamandua	Pilosa	Myrmecophagous	Female	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924964	SRR23924963

[a]Diet was determined based on field observations (i.e., dissections) and the literature (11, 74–78).
[b]Long- and short-reads are available on SRA under BioProject PRJNA942254 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA942254/).

NovaSeq instrument were outsourced to Novogene Europe (Cambridge, UK) to generate metagenomes using 50 million of 150 bp paired-end reads each (15 Gb of raw data per sample) (Table S1A). Two samples of two *Dasypus* species (DASY M1746 and DASY VLD168) could not be sequenced using short reads due to low quantity of starting material and were not included in the following analyses; resulting in 31 samples being analyzed. Short-read metagenomic sequencing of soil samples was performed following the same protocol.

Data filtering

Long-read data

Sequencing adapters were removed with Porechop v0.2.4 (81) used with default parameters. Reads shorter than 200 bp were removed using Filtlong v0.2.1 (82). No quality filtering was performed at this stage as it was previously done during basecalling (Qscore >10).

Host mitogenomes were downloaded for the nine myrmecophagous species of our dataset from the National Center for Biotechnology Information (NCBI) Genbank database (Table S1A). Long-read metagenomes were mapped to the host mitogenome with Minimap2 v2.17 (83, 84) with the ONT preset (*-ax map-ont*) for Oxford Nanopore reads. This allowed us to confirm the host species. Host mitogenomes were then assembled from the mapping reads with Flye v2.8.3 (85) with default parameters for ONT raw reads (*--nano-raw*). Mapping of host mitogenomic reads on host mitogenomes was visually checked with Geneious Prime 2022.0.2 (86), and mitochondrial reads were removed from the metagenomes for downstream analyses.

The same was done to remove host nuclear reads by mapping long-read metagenomes against the host genome using Minimap2 v2.17 (83, 84) using the Nanopore read option (*-ax map-ont*) (Table S1A). The *Dasypus novemcinctus* RefSeq genome assembly (Dasnov 3.0; GCF_000208655.1) was downloaded from Genbank. The *Orycteropus afer* genome was downloaded from the DNA Zoo database (87) (HiC assembly based on the draft assembly of Di Palma et al., unpublished). *Myrmecophaga tridactyla*, *Tamandua tetradactyla*, *Smutsia gigantea,* and *Proteles cristatus* genomes have been previously reconstructed using long (ONT) and short (Illumina) reads assembled with MaSuRCA v3.2.9 (88) [see (89) for a detailed description of the hybrid assembly process]. For *Cabassous unicinctus*, we used the Discovar draft genome assembly generated by (90). For species lacking an available reference genome, the genome of the closest relative in our dataset was used: *Smutsia gigantea* for *S. temminckii* and *Dasypus novemcinctus* for *D. kappleri* and *Dasypus* sp. nov. FG. Host reads were removed from the metagenomes for downstream analyses. One southern naked-tailed armadillo sample (CAB M3141) and one giant anteater sample (MYR M5295) presented a high proportion of host reads (>95%) and were excluded from the following analyses. The final dataset thus included 29 samples.

Finally, contaminant human reads were removed following the same approach and using the telomere-to-telomere human genome assembly (GCA009914755.3) as reference (Table S1A).

Short-read data

Illumina sequencing adapters were removed using FASTP v0.20.0 with default quality filtering parameters (91). Host and human reads were removed using the same approach as for long-read metagenomes but the mapping was done with bowtie2 v2.3.5 (92) with default parameters (Table S1A).

Samtools v1.7 (93) was used to manipulate long- and short-read mapping files.

Metagenome assembly

Long-read metagenomes were assembled for each sample using metaFlye v2.9 (94) with default parameters for ONT reads basecalled with Guppy v5+ (*--nano-hq*, recommended

mode when <5% of sequencing errors is expected) and the strain mode (*--keep-haplotypes*), which prevents closely related strains represented by different paths in the assembly graph to be collapsed. Long-read metagenome assemblies were then polished with short-reads using Pilon v1.24 (95) with default parameters. Short-read metagenomes were assembled using metaSPAdes v3.11.0 (96) and MEGAHIT v1.1.2 (97) with default parameters.

Assembly statistics were computed with anvi'o v7 (41) (Tables S1B and C). Anvi'o first generates a contig database in which *k*-mer frequencies for each contig are computed ($k = 4$). Prodigal v2.6.3 (98) is then used to predict archaeal and bacterial open reading frames and estimate the number of genes. Finally, microbial single-copy core genes (SCGs) are searched using HMMER (99) with the hmmscan v3.2.1 program (default in anvi'o).

Reconstruction of bacterial genomes

Genome binning and bins selection

Prior to binning, reads were mapped to metagenome assemblies to obtain the coverage information needed for binning. This was done using Bowtie2 v2.2.9 (92) for short reads and Minimap2 v2.17 (83, 84) for long reads. Genome binning was conducted on single long-read polished assemblies and short-read assemblies separately (trimmed to keep contigs longer than 1000 bp) using metaBAT2 v2.15 (100) with default parameters (minimum size of contigs set to 2,500 bp and bin minimal length set to 200 kb).

Completeness and redundancy of bins were estimated using anvi'o v7 (Tables S1D and E), which relies on the detection of SCGs. Bins with more than 90% completeness and less than 5% redundancy were selected for downstream analyses. Based on these criteria, 239 genome bins were selected from the long-read assemblies and 254 from the short-read assemblies. Genome bins statistics were computed using anvi'o v7 as done for the metagenome assemblies (Tables S1D and E).

Genome bins dereplication

To remove potentially redundant bins, a dereplication step was performed on the set of selected bins using dRep v3.3.0 (101). dRep first filters genomes based on their length (>50 000 by default) and their completeness and redundancy as computed with CheckM v1.1.11 (102) (>75% completeness, <25% redundancy by default). dRep then clusters genomes based on an average nucleotide identity (ANI) threshold of 90% using the Mash algorithm v2.3 (103). Genomes having at least 90% ANI are then clustered using fastANI v1.33 (104), here with a secondary ANI threshold of 98% (recommended to avoid mis-mapping of metagenomic reads against bins, e.g., to study their distribution across metagenomes). By default, at least 10% of the genome is compared (minimum alignment fraction). Two bins did not pass dRep quality filtering in each dataset and were not included in the analysis. In the long-read dataset, 38 genomes were removed by dRep with an ANI of 98%, resulting in 201 unique genome bins. In the short-read dataset, 48 genomes were removed, resulting in 206 unique genome bins. Selected genome bin statistics were compared between bins reconstructed from long-read and short-read assemblies using anvi'o v7 (see supplementary results part 1 available via Zenodo).

To determine whether similar genome bins were reconstructed using long-read polished assemblies ($n = 201$) or short-read assemblies ($n = 206$), dRep was also run on the set of all selected bins ($n = 407$) with 98% ANI to remove redundant genome bins (93 genomes removed, resulting in 314 nonredundant high-quality bins). To capture the majority of the bacterial taxa present in our samples, we decided to combine high-quality selected bins reconstructed from both long- and short-read assemblies dereplicated with an ANI of 98% in our final dataset for downstream analyses ($n = 314$; Table S1F) (see supplementary results part 1 and dRep output results available via Zenodo).

Taxonomic assignment of selected bins

The taxonomy of selected genome bins reconstructed from long- and short-read assemblies was assessed using anvi'o v7, which uses 22 SCGs and the taxonomy of the genomes defined by the Genome Taxonomy Database release 7 (GTDB) (38) from which these genes have been extracted (Table S1F).

The 314 selected bins were placed in a phylogeny of 2566 reference prokaryotic genomes downloaded from Genbank using PhyloPhlAn v3.0.58 (42). PhyloPhlAn first searches universal prokaryotic marker genes (option *-d* phylophlan, a database comprising 400 markers) (105), here using Diamond v2.0.6 (106). Marker genes were then aligned using MAFFT v7.475 (107), and cleaned using trimAl v1.4 (108). Next, marker gene alignments were concatenated. Finally, the phylogeny was inferred using IQ-TREE v2.0.3 (109) under the LG model. PhyloPhlAn was run with the *–diversity* high and *–fast* options, which together set a range of parameters for the reconstruction of high-ranked taxonomic-level phylogenies (42). By default, at least 100 markers of the PhyloPhlAn database must be present in a genome for it to be included in the analysis (*--min_num_markers* 100) and each marker should be found in at least four genomes to be included in the analysis (*--min_num_entries* 4). Seventy of the reference genomes had less than 100 markers and were excluded from the analysis resulting in a total of 2810 genomes in the final reconstructed tree. The phylogeny was rooted with Archaea so the bacteria were monophyletic. Using the same PhyloPhlAn parameters, a phylogeny of the set of the 314 selected bacterial genome bins was also inferred (Fig. S1). This phylogeny was rooted according to (110), placing the root of the bacterial tree separating Terrabacteria (e.g., Firmicutes and Actinobacteriota) and Gracilicutes (e.g., Proteobacteria, Desulfobacterota, Campylobacterota, and Fibrobacteres-Chlorobi-Bacteroidetes group).

Distribution of selected bins in gut metagenomes

Short-read metagenomes were mapped against the 314 selected bins using bowtie2 v2.3.5 with local sensitive alignment parameters. The distribution of bins across samples was explored using anvi'o v7. Anvi'o enables us to compute the percentage of mapped reads on each bin for each metagenome and computes statistics such as the coverage, percentage of recruited reads, abundance, and detection (horizontal coverage) of bins across metagenomes (samples). The distribution across samples was studied using the anvi'o interface by visualizing the detection of selected genome bins across samples. A detection threshold of 0.25% of the reference covered by at least one read was chosen to consider a bin detected in a sample. The number of bins shared among the different samples and host species and bins specific to a sample or host species were calculated. Applying the same approach, the distribution of selected bins reconstructed from the aardvark, ground pangolin, and southern aardwolf samples (n = 140) was explored in the eight soil samples collected in South Africa near feces sampling sites (Fig. S4).

Identification of bacterial chitinase genes

Chitinase genes were searched in selected genome bins. More specifically, sequences of enzymes belonging to the glycoside hydrolase 18 (GH18) family (comprising chitinases and chitin-binding proteins) as determined by the classification of carbohydrate active enzymes (CAZymes) (31) were scanned using dbCAN2 (111) with default parameters. dbCAN2 is designed to annotate CAZymes sequences in genomes using an HMMER search (99) against the dbCAN HMM database containing HMM models for each CAZyme family. Proteins were predicted using Prodigal (98).

The amino acid sequences of 420 identified GH18 genes were imported in Geneious Prime 2022.0.2. Sequences were aligned using MAFFT v7.450 with default parameters. The alignment was cleaned by removing sites not present in at least 50% of the sequences, resulting in an alignment of 390 amino acids. The chitinase gene tree was inferred with RAxML v8.2.11 (112) under the LG + G model and the rapid hill-climbing algorithm for topology rearrangements. After removing 26 sequences that did not align

correctly and had very long branches, the final alignment included 394 sequences. The phylogeny was rooted with the clades containing sequences without chitinolytic sites as they were divergent from the other sequences. The conserved chitinolytic site (DXXDXDXE) (48, 49), typical of chitin-degrading enzymes, was searched in the aligned sequences. BLAST searches against the NCBI nonredundant protein database were conducted to assess whether identified GH18 genes were similar to known microbial chitinases.

ACKNOWLEDGMENTS

We are indebted to Lionel Hautier, Rémi Allio, Fabien Condamine, Sérgio Ferreira-Cardoso, Pierre-Henri Fabre, Quentin Martinez, Mathilde Barthe, Aude Caizergues, Nathalie Delsuc, Jessica Briner, Tshediso Putsane, Baptiste Chenet, Edith Guilloton, Solène Lefort, Margo Traimond, Michel Blanc, Bertrand Goguillon, and Antoine Baglan for their help with sampling. Fieldwork sessions at Tswalu Kalahari Reserve (South Africa) were conducted under the auspices of the Tswalu Foundation thanks to Duncan MacFadyen (Oppenheimer Generations Research and Conservation), Dylan Smith, and Gus van Dyk. We thank two anonymous reviewers for their helpful comments. Computational analyses benefited from the Montpellier Bioinformatics Biodiversity (MBB) platform.

This work has been supported by grants from the European Research Council (ConvergeAnt project: ERC-2015-CoG-683257) and Investissements d'Avenir of the Agence Nationale de la Recherche (CEBA: ANR - 10 - LABX - 25 - 01; CEMEB: ANR - 10 - LABX - 0004). The JAGUARS collection, managed by Kwata NGO, is supported by the Collectivité Territoriale de Guyane and the Direction Générale des Territoires et de la Mer de Guyane.

This is contribution ISEM 2023-121-SUD of the Institut des Sciences de l'Evolution de Montpellier.

AUTHOR AFFILIATIONS

[1]Institut des Sciences de l'Evolution de Montpellier (ISEM), Univ Montpellier, CNRS, IRD, Montpellier, France
[2]CIC AG/Inserm 1424, Centre Hospitalier de Cayenne Andrée Rosemon, Cayenne, French Guiana, France
[3]Tropical Biome and immunopathology, Université de Guyane, Labex CEBA, DFR Santé, Cayenne, French Guiana, France
[4]Brain Function Research Group, School of Physiology, University of the Witwatersrand, Johannesburg, South Africa
[5]Centre for African Ecology, School of Animals, Plant, and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa
[6]Department of Biology, Valdosta State University, Valdosta, Georgia, USA
[7]National Museum and Centre for Environmental Management, University of the Free State, Bloemfontein, South Africa
[8]Institut Pasteur de la Guyane, Cayenne, French Guiana, France
[9]Kwata NGO, Cayenne, French Guiana, France
[10]Evolutionary Biology of the Microbial Cell, Institut Pasteur, Université Paris Cité, Paris, France

AUTHOR ORCIDs

Sophie Teullet http://orcid.org/0000-0003-2693-1797
Marie-Ka Tilak http://orcid.org/0000-0001-8995-3462
Roxane Schaub http://orcid.org/0000-0001-5742-3280
Nora M. Weyer http://orcid.org/0000-0002-8753-9222
Wendy Panaino http://orcid.org/0000-0002-8494-744X
Andrea Fuller http://orcid.org/0000-0001-6370-8151

W. J. Loughry http://orcid.org/0000-0001-8214-9893
Nico L. Avenant http://orcid.org/0000-0002-5390-9010
Benoit de Thoisy http://orcid.org/0000-0002-8420-5112
Guillaume Borrel http://orcid.org/0000-0003-4893-8180
Frédéric Delsuc http://orcid.org/0000-0002-6501-6287

FUNDING

Funder	Grant(s)	Author(s)
EC | European Research Council (ERC)	683257	Sophie Teullet Amandine Magdeleine Frédéric Delsuc
Agence Nationale de la Recherche (ANR)	ANR-10-LABX-0025	Sophie Teullet Marie-Ka Tilak Amandine Magdeleine Roxane Schaub Benoit de Thoisy Frédéric Delsuc
Agence Nationale de la Recherche (ANR)	ANR-10-LABX-0004	Sophie Teullet Marie-Ka Tilak Amandine Magdeleine Frédéric Delsuc

AUTHOR CONTRIBUTIONS

Sophie Teullet, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Conceptualization | Marie-Ka Tilak, Investigation, Resources, Supervision, Writing – review and editing | Amandine Magdeleine, Investigation, Resources | Roxane Schaub, Resources, Writing – review and editing | Nora M. Weyer, Resources, Writing – review and editing | Wendy Panaino, Resources, Writing – review and editing | Andrea Fuller, Resources, Supervision, Writing – review and editing | W. J. Loughry, Resources, Writing – review and editing | Nico L. Avenant, Resources, Writing – review and editing | Benoit de Thoisy, Resources, Supervision, Writing – review and editing | Guillaume Borrel, Formal analysis, Writing – review and editing | Frédéric Delsuc, Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Resources, Supervision, Writing – review and editing, Funding acquisition, Project administration

DATA AVAILABILITY

Raw RNAseq short Illumina and long ONT reads have been submitted to the Short Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) and are available under BioProject number PRJNA942254. Metagenome assemblies, genome bins (all and selected), phylogenetic datasets (alignments and corresponding trees), and other supplementary materials and results are available from Zenodo 7995394.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Fig. S1 (mSystems00388-23-S0001.pdf). Phylogeny of the 314 high-quality selected bins reconstructed from long- and short-read assemblies.
Fig. S2 (mSystems00388-23-S0002.pdf). Myrmecophagous-specific clades within Bacteroidetes and Proteobacteria.

Fig. S3 (mSystems00388-23-S0003.svg). Bacterial families of host-species-specific and shared genome bins carrying or not carrying GH18 genes.
Fig. S4 (mSystems00388-23-S0004.svg). Detection of the 140 high-quality selected bins reconstructed from aardvark, ground pangolin, and southern aardwolf samples.
Table S1 (mSystems00388-23-S0005.xlsx). Raw results of the different analyses conducted on each gut metagenome to reconstruct high-quality genome bins from raw metagenomic data for each dataset (long- and short-reads).
Table S2 (mSystems00388-23-S0006.xlsx). Presence/absence of the 314 high-quality selected genome bins across the 29 gut metagenomes of the nine focal myrmecophagous species.
Table S3 (mSystems00388-23-S0007.pdf). Proportion of chitinolytic genome bins (genomes having at least one GH18 with an active chitinolytic site) detected in the nine focal myrmecophagous species.
Table S4 (mSystems00388-23-S0008.pdf). Detailed sample information for the eight soil samples collected in South Africa.

Open Peer Review

PEER REVIEW HISTORY (review-history.pdf). An accounting of the reviewer comments and feedback.

REFERENCES

1. Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol 6:776–788. https://doi.org/10.1038/nrmicro1978
2. Moeller AH, Sanders JG. 2020. Roles of the gut microbiota in the adaptive evolution of mammalian species. Philos Trans R Soc Lond B Biol Sci 375:20190597. https://doi.org/10.1098/rstb.2019.0597
3. Alessandri G, Rizzo SM, Ossiprandi MC, van Sinderen D, Ventura M. 2022. Creating an atlas to visualize the biodiversity of the mammalian gut microbiota. Curr Opin Biotechnol 73:28–33. https://doi.org/10.1016/j.copbio.2021.06.028
4. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI. 2008. Evolution of mammals and their gut microbes. Science 320:1647–1651. https://doi.org/10.1126/science.1155725
5. Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, Henrissat B, Knight R, Gordon JI. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science 332:970–974. https://doi.org/10.1126/science.1198719
6. Groussin M, Mazel F, Sanders JG, Smillie CS, Lavergne S, Thuiller W, Alm EJ. 2017. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. Nat Commun 8:14319. https://doi.org/10.1038/ncomms14319
7. Nishida AH, Ochman H. 2018. Rates of gut microbiome divergence in mammals. Mol Ecol 27:1884–1897. https://doi.org/10.1111/mec.14473
8. Youngblut ND, Reischer GH, Walters W, Schuster N, Walzer C, Stalder G, Ley RE, Farnleitner AH. 2019. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. 1. Nat Commun 10:2200. https://doi.org/10.1038/s41467-019-10191-3
9. Song SJ, Sanders JG, Delsuc F, Metcalf J, Amato K, Taylor MW, Mazel F, Lutz HL, Winker K, Graves GR, Humphrey G, Gilbert JA, Hackett SJ, White KP, Skeen HR, Kurtis SM, Withrow J, Braile T, Miller M, McCracken KG, Maley JM, Ezenwa VO, Williams A, Blanton JM, McKenzie VJ, Knight R, Graf J. 2020. Comparative analyses of vertebrate gut microbiomes reveal convergence between birds and bats. mBio 11:e02901-19. https://doi.org/10.1128/mBio.02901-19
10. Thomas CM, Desmond-Le Quéméner E, Gribaldo S, Borrel G. 2022. Factors shaping the abundance and diversity of the gut archaeome across the animal kingdom. 1. Nat Commun 13:3358. https://doi.org/10.1038/s41467-022-31038-4
11. Redford KH. 1987. Ants and termites as food, p. 349–399. In Genoways, HH (ed.), Current Mammalogy. Springer US, Boston, MA.
12. Delsuc F, Scally M, Madsen O, Stanhope MJ, de Jong WW, Catzeflis FM, Springer MS, Douzery EJP. 2002. Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. Mol Biol Evol 19:1656–1671. https://doi.org/10.1093/oxfordjournals.molbev.a003989
13. Meredith RW, Janečka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simão TLL, Stadler T, Rabosky DL, Honeycutt RL, Flynn JJ, Ingram CM, Steiner C, Williams TL, Robinson TJ, Burk-Herrick A, Westerman M, Ayoub NA, Springer MS, Murphy WJ. 2011. Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334:521–524. https://doi.org/10.1126/science.1211028
14. Álvarez-Carretero S, Tamuri AU, Battini M, Nascimento FF, Carlisle E, Asher RJ, Yang Z, Donoghue PCJ, Dos Reis M. 2022. A species-level timeline of mammal evolution integrating phylogenomic data. Nature 602:263–267. https://doi.org/10.1038/s41586-021-04341-1
15. Emerling CA, Delsuc F, Nachman MW. 2018. Chitinase genes (*CHIA*s) provide genomic footprints of a post-cretaceous dietary radiation in placental mammals. Sci Adv 4:eaar6478. https://doi.org/10.1126/sciadv.aar6478
16. Ma J-E, Li L-M, Jiang H-Y, Zhang X-J, Li J, Li G-Y, Yuan L-H, Wu J, Chen J-P. 2017. Transcriptomic analysis identifies genes and pathways related to myrmecophagy in the Malayan pangolin (*Manis javanica*). PeerJ 5:e4140. https://doi.org/10.7717/peerj.4140
17. Ma J-E, Jiang H-Y, Li L-M, Zhang X-J, Li H-M, Li G-Y, Mo D-Y, Chen J-P. 2019. SMRT sequencing of the full-length transcriptome of the Sunda pangolin (*Manis Javanica*). Gene 692:208–216. https://doi.org/10.1016/j.gene.2019.01.008
18. Allio R, Teullet S, Lutgen D, Magdeleine A, Koual R, Tilak M-K, Thoisy B de, Emerling CA, Lefébure T, Delsuc F. 2023. Comparative transcriptomics reveals divergent paths of Chitinase evolution underlying dietary convergence in ant-eating mammals. bioRxiv. https://doi.org/10.1101/2022.11.29.518312
19. Delsuc F, Metcalf JL, Wegener Parfrey L, Song SJ, González A, Knight R. 2014. Convergence of gut microbiomes in myrmecophagous mammals. Mol Ecol 23:1301–1317. https://doi.org/10.1111/mec.12501
20. Ma J-E, Jiang H-Y, Li L-M, Zhang X-J, Li G-Y, Li H-M, Jin X-J, Chen J-P. 2018. The fecal metagenomics of Malayan pangolins identifies an extensive adaptation to myrmecophagy. Front Microbiol 9:2793. https://doi.org/10.3389/fmicb.2018.02793
21. Cheng S-C, Liu C-B, Yao X-Q, Hu J-Y, Yin T-T, Lim BK, Chen W, Wang G-D, Zhang C-L, Irwin DM, Zhang Z-G, Zhang Y-P, Yu L. 2023. Hologenomic

insights into mammalian adaptations to myrmecophagy. Natl Sci Rev 10:wac174. https://doi.org/10.1093/nsr/nwac174
22. Gooday GW. 1990. The ecology of chitin degradation, p. 387–430. In Marshall, KC (ed.), Advances in Microbial Ecology. Springer US, Boston, MA.
23. Beier S, Bertilsson S. 2013. Bacterial chitin degradation-mechanisms and ecophysiological strategies. Front Microbiol 4:149. https://doi.org/10.3389/fmicb.2013.00149
24. Swiontek Brzezinska M, Jankiewicz U, Burkowska A, Walczak M. 2014. Chitinolytic microorganisms and their possible application in environmental protection. Curr Microbiol 68:71–81. https://doi.org/10.1007/s00284-013-0440-4
25. Raimundo I, Silva R, Meunier L, Valente SM, Lago-Lestón A, Keller-Costa T, Costa R. 2021. Functional metagenomics reveals differential chitin degradation and utilization features across free-living and host-associated marine microbiomes. Microbiome 9:43. https://doi.org/10.1186/s40168-020-00970-2
26. Stevens CE, Hume ID. 1998. Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. Physiol Rev 78:393–427. https://doi.org/10.1152/physrev.1998.78.2.393
27. Simůnek J, Hodrová B, Bartonová H, Kopecný J. 2001. Chitinolytic bacteria of the mammal digestive tract. Folia Microbiol (Praha) 46:76–78. https://doi.org/10.1007/BF02825892
28. Whitaker JO, Dannelly HK, Prentice DA. 2004. Chitinase in insectivorous bats. J Mammal 85:15–18. https://doi.org/10.1644/1545--1542(2004)085<0015:CIIB>2.0.CO;2
29. Irulan A, Nathan PT, Priya YS, Marimuthu S, Elangovan V. 2011. Isolation and characterization of chitinase producing gut microflora of insectivorous bats. Trends in Biosciences 4:8–11.
30. Macdonald C, Barden S, Foley S. 2014. Isolation and characterization of chitin-degrading micro-organisms from the faeces of Goeldi's monkey, *Callimico goeldii*. J Appl Microbiol 116:52–59. https://doi.org/10.1111/jam.12338
31. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res 50:D571–D577. https://doi.org/10.1093/nar/gkab1045
32. Deeba F, Shakir H, Irfan M, Qazi J. 2016. Chitinase production in organisms: a review. Punjab Univ J Zool 31:101–106.
33. Adrangi S, Faramarzi MA. 2013. From bacteria to human: a journey into the world of Chitinases. Biotechnol Adv 31:1786–1795. https://doi.org/10.1016/j.biotechadv.2013.09.012
34. Yan Q, Fong SS. 2015. Bacterial chitinase: nature and perspectives for sustainable bioproduction. Bioresour Bioprocess 2:31. https://doi.org/10.1186/s40643-015-0057-5
35. Rathore AS, Gupta RD. 2015. Chitinases from bacteria to human: properties, applications, and future perspectives. Enzyme Res 2015:791907. https://doi.org/10.1155/2015/791907
36. Sanders JG, Beichman AC, Roman J, Scott JJ, Emerson D, McCarthy JJ, Girguis PR. 2015. Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. Nat Commun 6:8285. https://doi.org/10.1038/ncomms9285
37. Kohl KD, Dieppa-Colón E, Goyco-Blas J, Peralta-Martínez K, Scafidi L, Shah S, Zawacki E, Barts N, Ahn Y, Hedayati S, Secor SM, Rowe MP. 2022. Gut microbial ecology of five species of sympatric desert rodents in relation to herbivorous and Insectivorous feeding strategies. Integr Comp Biol 62:237–251. https://doi.org/10.1093/icb/icac045
38. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2021. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res 50:D785–D794. https://doi.org/10.1093/nar/gkab776
39. Buist G, Steen A, Kok J, Kuipers OP. 2008. LysM, a widely distributed protein motif for binding to (peptido)glycans. Mol Microbiol 68:838–847. https://doi.org/10.1111/j.1365-2958.2008.06211.x
40. Musacchio A, Wilmanns M, Saraste M. 1994. Structure and function of the SH3 domain. Prog Biophys Mol Biol 61:283–297. https://doi.org/10.1016/0079-6107(94)90003-5
41. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, Trigodet F, Watson AR, Esen ÖC, Moore RM, Clayssen Q, Lee MD, Kivenson V, Graham ED, Merrill BD, Karkman A, Blankenberg D, Eppley JM, Sjödin A, Scott JJ, Vázquez-Campos X, McKay LJ, McDaniel EA, Stevens SLR, Anderson RE, Fuessel J, Fernandez-Guerra A, Maignien L, Delmont TO, Willis AD. 2021. Community-led, integrated, reproducible multi-omics with Anvi'o. Nat Microbiol 6:3–6. https://doi.org/10.1038/s41564-020-00834-3
42. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, Zhu Q, Bolzan M, Cumbo F, May U, Sanders JG, Zolfo M, Kopylova E, Pasolli E, Knight R, Mirarab S, Huttenhower C, Segata N. 2020. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat Commun 11:2500. https://doi.org/10.1038/s41467-020-16366-7
43. Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x
44. Suzuki TA. 2017. Links between natural variation in the microbiome and host fitness in wild mammals. Integr Comp Biol 57:756–769. https://doi.org/10.1093/icb/icx104
45. Smith S. 1996. Isolation and characterization of chitinases from *Dasypus novemcinctus*, the nine-banded armadillo. MSU Graduate Theses 184.
46. Olsen MA, Blix AS, Utsi T, Sørmo W, Mathiesen SD. 1999. Chitinolytic bacteria in the minke whale forestomach. Can J Microbiol 46:85–94. https://doi.org/10.1139/w99-112
47. Wardman JF, Bains RK, Rahfeld P, Withers SG. 2022. Carbohydrate-active enzymes (CAZymes) in the gut microbiome. Nat Rev Microbiol 20:542–556. https://doi.org/10.1038/s41579-022-00712-1
48. Olland AM, Strand J, Presman E, Czerwinski R, Joseph-McCarthy D, Krykbaev R, Schlingmann G, Chopra R, Lin L, Fleming M, Kriz R, Stahl M, Somers W, Fitz L, Mosyak L. 2009. Triad of polar residues implicated in pH specificity of acidic mammalian chitinase. Protein Sci 18:569–578. https://doi.org/10.1002/pro.63
49. Hamid R, Khan MA, Ahmad M, Ahmad MM, Abdin MZ, Musarrat J, Javed S. 2013. Chitinases: an update. J Pharm Bioallied Sci 5:21–29. https://doi.org/10.4103/0975-7406.106559
50. Pel R, Hessels G, Aalfs H, Gottschal JC. 1989. Chitin degradation by *Clostridium* sp. strain 9.1 in mixed cultures with Saccharolytic and sulfate-reducing bacteria. FEMS Microbiol Lett 62:191–200. https://doi.org/10.1111/j.1574-6968.1989.tb03693.x
51. Simůnek J, Kopecný J, Hodrová B, Bartonová H. 2002. Identification and characterization of *Clostridium paraputrificum*, a chitinolytic bacterium of human digestive tract. Folia Microbiol (Praha) 47:559–564. https://doi.org/10.1007/BF02818798
52. Sebaihia M, Peck MW, Minton NP, Thomson NR, Holden MTG, Mitchell WJ, Carter AT, Bentley SD, Mason DR, Crossman L, Paul CJ, Ivens A, Wells-Bennik MHJ, Davis IJ, Cerdeño-Tárraga AM, Churcher C, Quail MA, Chillingworth T, Feltwell T, Fraser A, Goodhead I, Hance Z, Jagels K, Larke N, Maddison M, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, White B, Whithead S, Parkhill J. 2007. Genome sequence of a proteolytic (group I) *Clostridium botulinum* strain hall A and comparative analysis of the clostridial genomes. Genome Res 17:1082–1092. https://doi.org/10.1101/gr.6282807
53. Leisner JJ, Larsen MH, Ingmer H, Petersen BO, Duus JØ, Palcic MM. 2009. Cloning and comparison of phylogenetically related chitinases from listeria monocytogenes EGD and *Enterococcus faecalis* V583. J Appl Microbiol 107:2080–2087. https://doi.org/10.1111/j.1365-2672.2009.04420.x
54. Vaaje-Kolstad G, Bøhle LA, Gåseidnes S, Dalhus B, Bjørås M, Mathiesen G, Eijsink VGH. 2012. Characterization of the chitinolytic machinery of *Enterococcus faecalis* V583 and high-resolution structure of its oxidative CBM33 enzyme. J Mol Biol 416:239–254. https://doi.org/10.1016/j.jmb.2011.12.033
55. Vacca M, Celano G, Calabrese FM, Portincasa P, Gobbetti M, De Angelis M. 2020. The controversial role of human gut Lachnospiraceae. Microorganisms 8:573. https://doi.org/10.3390/microorganisms8040573
56. La Reau AJ, Meier-Kolthoff JP, Suen G. 2016. Sequence-based analysis of the genus *Ruminococcus* resolves its phylogeny and reveals strong host association. Microb Genom 2:e000099. https://doi.org/10.1099/mgen.0.000099
57. Bøhle LA, Mathiesen G, Vaaje-Kolstad G, Eijsink VGH. 2011. An endo-β-N-acetylglucosaminidase from *Enterococcus faecalis* V583 responsible for the hydrolysis of high-mannose and hybrid-type N-linked glycans.

FEMS Microbiol Lett 325:123–129. https://doi.org/10.1111/j.1574-6968.2011.02419.x
58. Spaink HP. 2004. Specific recognition of bacteria by plant LysM domain receptor kinases. Trends Microbiol 12:201–204. https://doi.org/10.1016/j.tim.2004.03.001
59. Ferreira-Cardoso S, Delsuc F, Hautier L. 2019. Evolutionary tinkering of the mandibular canal linked to convergent regression of teeth in placental mammals. Curr Biol 29:468–475. https://doi.org/10.1016/j.cub.2018.12.023
60. Ferreira-Cardoso S, Fabre P-H, de Thoisy B, Delsuc F, Hautier L. 2020. Comparative masticatory myology in anteaters and its implications for interpreting morphological convergence in myrmecophagous placentals. PeerJ 8:e9690. https://doi.org/10.7717/peerj.9690
61. Ferreira-Cardoso S, Claude J, Goswami A, Delsuc F, Hautier L. 2022. Flexible conservatism in the skull modularity of convergently evolved myrmecophagous placental mammals. BMC Ecol Evol 22:87. https://doi.org/10.1186/s12862-022-02030-9
62. Perez-Lamarque B, Sommeria-Klein G, Duret L, Morlon H. 2023. Phylogenetic comparative approach reveals evolutionary conservatism, ancestral composition, and integration of vertebrate gut microbiota. Mol Biol Evol 40:msad144. https://doi.org/10.1093/molbev/msad144
63. Mallott EK, Amato KR. 2021. Host specificity of the gut microbiome. Nat Rev Microbiol 19:639–653. https://doi.org/10.1038/s41579-021-00562-3
64. Mazel F, Davis KM, Loudon A, Kwong WK, Groussin M, Parfrey LW, Bik H. 2018. Is host filtering the main driver of phylosymbiosis across the tree of life?. mSystems 3:e00097-18. https://doi.org/10.1128/mSystems.00097-18
65. Gull JM, Stahl M, Osmann C, Ortmann S, Kreuzer M, Hatt J-M, Clauss M. 2015. Digestive physiology of captive giant anteaters (*Myrmecophaga tridactyla*): determinants of faecal dry matter content. J Anim Physiol Anim Nutr (Berl) 99:565–576. https://doi.org/10.1111/jpn.12223
66. Bornbusch SL, Greene LK, Rahobilalaina S, Calkins S, Rothman RS, Clarke TA, LaFleur M, Drea CM. 2022. Gut microbiota of ring-tailed Lemurs (*Lemur catta*) vary across natural and captive populations and correlate with environmental microbiota. Anim Microbiome 4:29. https://doi.org/10.1186/s42523-022-00176-x
67. Perry T, West E, Eisenhofer R, Stenhouse A, Wilson I, Laming B, Rismiller P, Shaw M, Grützner F. 2022. Characterising the gut microbiomes in wild and captive short-beaked echidnas reveals diet-associated changes. Front Microbiol 13:687115. https://doi.org/10.3389/fmicb.2022.687115
68. Banerjee S, van der Heijden MGA. 2023. Soil microbiomes and one health 1. Nat Rev Microbiol 21:6–20. https://doi.org/10.1038/s41579-022-00779-w
69. Hehemann J-H, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. 2010. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature 464:908–912. https://doi.org/10.1038/nature08937
70. Zilber-Rosenberg I, Rosenberg E. 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiol Rev 32:723–735. https://doi.org/10.1111/j.1574-6976.2008.00123.x
71. Zepeda Mendoza ML, Xiong Z, Escalera-Zamudio M, Runge AK, Thézé J, Streicker D, Frank HK, Loza-Rubio E, Liu S, Ryder OA, Samaniego Castruita JA, Katzourakis A, Pacheco G, Taboada B, Löber U, Pybus OG, Li Y, Rojas-Anaya E, Bohmann K, Carmona Baez A, Arias CF, Liu S, Greenwood AD, Bertelsen MF, White NE, Bunce M, Zhang G, Sicheritz-Pontén T, Gilbert MPT. 2018. Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. Nat Ecol Evol 2:659–668. https://doi.org/10.1038/s41559-018-0476-8
72. Janiak MC, Chaney ME, Tosi AJ. 2018. Evolution of acidic mammalian chitinase genes (*CHIA*) is related to body mass and insectivory in primates. Mol Biol Evol 35:607–622. https://doi.org/10.1093/molbev/msx312
73. Eizirik E, Murphy WJ, Koepfli K-P, Johnson WE, Dragoo JW, Wayne RK, O'Brien SJ. 2010. Pattern and timing of diversification of the mammalian order carnivora inferred from multiple nuclear gene sequences. Mol Phylogenet Evol 56:49–63. https://doi.org/10.1016/j.ympev.2010.01.033
74. Redford KH. 1985. Food habits of armadillos (Xenarthra, Dasypodidae), p. 429–438. In G.G. Montgomery (ed.), The evolution and ecology of armadillos, sloths, and vermilinguas. Smithsonian Institution Press, Washington, DC.
75. Redford KH. 1986. Dietary specialization and variation in two mammalian myrmecophages (variation in mammalian myrmecophagy). Revista Chilena de Historia Natural 59:201–208.
76. Anacleto TC da S. 2007. Food habits of four armadillo species in the Cerrado area Mato Grosso, Brazil. Zoological Studies 46:529–537.
77. Vaz VC, Santori RT, Jansen AM, Delciellos AC, D'Andrea PS. 2012. Notes on food habits of armadillos (Cingulata, Dasypodidae) and Anteaters (Pilosa, Myrmecophagidae) at Serra da capivara national park (Piauí state, Brazil). Edentata 13:84–89. https://doi.org/10.5537/020.013.0107
78. Aya-Cuero C, Chacón-Pacheco J, Anacleto TCS, Solari S, Hamilton MJ. 2019. *Dasypus kappleri* (Cingulata: Dasypodidae). Mammalian Species 51:51–60. https://doi.org/10.1093/mspecies/sez009
79. Magdeleine A, Tilak M-K, Teullet S, Delsuc F. 2021. High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing. https://doi.org/10.17504/protocols.io.bvyin7ue
80. Leger A, Leonardi T. 2019. pycoQC, interactive quality control for oxford nanopore sequencing. JOSS 4:1236. https://doi.org/10.21105/joss.01236
81. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex minion sequencing. Microb Genom 3:e000132. https://doi.org/10.1099/mgen.0.000132
82. Wick RR. 2021. Filtlong (0.2.1). Available from: https://doi.org/10.1186/s13059-021-02483-z.
83. Li H, Birol I. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191
84. Li H, Alkan C. 2021. New strategies to improve minimap2 alignment accuracy. Bioinformatics 37:4572–4574. https://doi.org/10.1093/bioinformatics/btab705
85. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8
86. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199
87. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356:92–95. https://doi.org/10.1126/science.aal3327
88. Zimin AV, Puiu D, Luo M-C, Zhu T, Koren S, Marçais G, Yorke JA, Dvořák J, Salzberg SL. 2017. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res 27:787–792. https://doi.org/10.1101/gr.213405.116
89. Allio R. 2021. Phylogenomics and comparative genomics in ant-eating mammals. Doctoral dissertation. University of Montpellier.
90. Emerling CA, Gibb GC, Tilak M-K, Hughes JJ, Kuch M, Duggan AT, Poinar HN, Nachman MW, Delsuc F. 2022. Genomic data suggest parallel dental vestigialization within the xenarthran radiation. bioRxiv. https://doi.org/10.1101/2022.12.09.519446
91. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560
92. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2.4. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923
93. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
94. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17:1103–1110. https://doi.org/10.1038/s41592-020-00971-x

95. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963
96. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116
97. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033
98. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119
99. Eddy SR, Pearson WR. 2011. Accelerated profile HMM searches. PLOS Comput Biol 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195
100. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7:e7359. https://doi.org/10.7717/peerj.7359
101. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11:2864–2868. https://doi.org/10.1038/ismej.2017.126
102. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114
103. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 17:132. https://doi.org/10.1186/s13059-016-0997-x
104. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9
105. Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun 4:2304. https://doi.org/10.1038/ncomms3304
106. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176
107. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010
108. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348
109. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300
110. Coleman GA, Davín AA, Mahendrarajah TA, Szánthó LL, Spang A, Hugenholtz P, Szöllősi GJ, Williams TA. 2021. A rooted phylogeny resolves early bacterial evolution. Science 372:eabe0511. https://doi.org/10.1126/science.abe0511
111. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/nar/gky418
112. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Manuscript associated supplementary material

Table S1 Raw results of the different analyses conducted on each gut metagenome to reconstruct high-quality genome bins from raw metagenomic data for each dataset (long- and short-reads). Table available via Zenodo DOI: 10.5281/zenodo.7995394 (https://zenodo.org/record/7995394).

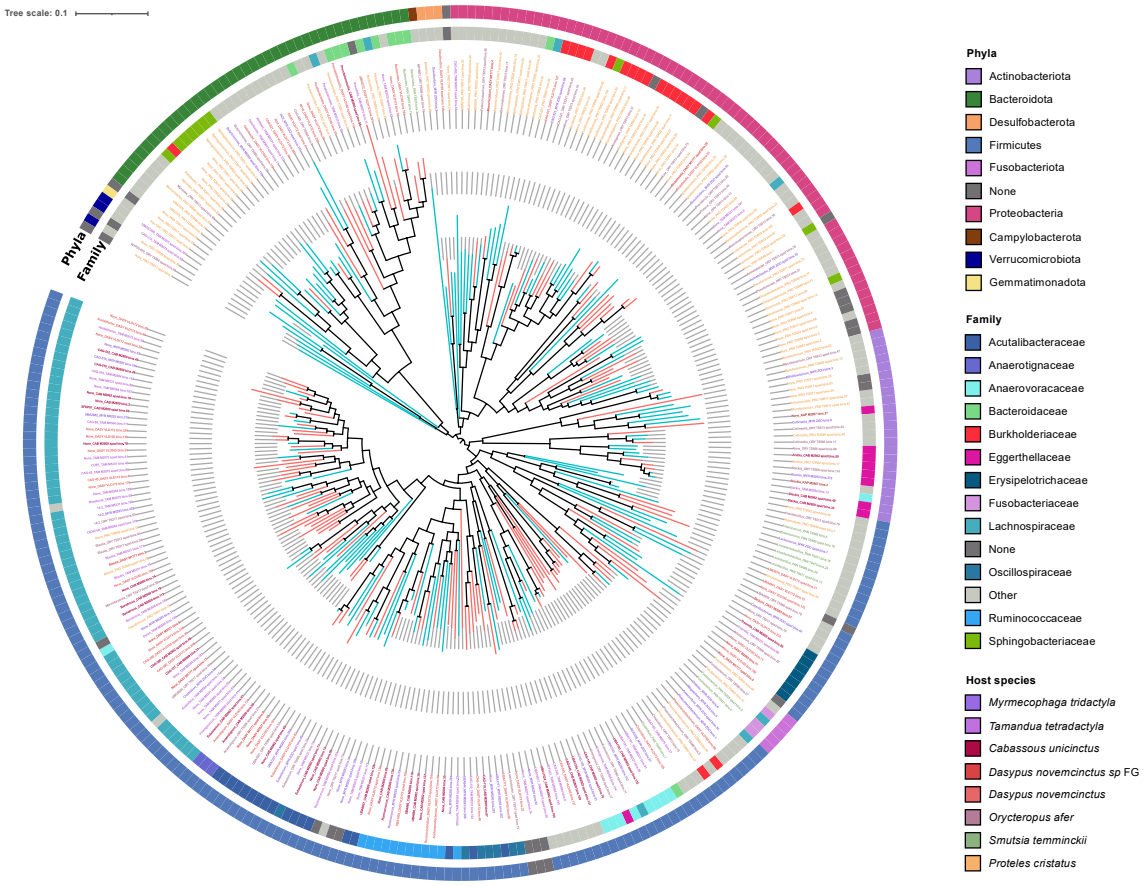


Fig S1 Phylogeny of the 314 high-quality selected bins reconstructed from long-read assemblies (n = 156; red branches) **and short-read assemblies** (n = 158; blue branches)**.** Circles respectively indicate (from inner to outer circles): the bacterial family and phyla the bin was assigned to based on the Genome Taxonomy Database (GTDB) release 7 (Parks *et al*, 2021). Colored sequence names indicate the host species. Bins' names of the myrmecophagous-specific clades are indicated at leaves of the phylogenetic tree together with the genus to which they were assigned to.

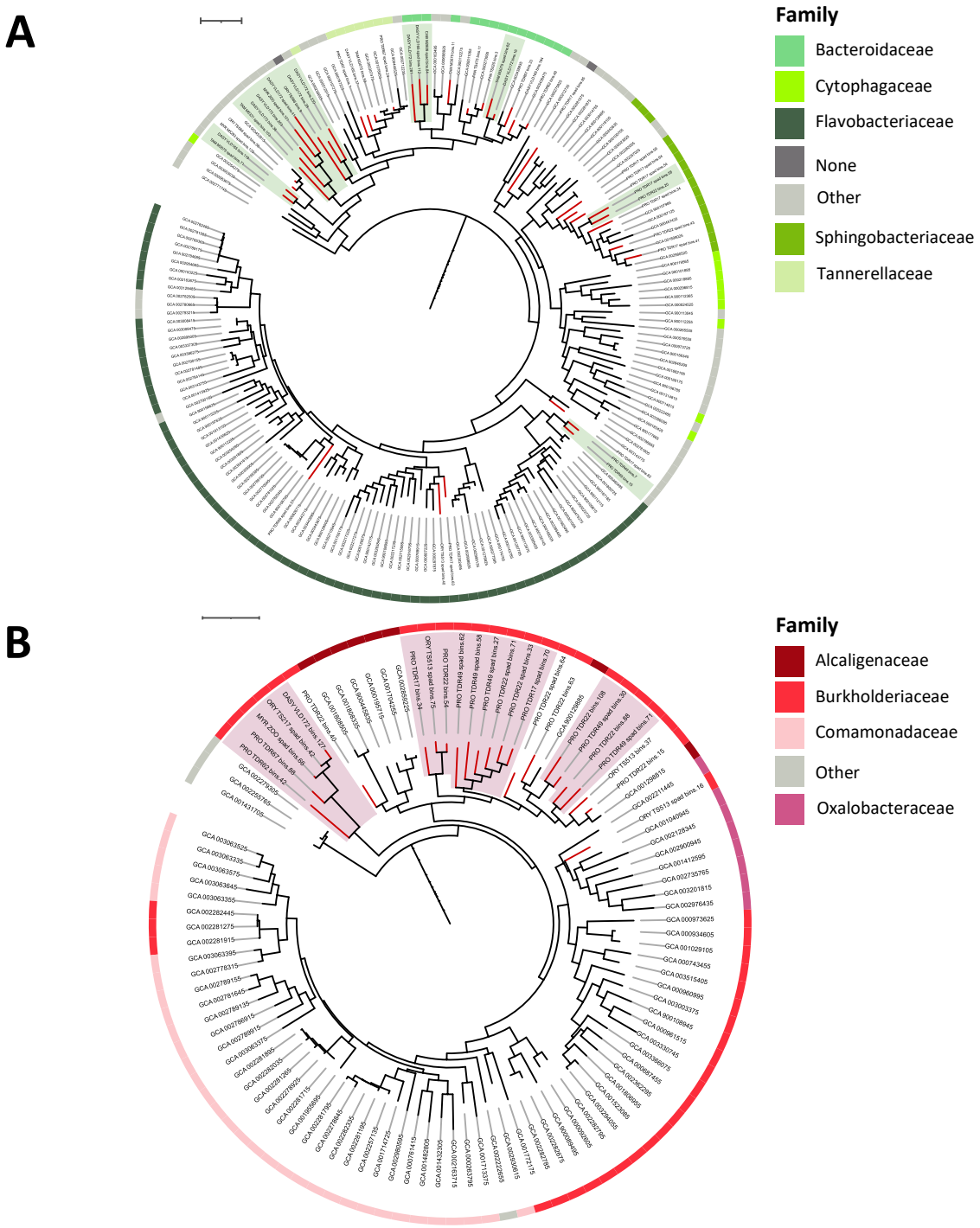


Fig S2 Myrmecophagous-specific clades within Bacteroidetes (A; green highlights) and **Proteobacteria** (B; pink highlights). The two trees are subtrees of Fig. 1. Outer circles indicate the bacterial family to which these genome bins were assigned based on the Genome Taxonomy Database (GTDB) release 7 (Parks *et al*, 2021).

Table S2 Presence/absence of the 314 high-quality selected genome bins across the 29 gut metagenomes of the nine focal myrmecophagous species. The taxonomy of each bin and the number of GH18 (including those with an active chitinolytic site) identified in each bin is also detailed. Table available via Zenodo DOI: 10.5281/zenodo.7995394 (https://zenodo.org/record/7995394).

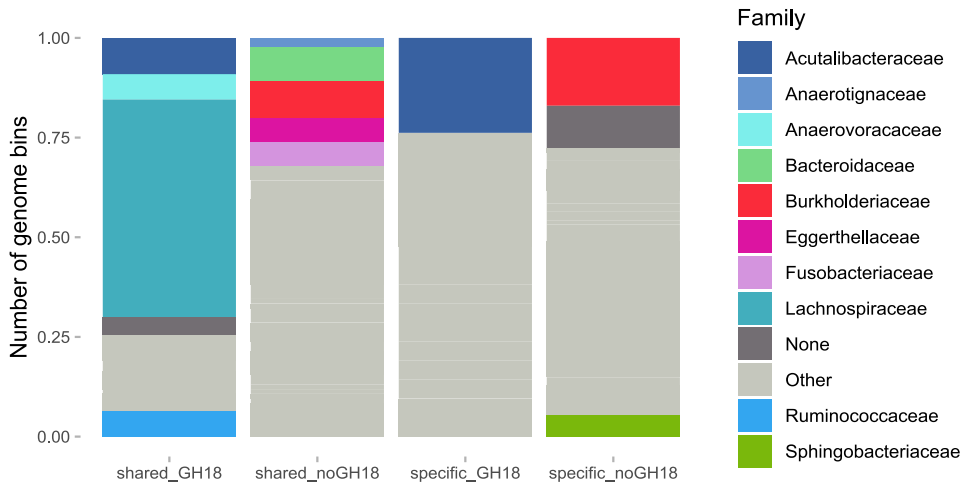


Fig S3 Bacterial families of host-species specific and shared genome bins carrying or not carrying GH18 genes.

Table S3 Proportion of chitinolytic genome bins (*i.e.*, genomes having at least one GH18 with an active chitinolytic site) **detected in the nine focal myrmecophagous species.**

Species	# chitinolytic bins detected	total # detected bins	# chitinolytic bins / total # bins detected
Cabassous unicinctus	46	113	0.407079646
***Dasypus sp. nov.* FG**	48	105	0.457142857
Dasypus novemcinctus	50	129	0.387596899
Dasypus kappleri	2	25	0.08
Myrmecophaga tridactyla	48	141	0.340425532
Orycteropus afer	15	97	0.154639175
Smutsia temminckii	5	18	0.277777778
Proteles cristatus	12	104	0.115384615
Tamandua tetradactyla	37	121	0.305785124

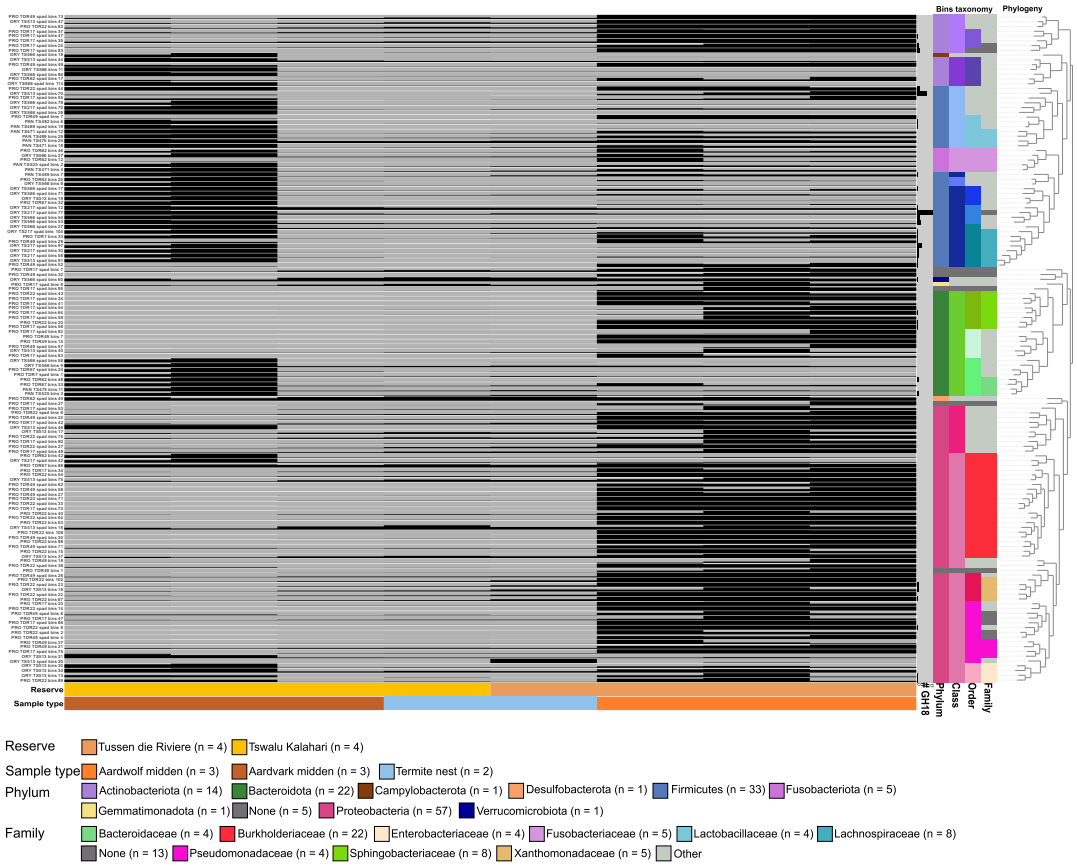


Fig S4 Detection of the 140 high-quality selected bins reconstructed from the aardvark, ground pangolin, and southern aardwolf samples (lines) **in the eight soil samples** (columns) **collected in South Africa near feces sampling sites.** Each square indicates the detection of a bin in a sample as estimated by anvi'o v7 (Eren *et al*, 2021). Bin names are indicated on the left. The provenance of each sample is indicated by different colors at the bottom of the graph. Columns on the right indicate (from left to right): the number of GH18 sequences identified in each bin (from 0 to 17), the bin's taxonomic phylum, class, order, and family. The phylogeny of the 140 selected bins inferred with PhyloPhlAn v3.0.58 (Asnicar *et al*, 2020) is also represented on the right of the graph. See detection table available via Zenodo (https://zenodo.org/record/7995394) for detailed values of detection.

Table S4 Detailed sample information for the eight soil samples collected in South Africa.

Sample	Sample type	Species	Common name	Class	Order	Family	Location
TDR012	Soil from Midden #1	*Proteles cristatus*	Southern ardwolf	Mammalia	Carnivora	Hyaenidae	Tussen Die Riviere Reserve
TDR014	Nest fragments and soil	*Trinervitermes sp.*	Snouted termites	Insecta	Blattodea	Termitidae	Tussen Die Riviere Reserve
TDR019	Soil from Midden #2	*Proteles cristatus*	Southern ardwolf	Mammalia	Carnivora	Hyaenidae	Tussen Die Riviere Reserve
TDR023	Soil from Midden #3	*Proteles cristatus*	Southern ardwolf	Mammalia	Carnivora	Hyaenidae	Tussen Die Riviere Reserve
TS218	Soil from midden	*Orycteropus afer*	Aardvark	Mammalia	Tubulidentata	Orycteropodidae	Tswalu Kalahari Reserve
TS274	Nest fragments and soil	*Trinervitermes sp.*	Snouted termites	Insecta	Blattodea	Termitidae	Tswalu Kalahari Reserve
TS281	Soil from midden	*Orycteropus afer*	Aardvark	Mammalia	Tubulidentata	Orycteropodidae	Tswalu Kalahari Reserve
TS298	Soil from midden	*Orycteropus afer*	Aardvark	Mammalia	Tubulidentata	Orycteropodidae	Tswalu Kalahari Reserve

II.3. Adaptations for chitin digestion in the gut microbiota of placental mammals

Context and short introduction

Studying the gut microbiota of myrmecophagous placental mammals revealed its potential to digest social insect prey (Teullet *et al*, 2023; see part II.2). Indeed, using genome-resolved metagenomics, genomes of chitinolytic bacteria were recovered and chitinase genes identified suggesting that such bacteria have the potential to digest the chitinous exoskeleton of insects. A diversity of bacterial chitinase genes have been retrieved and several genes identified in the same genome suggesting complex bacterial chitin-degrading pathways, which might participate in prey digestion. Comparing the distribution of such chitinolytic bacteria among myrmecophagous species revealed interesting patterns of shared and specific bacteria highlighting the potential influence of host phylogeny as well as diet and environment in shaping mammalian gut microbiota. Overall, this study shed light on the role of the gut microbiota in the convergent adaptation to myrmecophagy in placentals and the different microbial adaptations (*i.e*., similar or different microbial taxa) involved between the focal myrmecophagous host species.

Yet, to fully understand the specialization toward myrmecophagy in mammals, we need to identify potential patterns of taxonomic and functional convergence in the gut microbiota of myrmecophagous species when compared to non-myrmecophagous species. Comparing gut microbiota composition observed among myrmecophagous species and with their non-myrmecophagous sister-species should notably help understanding whether the host phylogeny (*i.e*., resulting in similar patterns between myrmecophagous species and their sister species), diet (*i.e*., resulting in similar patterns among myrmecophagous species) or even both, have shaped the gut microbiota of these species. Taxonomic comparisons based on 16S rRNA barcoding data between several ant- and termite-eating species and their sister-species have already revealed compositional convergences among myrmecophagous species gut microbiota and differences in abundance of some microbial taxa in myrmecophagous hosts (Delsuc *et al*, 2014). Such kind of study nevertheless lacks information on the functions carried by the microbial symbionts and whether microbial taxa that are more abundant in

myrmecophagous species ensure similar functions. Metagenomic studies on the Malayan pangolin and giant anteater partially answered this question by demonstrating that chitin-degrading enzymes are present in the gut microbiota of these ant-eating species (Ma *et al*, 2018; Cheng *et al*, 2023). Moreover, comparing gut metagenomes of the giant anteater, Malayan pangolin, and dog revealed that bacteria that are more abundant in myrmecophagous species carry chitinases (Cheng *et al*, 2023), which suggest that they represent an adaptation to myrmecophagy in mammals.

Here, to fully understand patterns of taxonomic and functional convergence in the gut microbiota of myrmecophagous species, we will study gut metagenomes of species representatives of the five myrmecophagous orders as well as their close relatives (*e.g*., sloths for anteaters, hyaenas for the aardwolf) and focus on microbial chitin-degrading pathways. In addition, including gut metagenomes of other mammalian species with a chitin-rich diet, such as insectivorous species (*e.g*., meerkats, elephant-shrews, some rodents) and marine mammals ingesting crustaceans (*e.g*., baleen whales), will allow further understanding the role played by the gut microbiota in chitin digestion and help potentially revealing convergent adaptations between species with similar diets.

This work is conducted in collaboration with Guillaume Borrel (Institut Pasteur, Paris, France). The dataset was finalized and the metagenome-assembled genomes reconstructed close to the end of my PhD. Thus, only the dataset will be presented here. Some of the analyses that could be conducted are exposed in the following parts.

Preliminary material and methods

➔ **Building a dataset of mammalian gut metagenomes**

Shotgun metagenomic short-read data from published studies such as the Earth Microbiome Project 500 (EMP500) consortium (Shaffer *et al*, 2022), collaborators (Guillaume Borrel), and data generated as part of the ConvergeAnt project will be used for this study (Table S II.1).

This dataset comprises a total of 328 samples representing 115 mammalian species (Table S II.1). Metagenomic data generated as part of the ConvergeAnt project were obtained from field-collected fecal samples of myrmecophagous species and non-myrmecophagous

closely related species sampled in French Guiana, South Africa, and the USA. Feces were preserved in 96% ethanol at -20°C. Total DNA was extracted using the "Genomic DNA from soil" extraction kit (Macherey-Nagel) following the DNA extraction steps of the Magdeleine *et al* (2021) protocol presented in part II.1.2. Library preparation and short-read 150 PE sequencing on an Illumina NovaSeq instrument were outsourced to Novogene Europe (Cambridge, UK). In total, 68 gut metagenomes of nine myrmecophagous and 12 non-myrmecophagous species were generated (Table S II.1) including 31 samples already published in the study by Teullet *et al* (2023) presented in part II.2.

Publicly available myrmecophagous gut metagenomes for the Malayan pangolin (*Manis javanica*; n = 8) and the giant anteater (*Myrmecophaga tridactyla*; n = 6) (Ma *et al*, 2018; Cheng *et al*, 2023) were added to the dataset (Table S II.1). Ant- and termite-eating species included in this dataset represent the five convergent myrmecophagous placental orders:

- Pangolins (Pholidota): *M. javanica* and the ground pangolin (*Smutsia temminckii*).
- Anteaters (Pilosa): *M. tridactyla* and the southern tamandua (*Tamandua tetradactyla*).
- Armadillos (Cingulata): the southern naked-tailed armadillo (*Cabassous unicinctus*), the nine-banded armadillo (*Dasypus novemcinctus*), the Guianan long-nosed armadillo (*Dasypus* sp. nov. SG), the greater long-nosed armadillo (*Dasypus kappleri*), and the southern three-banded armadillo (*Tolypeutes matacus*).
- Aardvark (Tubulidentata): *Orycteropus afer.*
- Aardwolf (Carnivora): *Proteles cristatus*.

Some of these species such as the long-nosed armadillos are not fully myrmecophagous species as their diet can include other food sources (*e.g.*, insects, worms). Closely related species for comparisons include sloths for anteaters (*Bradypus tridactylus*, *Choloepus didactylus*, *Choloepus hoffmanni*), Afrotherian species (*e.g.*, *Macroscelides proboscideus*, *Loxodonta Africana*, *Procavia capensis*) for the aardvark, and carnivoran species (*e.g.*, *Suricata suricatta*, *Puma yagouaroundi*, *Potos flavus*, *Canis mesomelas*) for the pangolins and the aardwolf. As they were available from Youngblut *et al* (2020), two samples of the short-beaked echidna (*Tachyglossus aculeatus*, Monotremata), which includes important quantities of ants in its diet, were added to this dataset for comparison with placental myrmecophagous species.

Other publicly available gut metagenomes were included in the dataset to maximize phylogenetic diversity at the mammalian scale. Notably, several mammalian species of the vertebrate gut metagenomes dataset of Youngblut *et al* (2020) and the EMP500 (Shaffer *et al*, 2022; https://earthmicrobiome.org/emp500/) were incorporated as well as samples from black rhinos (Gibson *et al*, 2019), giant pandas, Asiatic black bears, and bamboo rats (Guo *et al*, 2018), several primate species (Orkin *et al*, 2019; Sharma *et al*, 2020; Yan *et al*, 2021), and pigs (Xiao *et al*, 2016).

Non-myrmecophagous species with a chitin-rich diet included in this dataset are notably represented by ten metagenomes of four species of whales (from Sanders *et al*, 2015) as well as insectivorous species such as the meerkat (*Suricata suricatta*; n = 4), the round-eared elephant shrew (*Macroscelides proboscideus*; n = 1), the lesser hedgehog tenrec (*Echinops telfairi*; n = 1), the dwarf mongoose (*Helogale parvula*; n = 1), the banded mongoose (*Mungos mungos*; n = 1), and the sloth bear (*Melursus ursinus*; n = 1) which is also known to include ants in its diet.

➔ **Metagenomes assembly and genome binning**

Short-read metagenomes were assembled using metaSPAdes v3.11.0 (Nurk *et al*, 2017) and MEGAHIT v1.1.2 (Li *et al*, 2015) with default parameters. Genome binning was conducted using MetaBAT v1.2.15 (Kang *et al*, 2015) and METABAT2 v2.15 (Kang *et al*, 2019) with default parameters. To select bins with the highest quality score between the different combinations of assemblers and genome binners used, a first dereplication was performed with dRep v3.2.2 (Olm *et al*, 2017). Prior to dereplication, reads were mapped against genome bins and reassembled with the *reassemble_bins* module of MetaWRAP v1.3.2 (Uritskiy *et al*, 2018) to improve completion and reduce redundancy. A second dereplication was done on the set of all selected genome bins to remove redundant bins with an Average Nucleotide Identity set to 95, 99, and 100%. Taxonomy, completeness, and redundancy of genome bins were assessed with CheckM v1.0.7 (Parks *et al*, 2015).

In total, 17 624 genome bins were reconstructed. Among them, 3 943 genome bins were dereplicated with an ANI of 95% (the value usually used for species-level dereplication; Olm *et al*, 2020), and had a completeness higher than 90% and a redundancy smaller than 5%,

representing high-quality genome bins. Their completeness varied from 90.02% to 100% (mean = 95.25%) and their redundancy from 0% to 4.94% (mean = 0.6569%). Their genome length ranged from 614 983 bp to 8 322 128 bp (mean = 2 350 899 bp) and their scaffold N50 from 5 811 bp to 4 131 223 bp (mean = 106 347 bp). All these high-quality genome bins were taxonomically assigned to Bacteria and mainly to the following phyla (more than 100 genomes reconstructed): Firmicutes (n = 2 337), Bacteroidota (n = 915), Proteobacteria (n = 209), Actinobacteria (n = 170), and Spirochaetota (n = 115).

Prospects: future analyses

With this set of high-quality metagenome-assembled genomes, the aim will be to investigate the presence and study the distribution of chitinolytic bacteria between host species to understand the role of the gut microbiota in chitin digestion in placental species with diverse diets.

One of the main objectives of this study is to understand whether myrmecophagous species carry specific chitinolytic bacteria that are found uniquely in these species and not in other species, which would highlight potential microbial taxa associated with the myrmecophagous diet. Besides, understanding whether the same bacteria have been convergently recruited in myrmecophagous species will help to decipher the underlying adaptive mechanisms involved in insect prey digestion in these species. Indeed, because of their divergent evolutionary histories (*e.g.*, phylogenetic constraints, environmental influences), myrmecophagous species might have inherited and acquired different bacteria. Besides, having few highly abundant chitinolytic bacteria might have been selected in some species whereas in others having diverse chitinolytic bacteria might be more efficient to digest prey. Importantly, characterizing the presence of chitinolytic bacteria in other species having chitin-rich diets (*e.g.*, crustacean-eating species like baleen whales, or less specialized insectivorous species such as meerkats or sloth bears) will further shed light on the microbial adaptations involved in chitin digestion in placentals. Indeed, the common ancestor of mammals potentially had a gut microbiota composition close to the one of insectivorous species (Perez-Lamarque *et al*, 2023) suggesting its composition changed as placentals diversified. Therefore, some bacteria shared between different species with chitin-rich diets

may correspond to ancient bacteria inherited from their common ancestor that have been retained because of their important role for prey digestion.

Moreover, chitinolytic bacteria have been identified in mammalian species having diverse diets (*e.g*., Šimůnek *et al*, 2001; Macdonald *et al*, 2014; Whitaker *et al*, 2004; Sanders *et al*, 2015), therefore genomes of chitin-degrading bacteria might be shared between several placental species. The main question will then reside in understanding whether myrmecophagous species (and other insectivorous species) possess more chitinolytic bacteria in their gut microbiota (as they might need them to ensure prey digestion) than other non-myrmecophagous species. Furthermore, to disentangle the effects of the host species diet and phylogeny, it is necessary to compare the composition of the gut microbiota of myrmecophagous species with the one of their non-myrmecophagous sister-species to reveal potential patterns of convergence among ant- and termite-eating species. Indeed, phylosymbiosis is particularly strong in mammals (*e.g*., Song *et al*, 2020; Mallott and Amato, 2021) meaning that closely related species share more similar microbiomes than distantly related species.

To answer such questions, similar analyses as the ones performed in Teullet *et al* (2023) could be done to explore the dataset. For instance, GH18 genes could be identified in the reconstructed bacterial genomes to first identify putative chitinolytic bacteria. Such preliminary analyses will unravel the diversity of chitinolytic bacteria present in placental gut microbiota. Co-diversification (*i.e*., diversification of microbial symbionts following host species diversification) analyses will allow the identification of shared and host-specific bacterial taxa which would help assessing whether some chitinolytic bacteria are found only in myrmecophagous species or are shared with other species, notably closely related ones. Moreover, it could help us to assess whether myrmecophagous species carry more diverse chitinolytic bacteria than other species and whether some are shared among ant- and termite-eating species. More generally, such analysis will help us understand the evolution of chitin-degrading bacteria in the gut microbiota of placentals. For instance, Sanders *et al* (2023) used this approach on a set of primate gut MAGs and demonstrated that bacterial taxa, especially within the Firmicutes, Actinobacteria, and Bacteroidetes, co-diversified with their hosts but many were lost in humans which could be associated with lifestyle changes. Similar patterns could be observed in our dataset, for instance if some chitinolytic bacteria are inherited from

a common ancestor, insectivorous species might have retained them while non-insectivorous sister-species might have lost these microbial taxa. Additionally, such analysis could reveal potential bacterial diversification within specific host species.

Having high-quality bacterial genomes should also allow the comparative study of chitin-degrading microbial pathways. For instance, biosynthetic gene clusters (BGCs) could be identified and associated to specific functional categories (*e.g*., Cluster of Orthologous Groups (COGs), carbohydrate active enzymes families (CAZymes), KEGG pathways) to identify functions carried by bacteria and investigate their distribution in the different host species. For instance, such analysis might highlight enrichment of certain CAZymes families known to contain bacterial chitin-degrading or chitin-binding enzymes (*i.e*., GH18, GH19, or GH20) in host species with chitin-rich diets. Youngblut *et al* (2020) performed such analysis on a set of 1 522 species-level genome bins recovered from vertebrate gut metagenomes. They identified 1 986 BGCs and demonstrated that different functions are enriched in host-associated compared to environment-associated bacteria.

Finally, results from such study could be discussed in the light of chitinase genes (*CHIAs*) found in the host genome (see Chapter I). Indeed, it has been shown that from a placental ancestor probably carrying five *CHIA* genes, several gene losses occurred in non-insectivorous species during the placental radiation (Emerling *et al*, 2018). Those genes could participate in prey digestion in species with a chitin-rich diet as they were found expressed in several digestive organs in two ant-eating species (Allio *et al*, 2023; see chapter I). Combining metagenomic and genomic data should thus allow disentangling the respective contributions of genomic adaptations of the host and its associated gut microbiome in ensuring one function (*i.e*., chitin digestion). In the case of myrmecophagous species, it will be of particular interest for species like the aardwolf, which adopted this specialized diet relatively recently (< 10 Mya). Indeed, this species does not show striking morphological adaptations to myrmecophagy and possesses only one functional *CHIA* gene (*CHIA5*) in its genome (Allio *et al*, 2023). This gene does not seem to be expressed in its salivary glands unlike other myrmecophagous species like the southern tamandua and Malayan pangolin (Allio *et al*, 2023). Therefore, the aardwolf might potentially rely more on its gut microbiota to digest its preys than other more ancient myrmecophagous species such as pangolins or anteaters, as shifts in the gut microbiota composition could have happened more easily and rapidly than genomic adaptations.

Combining genomic and metagenomic data could therefore help us to better understand the complex adaptive mechanisms involved in the convergent evolution of myrmecophagy in mammals and the role the holobiont (*i.e.*, the host and its associated microorganisms) played in their specialization to this diet (see the general discussion).

Supplementary data

Table S II.1. Detailed information of the 329 mammalian samples used for the comparative study of gut metagenomes.

Table S II.1. Detailed information of the 329 mammalian samples used for the comparative study of gut metagenomes.

Sample name	Order	Family	Genus	Species	Host common name	Diet percentage of invertebrates	Captive/wild	Country of origin	Sample type	Sex	Age	Study	Study accession
Mp01	Pholidota	Manidae	*Manis*	*Manis javanica*	Malayan pangolin	100	Captive	China	feces	NA	NA	Cheng et al 2023 NatlSciRev	PRJCA005131
Mp02	Pholidota	Manidae	*Manis*	*Manis javanica*	Malayan pangolin	100	Captive	China	feces	NA	NA	Cheng et al 2023 NatlSciRev	PRJCA005131
Mp02_Colon	Pholidota	Manidae	*Manis*	*Manis javanica*	Malay pangolin	100	Captive	China	gut	NA	NA	Cheng et al 2023 NatlSciRev	PRJCA005131
Mp03	Pholidota	Manidae	*Manis*	*Manis javanica*	Malayan pangolin	100	Captive	China	feces	NA	NA	Cheng et al 2023 NatlSciRev	PRJCA005131
GA01	Pilosa	Myrmecophagidae	*Myrmecophaga*	*Myrmecophaga tridactyla*	Giant anteater	100	Captive	China	feces	NA	NA	Cheng et al 2023 NatlSciRev	PRJCA005131
GA02	Pilosa	Myrmecophagidae	*Myrmecophaga*	*Myrmecophaga tridactyla*	Giant anteater	100	Captive	China	feces	NA	NA	Cheng et al 2023 NatlSciRev	PRJCA005131
GA03	Pilosa	Myrmecophagidae	*Myrmecophaga*	*Myrmecophaga tridactyla*	Giant anteater	100	Captive	China	feces	NA	NA	Cheng et al 2023 NatlSciRev	PRJCA005131
GA04	Pilosa	Myrmecophagidae	*Myrmecophaga*	*Myrmecophaga tridactyla*	Giant anteater	100	Captive	China	feces	NA	NA	Cheng et al 2023 NatlSciRev	PRJCA005131
GA05	Pilosa	Myrmecophagidae	*Myrmecophaga*	*Myrmecophaga tridactyla*	Giant anteater	100	Captive	China	feces	NA	NA	Cheng et al 2023 NatlSciRev	PRJCA005131
GA06	Pilosa	Myrmecophagidae	*Myrmecophaga*	*Myrmecophaga tridactyla*	Giant anteater	100	Captive	China	feces	NA	NA	Cheng et al 2023 NatlSciRev	PRJCA005131
Btaurus76S002	Artiodactyla	Bovidae	*Bos*	*Bos taurus*	Cattle	0	Captive	Malawi	feces	NA	NA	EMP500	PRJEB42019
Cprevosti51S013	Rodentia	Sciuridae	*Callosciurus*	*Callosciurus prevosti*	Prevost's squirrel	20	Captive	USA	feces	NA	NA	EMP500	PRJEB42019
Choffmanni51S019	Pilosa	Choloepodidae	*Choloepus*	*Choloepus hoffmanni*	Hoffmanns Two-toed sloth	0	Captive	USA	feces	NA	NA	EMP500	PRJEB42019
Dbicornis51S003	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis michaeli*	Eastern black rhinoceros	0	Captive	USA	feces	NA	NA	EMP500	PRJEB42019
Etelfairi51S014	Afrosoricida	Tenrecomorpha	*Echinops*	*Echinops telfairi*	Lesser hedgehog tenrec	60	Captive	USA	feces	NA	NA	EMP500	PRJEB42019
Efprzewalskii51S008	Perissodactyla	Equidae	*Equus*	*Equus ferus przewalskii*	Przewalski horse	0	Captive	USA	feces	Female	NA	EMP500	PRJEB42019
Fmargarita51S010	Carnivora	Felidae	*Felis*	*Felis margarita*	Sand cat	0	Captive	USA	feces	Female	NA	EMP500	PRJEB42019
Ggorilla51S016	Primates	Hominidae	*Gorilla*	*Gorilla gorilla gorilla*	Western gorilla	0	Captive	USA	feces	NA	NA	EMP500	PRJEB42019
Hparvulus51S009	Carnivora	Herpestidae	*Helogale*	*Helogale parvula*	Dwarf mongoose	70	Captive	USA	feces	NA	NA	EMP500	PRJEB42019
Hantinena51S006	Rodentia	Nesomyidae	*Hypogeomys*	*Hypogeomys antinena*	Malagasy giant rat	0	Captive	USA	feces	NA	NA	EMP500	PRJEB42019
Lcatta51S004	Primates	Lemuridae	*Lemur*	*Lemur catta*	Ringtailed lemur	0	Captive	USA	feces	Male	NA	EMP500	PRJEB42019
Lafricana51S018	Proboscidea	Elephantidae	*Loxodonta*	*Loxodonta africana africana*	Southern African bush elephant	0	Captive	USA	feces	Female	NA	EMP500	PRJEB42019
Mleucophaeus51S015	Primates	Cercopithecidae	*Mandrillus*	*Mandrillus leucophaeus*	Drill	10	Captive	USA	feces	NA	NA	EMP500	PRJEB42019
Mursinus51S007	Carnivora	Ursidae	*Melursus*	*Melursus ursinus*	Sloth bear	70	Captive	USA	feces	NA	NA	EMP500	PRJEB42019
Mmungos51S012	Carnivora	Herpestidae	*Mungos*	*Mungos mungos*	Banded mongoose	80	Captive	USA	feces	NA	NA	EMP500	PRJEB42019
Mtridactyla51S020	Pilosa	Myrmecophagidae	*Myrmecophaga*	*Myrmecophaga tridactyla*	Giant anteater	100	Captive	USA	feces	NA	NA	EMP500	PRJEB42019
Nprocyonoides51S002	Carnivora	Canidae	*Nyctereutes*	*Nyctereutes procyonoides*	Raccoon dog	20	Captive	USA	feces	Male	NA	EMP500	PRJEB42019
Sscrofa76S005	Artiodactyla	Suidae	*Sus*	*Sus scrofa*	Pig	10	Captive	Malawi	feces	NA	NA	EMP500	PRJEB42019
Ttetradactyla51S005	Pilosa	Myrmecophagidae	*Tamandua*	*Tamandua tetradactyla*	Southern tamandua	100	Captive	USA	feces	NA	NA	EMP500	PRJEB42019
Tmatacus51S005	Cingulata	Chlamyphoridae	*Tolypeutes*	*Tolypeutes matacus*	Southern three-banded armadillo	80	Captive	USA	feces	Male	NA	EMP500	PRJEB42019
Tlatirostris75S003	Sirenia	Trichechidae	*Trichechus*	*Trichechus manatus latirostris*	Florida manatee	0	Captive	USA	feces	Male	NA	EMP500	PRJEB42019
Tlatirostris75S005	Sirenia	Trichechidae	*Trichechus*	*Trichechus manatus latirostris*	Florida manatee	0	Captive	USA	feces	Male	NA	EMP500	PRJEB42019
Bborealis50S012	Artiodactyla	Balaenopteridae	*Balaenoptera*	*Balaenoptera borealis*	Sei whale	60	Wild	Canada	feces	NA	Adult	EMP500 Sanders et al 2015 NatCom	PRJEB42019
BphysalusFinBP1	Artiodactyla	Balaenopteridae	*Balaenoptera*	*Balaenoptera physalus*	Fin whale	70	Wild	USA	feces	NA	Adult	EMP500 Sanders et al 2015 NatCom	PRJEB42019
BphysalusFinBP2	Artiodactyla	Balaenopteridae	*Balaenoptera*	*Balaenoptera physalus*	Fin whale	70	Wild	USA	feces	NA	Adult	EMP500 Sanders et al 2015 NatCom	PRJEB42019
EglacialisG3	Artiodactyla	Balaenidae	*Eubalaena*	*Eubalaena glacialis*	North Atlantic right whale	100	Wild	Canada	feces	NA	Adult	EMP500 Sanders et al 2015 NatCom	PRJEB42019
MnovaeangliaeAK1	Artiodactyla	Balaenopteridae	*Megaptera*	*Megaptera novaeangliae*	Humpback whale	80	Wild	USA	feces	NA	Adult	EMP500 Sanders et al 2015 NatCom	PRJEB42019
MnovaeangliaeAK4	Artiodactyla	Balaenopteridae	*Megaptera*	*Megaptera novaeangliae*	Humpback whale	80	Wild	USA	feces	NA	Adult	EMP500 Sanders et al 2015 NatCom	PRJEB42019
R15	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Male	Adult	Gibson et al 2019 SciRep	PRJNA532626
R18	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Female	Senior_adult	Gibson et al 2019 SciRep	PRJNA532626
R19	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Male	Juvenile	Gibson et al 2019 SciRep	PRJNA532626
R20	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Male	Adult	Gibson et al 2019 SciRep	PRJNA532626
R25	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Captive	South_Africa	feces	Female	Adult	Gibson et al 2019 SciRep	PRJNA532626
R27	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Captive	South_Africa	feces	Female	Adult	Gibson et al 2019 SciRep	PRJNA532626
R03	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Male	Juvenile	Gibson et al 2019 SciRep	PRJNA532626
R08	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Female	Adult	Gibson et al 2019 SciRep	PRJNA532626
R12	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Male	Adult	Gibson et al 2019 SciRep	PRJNA532626
R16	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Female	Senior_adult	Gibson et al 2019 SciRep	PRJNA532626
R21	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Captive	South_Africa	feces	Female	Adult	Gibson et al 2019 SciRep	PRJNA532626
R23	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Captive	South_Africa	feces	Female	Adult	Gibson et al 2019 SciRep	PRJNA532626
R24	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Captive	South_Africa	feces	Male	Adult	Gibson et al 2019 SciRep	PRJNA532626
R26	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Captive	South_Africa	feces	Female	Adult	Gibson et al 2019 SciRep	PRJNA532626
R04	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Female	Senior_adult	Gibson et al 2019 SciRep	PRJNA532626
R02	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Female	Senior_adult	Gibson et al 2019 SciRep	PRJNA532626
R01	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Male	Senior_adult	Gibson et al 2019 SciRep	PRJNA532626
R07	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Female	Adult	Gibson et al 2019 SciRep	PRJNA532626
R06	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Male	Senior_adult	Gibson et al 2019 SciRep	PRJNA532626
R05	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Female	Senior_adult	Gibson et al 2019 SciRep	PRJNA532626
R11	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Male	Adult	Gibson et al 2019 SciRep	PRJNA532626
R14	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Male	Adult	Gibson et al 2019 SciRep	PRJNA532626
R17	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Wild	South_Africa	feces	Male	Adult	Gibson et al 2019 SciRep	PRJNA532626
R22	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Captive	South_Africa	feces	Male	Adult	Gibson et al 2019 SciRep	PRJNA532626
R28	Perissodactyla	Rhinocerotidae	*Diceros*	*Diceros bicornis*	Black rhino	0	Captive	South_Africa	feces	Female	Adult	Gibson et al 2019 SciRep	PRJNA532626
FY1.27	Carnivora	Ursidae	*Ailuropoda*	*Ailuropoda melanoleuca*	Giant panda	0	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
WG	Carnivora	Ursidae	*Ailuropoda*	*Ailuropoda melanoleuca*	Giant panda	0	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
HH	Carnivora	Ursidae	*Ailuropoda*	*Ailuropoda melanoleuca*	Giant panda	0	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
GZ1.9.2	Carnivora	Ursidae	*Ailuropoda*	*Ailuropoda melanoleuca*	Giant panda	0	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
ZM2.1.27	Carnivora	Ursidae	*Ailuropoda*	*Ailuropoda melanoleuca*	Giant panda	0	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
BR1	Rodentia	Spalacidae	*Rhizomys*	*Rhizomys sinensis*	Bamboo rat	0	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
BR2	Rodentia	Spalacidae	*Rhizomys*	*Rhizomys sinensis*	Bamboo rat	0	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
BR3	Rodentia	Spalacidae	*Rhizomys*	*Rhizomys sinensis*	Bamboo rat	0	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
BR4	Rodentia	Spalacidae	*Rhizomys*	*Rhizomys sinensis*	Bamboo rat	0	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
CB7	Carnivora	Ursidae	*Ursus*	*Ursus thibetanus*	Asiatic black bear	10	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
CB9	Carnivora	Ursidae	*Ursus*	*Ursus thibetanus*	Asiatic black bear	10	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
CB1	Carnivora	Ursidae	*Ursus*	*Ursus thibetanus*	Asiatic black bear	10	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
CB3	Carnivora	Ursidae	*Ursus*	*Ursus thibetanus*	Asiatic black bear	10	Captive	China	feces	NA	NA	Guo et al 2018 FrontiersMicrobiol	PRJNA407583
A-1	Pholidota	Manidae	*Manis*	*Manis javanica*	Malayan pangolin	100	Captive	China	feces	Female	Adult	Ma et al 2018 FrontiersMicrobiol	PRJNA476660
A-2	Pholidota	Manidae	*Manis*	*Manis javanica*	Malayan pangolin	100	Captive	China	feces	Female	Adult	Ma et al 2018 FrontiersMicrobiol	PRJNA476660
B-1	Pholidota	Manidae	*Manis*	*Manis javanica*	Malayan pangolin	100	Captive	China	feces	Male	Adult	Ma et al 2018 FrontiersMicrobiol	PRJNA476660
B-2	Pholidota	Manidae	*Manis*	*Manis javanica*	Malayan pangolin	100	Captive	China	feces	Female	Adult	Ma et al 2018 FrontiersMicrobiol	PRJNA476660

Table S II.1. Detailed information of the 329 mammalian samples used for the comparative study of gut metagenomes.

Sample name	Order	Family	Genus	Species	Host common name	Diet percentage of invertebrates	Captive/wild	Country of origin	Sample type	Sex	Age	Study	Study accession
Gp0404153	Rodentia	Castoridae	*Castor*	*Castor canadensis*	American beaver	0	Wild	Canada	gut	NA	NA	NA	PRJNA620694
Gp0404155	Rodentia	Castoridae	*Castor*	*Castor canadensis*	American beaver	0	Wild	Canada	gut	NA	NA	NA	PRJNA620696
Gp0404158	Rodentia	Castoridae	*Castor*	*Castor canadensis*	American beaver	0	Wild	Canada	gut	NA	NA	NA	PRJNA620698
Gp0404163	Rodentia	Castoridae	*Castor*	*Castor canadensis*	American beaver	0	Wild	Canada	gut	NA	NA	NA	PRJNA620703
Gp0404165	Rodentia	Castoridae	*Castor*	*Castor canadensis*	American beaver	0	Wild	Canada	gut	NA	NA	NA	PRJNA620705
Gp0404168	Rodentia	Castoridae	*Castor*	*Castor canadensis*	American beaver	0	Wild	Canada	gut	NA	NA	NA	PRJNA620708
Gp0404170	Rodentia	Castoridae	*Castor*	*Castor canadensis*	American beaver	0	Wild	Canada	gut	NA	NA	NA	PRJNA620710
Gp0404173	Rodentia	Castoridae	*Castor*	*Castor canadensis*	American beaver	0	Wild	Canada	gut	NA	NA	NA	PRJNA620713
Gp0404175	Rodentia	Castoridae	*Castor*	*Castor canadensis*	American beaver	0	Wild	Canada	gut	NA	NA	NA	PRJNA620715
Gp0404178	Rodentia	Castoridae	*Castor*	*Castor canadensis*	American beaver	0	Wild	Canada	gut	NA	NA	NA	PRJNA620718
M03	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Male	Adult	No study | Baboon feces metagenome	PRJNA271618
M05	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Male	Adult	No study | Baboon feces metagenome	PRJNA271618
F06	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Female	Adult	No study | Baboon feces metagenome	PRJNA271618
M06	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Male	Adult	No study | Baboon feces metagenome	PRJNA271618
M07	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Male	Adult	No study | Baboon feces metagenome	PRJNA271618
F11	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Female	Adult	No study | Baboon feces metagenome	PRJNA271618
M10	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Male	Adult	No study | Baboon feces metagenome	PRJNA271618
F17	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Female	Adult	No study | Baboon feces metagenome	PRJNA271618
M14	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Male	Adult	No study | Baboon feces metagenome	PRJNA271618
F19	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Female	Adult	No study | Baboon feces metagenome	PRJNA271618
F20	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Female	Adult	No study | Baboon feces metagenome	PRJNA271618
F21	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Female	Adult	No study | Baboon feces metagenome	PRJNA271618
F22	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Female	Adult	No study | Baboon feces metagenome	PRJNA271618
F23	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Female	Adult	No study | Baboon feces metagenome	PRJNA271618
F24	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Female	Adult	No study | Baboon feces metagenome	PRJNA271618
M17	Primates	Cercopithecidae	*Papio*	*Papio cynocephalus*	Yellow baboon	10	Wild	Kenya	feces	Male	Adult	No study | Baboon feces metagenome	PRJNA271618
C10	Primates	Cercopithecidae	*Rhinopithecus*	*Rhinopithecus roxellana*	Golden snub-nosed monkey	0	Captive	China	feces	NA	NA	No study | monkey metagenome	PRJNA436633
C2	Primates	Cercopithecidae	*Rhinopithecus*	*Rhinopithecus roxellana*	Golden snub-nosed monkey	0	Captive	China	feces	NA	NA	No study | monkey metagenome	PRJNA436633
W12	Primates	Cercopithecidae	*Rhinopithecus*	*Rhinopithecus roxellana*	Golden snub-nosed monkey	0	Wild	China	feces	NA	NA	No study | monkey metagenome	PRJNA436633
W10	Primates	Cercopithecidae	*Rhinopithecus*	*Rhinopithecus roxellana*	Golden snub-nosed monkey	0	Wild	China	feces	NA	NA	No study | monkey metagenome	PRJNA436633
P7	Primates	Cercopithecidae	*Rhinopithecus*	*Rhinopithecus roxellana*	Golden snub-nosed monkey	0	Wild	China	feces	NA	NA	No study | monkey metagenome	PRJNA436633
P5	Primates	Cercopithecidae	*Rhinopithecus*	*Rhinopithecus roxellana*	Golden snub-nosed monkey	0	Wild	China	feces	NA	NA	No study | monkey metagenome	PRJNA436633
C12	Primates	Cercopithecidae	*Rhinopithecus*	*Rhinopithecus roxellana*	Golden snub-nosed monkey	0	Captive	China	feces	NA	NA	No study | monkey metagenome	PRJNA436633
P11	Primates	Cercopithecidae	*Rhinopithecus*	*Rhinopithecus roxellana*	Golden snub-nosed monkey	0	Wild	China	feces	NA	NA	No study | monkey metagenome	PRJNA436633
P10	Primates	Cercopithecidae	*Rhinopithecus*	*Rhinopithecus roxellana*	Golden snub-nosed monkey	0	Wild	China	feces	NA	NA	No study | monkey metagenome	PRJNA436633
W2	Primates	Cercopithecidae	*Rhinopithecus*	*Rhinopithecus roxellana*	Golden snub-nosed monkey	0	Wild	China	feces	NA	NA	No study | monkey metagenome	PRJNA436633
N30110	Primates	Hominidae	*Pan*	*Pan troglodytes schweinfurthii*	Chimpanzee	10	Wild	Uganda	feces	Female	NA	No study | Pan exome project	PRJNA505752
N21602	Primates	Hominidae	*Pan*	*Pan troglodytes schweinfurthii*	Chimpanzee	10	Wild	Uganda	feces	Male	NA	No study | Pan exome project	PRJNA505752
N21608	Primates	Hominidae	*Pan*	*Pan troglodytes schweinfurthii*	Chimpanzee	10	Wild	Uganda	feces	Male	NA	No study | Pan exome project	PRJNA505752
N32207	Primates	Hominidae	*Pan*	*Pan troglodytes schweinfurthii*	Chimpanzee	10	Wild	Uganda	feces	Male	NA	No study | Pan exome project	PRJNA505752
N44907	Primates	Hominidae	*Pan*	*Pan troglodytes schweinfurthii*	Chimpanzee	10	Wild	Uganda	feces	Female	NA	No study | Pan exome project	PRJNA505752
N35006	Primates	Hominidae	*Pan*	*Pan troglodytes schweinfurthii*	Chimpanzee	10	Wild	Uganda	feces	Male	NA	No study | Pan exome project	PRJNA505752
N17311	Primates	Hominidae	*Pan*	*Pan troglodytes schweinfurthii*	Chimpanzee	10	Wild	Uganda	feces	Female	NA	No study | Pan exome project	PRJNA505752
N41207	Primates	Hominidae	*Pan*	*Pan troglodytes schweinfurthii*	Chimpanzee	10	Wild	Uganda	feces	Male	NA	No study | Pan exome project	PRJNA505752
Cebus_capucinus_imitator_PRJNA485217_SSR096	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR098	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR091	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR093	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR114	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR116	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR102	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR105	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR118	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR013	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR016	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR007	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR012	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR046	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR047	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR024	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR039	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR049	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR056	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
Cebus_capucinus_imitator_PRJNA485217_SSR056	Primates	Cebidae	*Cebus*	*Cebus imitator*	Panamanian white-faced capuchin	20	Wild	Costa_Rica	feces	NA	NA	Orkin et al 2019 ISME J	PRJNA485217
RightWhale.F12	Artiodactyla	Balaenidae	*Eubalaena*	*Eubalaena glacialis*	North Atlantic right whale	100	Wild	Canada	feces	NA	Adult	Sanders et al 2015 NatCom	PRJEB37953
RightWhale.F5	Artiodactyla	Balaenidae	*Eubalaena*	*Eubalaena glacialis*	North Atlantic right whale	100	Wild	Canada	feces	NA	Adult	Sanders et al 2015 NatCom	PRJEB37953
RightWhale.F8	Artiodactyla	Balaenidae	*Eubalaena*	*Eubalaena glacialis*	North Atlantic right whale	100	Wild	Canada	feces	NA	Adult	Sanders et al 2015 NatCom	PRJEB37953
RightWhale.F16	Artiodactyla	Balaenidae	*Eubalaena*	*Eubalaena glacialis*	North Atlantic right whale	100	Wild	Canada	feces	NA	Adult	Sanders et al 2015 NatCom	PRJEB37953
2011_85	Primates	Hominidae	*Gorilla*	*Gorilla gorilla*	Gorilla	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116
2011_79	Primates	Hominidae	*Gorilla*	*Gorilla gorilla*	Gorilla	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116
2009_109	Primates	Hominidae	*Gorilla*	*Gorilla gorilla*	Gorilla	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116
2011_66	Primates	Hominidae	*Gorilla*	*Gorilla gorilla*	Gorilla	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116
2011_56	Primates	Hominidae	*Gorilla*	*Gorilla gorilla*	Gorilla	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116
2011_149	Primates	Hominidae	*Gorilla*	*Gorilla gorilla*	Gorilla	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116
2011_119	Primates	Hominidae	*Gorilla*	*Gorilla gorilla*	Gorilla	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116
2009_265	Primates	Hominidae	*Gorilla*	*Gorilla gorilla*	Gorilla	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116
2009_107	Primates	Hominidae	*Gorilla*	*Gorilla gorilla*	Gorilla	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116
2009_193	Primates	Hominidae	*Gorilla*	*Gorilla gorilla*	Gorilla	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116
HU539_S13	Primates	Hominidae	*Homo*	*Homo sapiens*	Human	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116
HU537_S11	Primates	Hominidae	*Homo*	*Homo sapiens*	Human	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116

Table S II.1. Detailed information of the 329 mammalian samples used for the comparative study of gut metagenomes.

Sample name	Order	Family	Genus	Species	Host common name	Diet percentage of invertebrates	Captive/wild	Country of origin	Sample type	Sex	Age	Study	Study accession	
HU508_S26	Primates	Hominidae	*Homo*	*Homo sapiens*	Human	0	Wild	Central_African_Republic	feces	NA	NA	Sharma et al 2020 mSystems	PRJNA635116	
CAB_M2809	Cingulata	Dasypodidae	*Cabassous*	*Cabassous unicinctus*	Southern naked-tailed armadillo	100	Wild	French_Guiana	feces	Male	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
CAB_M2962	Cingulata	Dasypodidae	*Cabassous*	*Cabassous unicinctus*	Southern naked-tailed armadillo	100	Wild	French_Guiana	feces	Male	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
CAB_M3141	Cingulata	Dasypodidae	*Cabassous*	*Cabassous unicinctus*	Southern naked-tailed armadillo	100	Wild	French_Guiana	feces	Female	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
DASY_M1777	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	100	Wild	French_Guiana	feces	Male	Young	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
DASY_M2865	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	100	Wild	French_Guiana	feces	Female	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
DASY_M3021	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	100	Wild	French_Guiana	feces	Male	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
KAP_M2255	Cingulata	Dasypodidae	*Dasypus*	*Dasypus kappleri*	Greater long-nosed armadillo	100	Wild	French_Guiana	feces	NA	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
KAP_M2867	Cingulata	Dasypodidae	*Dasypus*	*Dasypus kappleri*	Greater long-nosed armadillo	100	Wild	French_Guiana	feces	Male	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
DASY_VLD165	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus*	Nine-banded armadillo	100	Wild	USA	feces	Female	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
DASY_VLD172	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus*	Nine-banded armadillo	100	Wild	USA	feces	Female	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
MYR_M5293	Pilosa	Myrmecophagidae	*Myrmecophaga*	*Myrmecophaga tridactyla*	Giant anteater	100	Wild	French_Guiana	feces	Female	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
MYR_M5295	Pilosa	Myrmecophagidae	*Myrmecophaga*	*Myrmecophaga tridactyla*	Giant anteater	100	Wild	French_Guiana	feces	NA	Juvenile	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
MYR_ZOO	Pilosa	Myrmecophagidae	*Myrmecophaga*	*Myrmecophaga tridactyla*	Giant anteater	100	Captive	France	feces	Female	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
ORY_TS217	Tubulidentata	Orycteropodidae	*Orycteropus*	*Orycteropus afer*	Aardvark	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
ORY_TS513	Tubulidentata	Orycteropodidae	*Orycteropus*	*Orycteropus afer*	Aardvark	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
ORY_TS566	Tubulidentata	Orycteropodidae	*Orycteropus*	*Orycteropus afer*	Aardvark	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
PRO_TDR17	Carnivora	Hyaenidae	*Proteles*	*Proteles cristatus*	Southern aardwolf	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
PRO_TDR22	Carnivora	Hyaenidae	*Proteles*	*Proteles cristatus*	Southern aardwolf	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
PRO_TDR49	Carnivora	Hyaenidae	*Proteles*	*Proteles cristatus*	Southern aardwolf	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
PRO_TDR62	Carnivora	Hyaenidae	*Proteles*	*Proteles cristatus*	Southern aardwolf	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
PRO_TDR67	Carnivora	Hyaenidae	*Proteles*	*Proteles cristatus*	Southern aardwolf	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
PRO_TDR7	Carnivora	Hyaenidae	*Proteles*	*Proteles cristatus*	Southern aardwolf	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
PAN_TS471	Pholidota	Manidae	*Smutsia*	*Smutsia temminkii*	Ground pangolin	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
PAN_TS475	Pholidota	Manidae	*Smutsia*	*Smutsia temminkii*	Ground pangolin	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
PAN_TS482	Pholidota	Manidae	*Smutsia*	*Smutsia temminkii*	Ground pangolin	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
PAN_TS488	Pholidota	Manidae	*Smutsia*	*Smutsia temminkii*	Ground pangolin	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
PAN_TS489	Pholidota	Manidae	*Smutsia*	*Smutsia temminkii*	Ground pangolin	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
PAN_TS525	Pholidota	Manidae	*Smutsia*	*Smutsia temminkii*	Ground pangolin	100	Wild	South_Africa	feces	NA	NA	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
TAM_M3075	Pilosa	Myrmecophagidae	*Tamandua*	*Tamandua tetradactyla*	Southern tamandua	100	Wild	French_Guiana	feces	Male	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
TAM_M5331	Pilosa	Myrmecophagidae	*Tamandua*	*Tamandua tetradactyla*	Southern tamandua	100	Wild	French_Guiana	feces	Male	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
TAM_M5584	Pilosa	Myrmecophagidae	*Tamandua*	*Tamandua tetradactyla*	Southern tamandua	100	Wild	French_Guiana	feces	Female	Adult	Teullet et al 2023 mSystems	PRJNA942254	ConvergeAnt project
208	Chiroptera	Phyllostomidae	*Artibeus*	*Artibeus jamaicensis*	Jamaican fruit bat	10	NA	Guadeloupe	feces	NA	NA	This study	NA	G. Borrel
AA43	Artiodactyla	Cervidae	*Axis*	*Axis axis*	Chital	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
BP33	Diprotodontia	Potoroidae	*Bettongia*	*Bettongia penicillata*	Woylie	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
BRAtri_M5146	Pilosa	Bradypodidae	*Bradypus*	*Bradypus tridactylus*	Pale-throated sloth	0	Wild	French_Guiana	feces	NA	Juvenile	This study	NA	ConvergeAnt project
BRAtri_M5147	Pilosa	Bradypodidae	*Bradypus*	*Bradypus tridactylus*	Pale-throated sloth	0	Wild	French_Guiana	feces	Female	Juvenile	This study	NA	ConvergeAnt project
CD44	Artiodactyla	Camelidae	*Camelus*	*Camelus dromedarius*	Camel	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
CANmes_TS226	Carnivora	Canidae	*Canis*	*Canis mesomelas*	Black-backed jackal	10	Wild	South_Africa	feces	NA	NA	This study	NA	ConvergeAnt project
CANmes_TS227	Carnivora	Canidae	*Canis*	*Canis mesomelas*	Black-backed jackal	10	Wild	South_Africa	feces	NA	NA	This study	NA	ConvergeAnt project
54	Artiodactyla	Tayassuidae	*Catagonus*	*Catagonus wagneri*	Chacoan tagua	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
CP29	Rodentia	Caviidae	*Cavia*	*Cavia porcellus*	Guinea pig	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
45	Perissodactyla	Rhinocerotidae	*Ceratotherium*	*Ceratotherium simum*	White Rinoceros	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
142b	Perissodactyla	Rhinocerotidae	*Ceratotherium*	*Ceratotherium simum*	White Rinoceros	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
204	Primates	Cercopithecidae	*Chlorocebus*	*Chlorocebus aethiops*	Grivet monkey	20	NA	Guadeloupe	feces	NA	NA	This study	NA	G. Borrel
CHOdid_M5164	Pilosa	Megalonychidae	*Choloepus*	*Choloepus didactylus*	Linnaeus two-toed sloth	0	Captive	French_Guiana	feces	Male	Adult	This study	NA	ConvergeAnt project
CUNpac_M5329	Rodentia	Agoutidae	*Cuniculus*	*Cuniculus paca*	Lowlands paca	0	Wild	French_Guiana	feces	Male	Juvenile	This study	NA	ConvergeAnt project
DASlep_M5324	Rodentia	Dasyproctidae	*Dasyprocta*	*Dasyprocta leporina*	Red-rumped agouti	10	Captive	French_Guiana	feces	NA	Adult	This study	NA	ConvergeAnt project
DASY_M2976	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	100	Wild	French_Guiana	feces	Male	Adult	This study	NA	ConvergeAnt project
DASY_M3042	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	100	Wild	French_Guiana	feces	Female	Adult	This study	NA	ConvergeAnt project
DASY_M3068	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	100	Wild	French_Guiana	feces	Female	Adult	This study	NA	ConvergeAnt project
DASY_M3070	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	100	Wild	French_Guiana	feces	Male	Adult	This study	NA	ConvergeAnt project
DASY_M3126	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	100	Wild	French_Guiana	feces	Female	Adult	This study	NA	ConvergeAnt project
KAP_M3109	Cingulata	Dasypodidae	*Dasypus*	*Dasypus kappleri*	Greater long-nosed armadillo	100	Wild	French_Guiana	feces	Male	Adult	This study	NA	ConvergeAnt project
KAP_M3462	Cingulata	Dasypodidae	*Dasypus*	*Dasypus kappleri*	Greater long-nosed armadillo	100	Wild	French_Guiana	feces	Male	Adult	This study	NA	ConvergeAnt project
KAP_M3474	Cingulata	Dasypodidae	*Dasypus*	*Dasypus kappleri*	Greater long-nosed armadillo	100	Wild	French_Guiana	feces	Male	Adult	This study	NA	ConvergeAnt project
KAP_M3477	Cingulata	Dasypodidae	*Dasypus*	*Dasypus kappleri*	Greater long-nosed armadillo	100	Wild	French_Guiana	feces	Female	Adult	This study	NA	ConvergeAnt project
DASY_VLD161	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus*	Nine-banded armadillo	100	Wild	USA	feces	Male	Adult	This study	NA	ConvergeAnt project
DASY_VLD166	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus*	Nine-banded armadillo	100	Wild	USA	feces	Male	Adult	This study	NA	ConvergeAnt project
DASY_VLD170	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus*	Nine-banded armadillo	100	Wild	USA	feces	Female	Adult	This study	NA	ConvergeAnt project
DASY_VLD174	Cingulata	Dasypodidae	*Dasypus*	*Dasypus novemcinctus*	Nine-banded armadillo	100	Wild	USA	feces	Male	Adult	This study	NA	ConvergeAnt project
DM36	Diprotodontia	Macropodidae	*Dendrolagus*	*Dendrolagus matschiei*	Matschie's tree-kangaroo	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
36	Rodentia	Caviidae	*Dolichotis*	*Dolichotis patagonum*	Patagonian Cavy	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
268	Perissodactyla	Equidae	*Equus*	*Equus przewalskii*	Przewalski's horse	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
176	Perissodactyla	Equidae	*Equus*	*Equus asinus*	Donkey	0	NA	Guadeloupe	feces	NA	NA	This study	NA	G. Borrel
EC50	Perissodactyla	Equidae	*Equus*	*Equus caballus*	Horse	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
43	Perissodactyla	Equidae	*Equus*	*Equus grevyi*	Grevy's Zebra	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
141b	Perissodactyla	Equidae	*Equus*	*Equus grevyi*	Grevy's Zebra	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
145	Primates	Lemuridae	*Eulemur*	*Eulemur rubriventer*	Red-bellied Lemur	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
CT04	Cingulata	Chlamyphoridae	*Euphractus*	*Euphractus sexcinctus*	Six-banded armadillo	50	NA	France	feces	NA	NA	This study	NA	G. Borrel
GALvit_M178	Carnivora	Mustelidae	*Galictis*	*Galictis vittata*	Greater grison	20	Wild	French_Guiana	feces	Female	Adult	This study	NA	ConvergeAnt project
85	Artiodactyla	Giraffidae	*Giraffa*	*Giraffa camelopardalis*	Giraffe	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
GG26	Primates	Hominidae	*Gorilla*	*Gorilla gorilla*	Gorilla	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
146	Primates	Lemuridae	*Hapalemur*	*Hapalemur aureus*	Golden bamboo lemur	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
146b	Primates	Lemuridae	*Hapalemur*	*Hapalemur aureus*	Golden bamboo lemur	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
HI30	Rodentia	Hystricidae	*Hystrix*	*Hystrix indica*	Indian crested porcupine	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
149	Primates	Atelidae	*Lagothrix*	*Lagothrix poeppigii*	Silvery woolly monkey	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
144	Primates	Lemuridae	*Lemur*	*Lemur catta*	Ring-tailed Lemur	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
144b	Primates	Lemuridae	*Lemur*	*Lemur catta*	Ring-tailed Lemur	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
LA18	Proboscidea	Elephantidae	*Loxodonta*	*Loxodonta africana*	African elephant	0	NA	France	feces	NA	NA	This study	NA	G. Borrel

Table S II.1. Detailed information of the 329 mammalian samples used for the comparative study of gut metagenomes.

Sample name	Order	Family	Genus	Species	Host common name	Diet percentage of invertebrates	Captive/wild	Country of origin	Sample type	Sex	Age	Study	Study accession	
Macaca61	Primates	Cercopithecidae	*Macaca*	*Macaca fascicularis*	Crab-eating macaque	50	NA	France	feces	NA	NA	This study	NA	G. Borrel
Macaca63	Primates	Cercopithecidae	*Macaca*	*Macaca fascicularis*	Crab-eating macaque	50	NA	France	feces	NA	NA	This study	NA	G. Borrel
Macaca66	Primates	Cercopithecidae	*Macaca*	*Macaca fascicularis*	Crab-eating macaque	50	NA	France	feces	NA	NA	This study	NA	G. Borrel
Macaca7s	Primates	Cercopithecidae	*Macaca*	*Macaca fascicularis*	Crab-eating macaque	50	NA	France	feces	NA	NA	This study	NA	G. Borrel
MP15	Macroscelidea	Macroscelididae	*Macroscelides*	*Macroscelides proboscideus*	Round-eared elephant shrew	80	NA	France	feces	NA	NA	This study	NA	G. Borrel
MAKdid_M1153	Rodentia	Echimyidae	*Makalata*	*Makalata didelphoides*	Red-nosed armored tree-rat	0	Wild	French_Guiana	feces	Female	NA	This study	NA	ConvergeAnt project
MEShis_M3108	Rodentia	Echimyidae	*Mesomys*	*Mesomys hispidus*	Ferreira's spiny tree rat	0	Wild	French_Guiana	feces	Male	Adult	This study	NA	ConvergeAnt project
CT05	Pilosa	Myrmecophagidae	*Myrmecophaga*	*Myrmecophaga tridactyla*	Giant anteater	100	NA	France	feces	NA	NA	This study	NA	G. Borrel
177	Lagomorpha	Leporidae	*Oryctolagus*	*Orcytolagus cuniculus*	Rabbit	0	NA	Guadeloupe	feces	NA	NA	This study	NA	G. Borrel
ORY_TS219	Tubulidentata	Orycteropodidae	*Orycteropus*	*Orycteropus afer*	Aardvark	100	Wild	South_Africa	feces	NA	NA	This study	NA	ConvergeAnt project
ORY_TS278	Tubulidentata	Orycteropodidae	*Orycteropus*	*Orycteropus afer*	Aardvark	100	Wild	South_Africa	feces	NA	NA	This study	NA	ConvergeAnt project
ORY_TS280	Tubulidentata	Orycteropodidae	*Orycteropus*	*Orycteropus afer*	Aardvark	100	Wild	South_Africa	feces	NA	NA	This study	NA	ConvergeAnt project
ORY_TS526	Tubulidentata	Orycteropodidae	*Orycteropus*	*Orycteropus afer*	Aardvark	100	Wild	South_Africa	feces	NA	NA	This study	NA	ConvergeAnt project
PT21	Primates	Hominidae	*Pan*	*Pan troglodytes*	Chimpanzee	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
55	Carnivora	Felidae	*Panthera*	*Panthera onca*	Jaguar	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
52	Primates	Pitheciidae	*Pithecia*	*Pithecia pithecia*	White-faced saki	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
203	Primates	Pitheciidae	*Pithecia*	*Pithecia pithecia*	White-faced saki	0	NA	Guadeloupe	feces	NA	NA	This study	NA	G. Borrel
269	Primates	Hominidae	*Pongo*	*Pongo pygmaeus*	Bornean orangutan	10	NA	France	feces	NA	NA	This study	NA	G. Borrel
PP22	Primates	Hominidae	*Pongo*	*Pongo pygmaeus*	Bornean orangutan	10	NA	France	feces	NA	NA	This study	NA	G. Borrel
POTfla_M1749	Carnivora	Procyonidae	*Potos*	*Potos flavus*	Kinkajou	10	Wild	French_Guiana	feces	Male	Adult	This study	NA	ConvergeAnt project
POTfla_M2756	Carnivora	Procyonidae	*Potos*	*Potos flavus*	Kinkajou	10	Wild	French_Guiana	feces	Female	NA	This study	NA	ConvergeAnt project
CT14	Hyracoidea	Procaviidae	*Procavia*	*Procavia capensis*	Cape rock hyrax	0	NA	South_Africa	feces	NA	NA	This study	NA	G. Borrel
PROguy_M1199	Rodentia	Echimyidae	*Proechimys*	*Proechimys guyannensis*	Guyenne spiny rat	0	Wild	French_Guiana	feces	Male	NA	This study	NA	ConvergeAnt project
PROguy_M1209	Rodentia	Echimyidae	*Proechimys*	*Proechimys guyannensis*	Guyenne spiny rat	0	Wild	French_Guiana	feces	Female	NA	This study	NA	ConvergeAnt project
84	Primates	Indriidae	*Propithecus*	*Propithecus coronatus*	Crowned Sifaka	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
140	Artiodactyla	Cervidae	*Pudu*	*Pudu puda*	Pudu	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
PUMyag_M1572	Carnivora	Felidae	*Puma*	*Puma yagouaroundi*	Jaguarundi	10	Wild	French_Guiana	feces	Male	NA	This study	NA	ConvergeAnt project
RU52	Perissodactyla	Rhinocerotidae	*Rhinoceros*	*Rhinoceros unicornis*	Greater one-horned rhinoceros	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
SE5	Primates	Cercopithecidae	*Semnopithecus*	*Semnopithecus entellus*	Northern plains gray langur	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
PAN_TS352	Pholidota	Manidae	*Smutsia*	*Smutsia temminkii*	Ground pangolin	100	Wild	South_Africa	feces	NA	NA	This study	NA	ConvergeAnt project
SURsur_TS222	Carnivora	Herpestidae	*Suricata*	*Suricata suricatta*	Meerkat	70	Wild	South_Africa	feces	NA	NA	This study	NA	ConvergeAnt project
SURsur_TS223	Carnivora	Herpestidae	*Suricata*	*Suricata suricatta*	Meerkat	70	Wild	South_Africa	feces	NA	NA	This study	NA	ConvergeAnt project
SURsur_TS224	Carnivora	Herpestidae	*Suricata*	*Suricata suricatta*	Meerkat	70	Wild	South_Africa	feces	NA	NA	This study	NA	ConvergeAnt project
274	Carnivora	Herpestidae	*Suricata*	*Suricata suricatta*	Meerkat	70	NA	France	feces	NA	NA	This study	NA	G. Borrel
172	Artiodactyla	Suidae	*Sus*	*Sus scrofa domesticus*	Pig	10	NA	Guadeloupe	feces	NA	NA	This study	NA	G. Borrel
TAM_M3074	Pilosa	Myrmecophagidae	*Tamandua*	*Tamandua tetradactyla*	Southern tamandua	100	Wild	French_Guiana	feces	Female	Adult	This study	NA	ConvergeAnt project
TI49	Perissodactyla	Tapiridae	*Tapirus*	*Tapirus indicus*	Malayan tapir	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
48	Perissodactyla	Tapiridae	*Tapirus*	*Tapirus terrestris*	Brazilian Tapir	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
143	Perissodactyla	Tapiridae	*Tapirus*	*Tapirus terrestris*	South American Tapir	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
138	Sirenia	Trichechidae	*Trichechus*	*Trichechus manatus*	American manatee	0	NA	France	feces	NA	NA	This study	NA	G. Borrel
Pig08	Artiodactyla	Suidae	*Sus*	*Sus scrofa domesticus*	Pig	10	Captive	NA	feces	NA	NA	Xiao et al 2016 NatMicrobio	PRJEB11755	
Pig21	Artiodactyla	Suidae	*Sus*	*Sus scrofa domesticus*	Pig	10	Captive	NA	feces	NA	NA	Xiao et al 2016 NatMicrobio	PRJEB11755	
Pig23	Artiodactyla	Suidae	*Sus*	*Sus scrofa domesticus*	Pig	10	Captive	NA	feces	NA	NA	Xiao et al 2016 NatMicrobio	PRJEB11755	
Pig88	Artiodactyla	Suidae	*Sus*	*Sus scrofa domesticus*	Pig	10	Captive	NA	feces	NA	NA	Xiao et al 2016 NatMicrobio	PRJEB11755	
Macaca mulatta_control_120720	Primates	Cercopithecidae	*Macaca*	*Macaca mulatta*	Rhesus monkey	10	Captive	USA	feces	Female	Young	Yan et al 2021 npj Biofilms Microbiomes	PRJNA574851	
Macaca mulatta_control_120718	Primates	Cercopithecidae	*Macaca*	*Macaca mulatta*	Rhesus monkey	10	Captive	USA	feces	Female	Young	Yan et al 2021 npj Biofilms Microbiomes	PRJNA574851	
Macaca mulatta_control_120714	Primates	Cercopithecidae	*Macaca*	*Macaca mulatta*	Rhesus monkey	10	Captive	USA	feces	Female	Young	Yan et al 2021 npj Biofilms Microbiomes	PRJNA574851	
Macaca mulatta_control_120708	Primates	Cercopithecidae	*Macaca*	*Macaca mulatta*	Rhesus monkey	10	Captive	USA	feces	Female	Young	Yan et al 2021 npj Biofilms Microbiomes	PRJNA574851	
Macaca mulatta_control_110807	Primates	Cercopithecidae	*Macaca*	*Macaca mulatta*	Rhesus monkey	10	Captive	USA	feces	Male	Young	Yan et al 2021 npj Biofilms Microbiomes	PRJNA574851	
Macaca mulatta_A53T_130996	Primates	Cercopithecidae	*Macaca*	*Macaca mulatta*	Rhesus monkey	10	Captive	USA	feces	Female	Young	Yan et al 2021 npj Biofilms Microbiomes	PRJNA574851	
Macaca mulatta_A53T_130919	Primates	Cercopithecidae	*Macaca*	*Macaca mulatta*	Rhesus monkey	10	Captive	USA	feces	Male	Young	Yan et al 2021 npj Biofilms Microbiomes	PRJNA574851	
Macaca mulatta_A53T_130916	Primates	Cercopithecidae	*Macaca*	*Macaca mulatta*	Rhesus monkey	10	Captive	USA	feces	Female	Young	Yan et al 2021 npj Biofilms Microbiomes	PRJNA574851	
Macaca mulatta_A53T_120666	Primates	Cercopithecidae	*Macaca*	*Macaca mulatta*	Rhesus monkey	10	Captive	USA	feces	Female	Young	Yan et al 2021 npj Biofilms Microbiomes	PRJNA574851	
X409_Impala	Artiodactyla	Bovidae	*Aepyceros*	*Aepyceros melampus*	Impala	0	Wild	South_Africa	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X410_Impala	Artiodactyla	Bovidae	*Aepyceros*	*Aepyceros melampus*	Impala	0	Wild	South_Africa	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X21_Yellow_necked_Field_Mouse	Rodentia	Muridae	*Apodemus*	*Apodemus flavicollis*	Yellow necked Field Mouse	20	Wild	Austria	gut	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X248_Yellow_necked_Field_Mouse	Rodentia	Muridae	*Apodemus*	*Apodemus flavicollis*	Yellow necked Field Mouse	20	Wild	Austria	gut	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X155_Long_tailed_Field_Mouse	Rodentia	Muridae	*Apodemus*	*Apodemus sylvaticus*	Long tailed Field Mouse	20	Wild	Austria	gut	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X137_Gaur	Artiodactyla	Bovidae	*Bos*	*Bos frontalis*	Gaur	0	Wild	Viet_Nam	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X138_Gaur	Artiodactyla	Bovidae	*Bos*	*Bos frontalis*	Gaur	0	Wild	Viet_Nam	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X221_Alpine_Ibex	Artiodactyla	Bovidae	*Capra*	*Capra ibex*	Alpine Ibex	0	Captive	Austria	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X129_Alpine_Ibex	Artiodactyla	Bovidae	*Capra*	*Capra ibex*	Alpine Ibex	0	Wild	Austria	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X130_Alpine_Ibex	Artiodactyla	Bovidae	*Capra*	*Capra ibex*	Alpine Ibex	0	Wild	Austria	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X371_Linnaeus_Two_toed_Sloth	Pilosa	Megalonychidae	*Choloepus*	*Choloepus didactylus*	Linnaeus Two toed sloth	0	Captive	Austria	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X355_Hoffmanns_Two_toed_Sloth	Pilosa	Megalonychidae	*Choloepus*	*Choloepus hoffmanni*	Hoffmanns Two toed sloth	0	Wild	NA	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X101_Horse	Perissodactyla	Equidae	*Equus*	*Equus caballus*	Horse	0	Captive	Austria	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X77_Onager	Perissodactyla	Equidae	*Equus*	*Equus hemionus*	Onager	0	Wild	Iran	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X394_Przewalski_horse	Perissodactyla	Equidae	*Equus*	*Equus przewalski*	Przewalski horse	0	Wild	Mongolia	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X424_Quagga	Perissodactyla	Equidae	*Equus*	*Equus quagga burchellii*	Plains Zebra	0	Wild	East_Africa	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X429_Giraffe	Artiodactyla	Giraffidae	*Giraffa*	*Giraffa camelopardalis*	Giraffe	0	Captive	Austria	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X430_Giraffe	Artiodactyla	Giraffidae	*Giraffa*	*Giraffa camelopardalis*	Giraffe	0	Captive	Austria	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X417_Giraffe	Artiodactyla	Giraffidae	*Giraffa*	*Giraffa camelopardalis*	Giraffe	0	Wild	East_Africa	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X16_Fat_Dormouse	Rodentia	Gliridae	*Glis*	*Glis glis*	Fat Dormouse	0	Captive	Austria	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X363_Desmarests_Spiny_Pocket_Mouse	Rodentia	Heteromyidae	*Heteromys*	*Heteromys desmarestianus*	Desmarests Spiny Pocket Mouse	0	Wild	Costa_Rica	gut	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X339_White_beaked_Dolphin	Artiodactyla	Delphinidae	*Lagenorhynchus*	*Lagenorhynchus albirostris*	White beaked Dolphin	40	Wild	England	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X301_European_Hare	Lagomorpha	Leporidae	*Lepus*	*Lepus europaeus*	European Hare	0	Wild	Austria	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X10_House_Mouse	Rodentia	Muridae	*Mus*	*Mus musculus*	House Mouse	20	Captive	Austria	gut	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X107_House_Mouse	Rodentia	Muridae	*Mus*	*Mus musculus*	House Mouse	20	Captive	Austria	gut	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X403_Chacma_Baboon	Primates	Cercopithecidae	*Papio*	*Papio ursinus*	Chacma Baboon	20	Wild	South_Africa	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	
X108_Striped_Desert_Hamster	Rodentia	Cricetidae	*Phodopus*	*Phodopus sungorus*	Striped Desert Hamster	0	Captive	Austria	gut	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078	

Table S II.1. Detailed information of the 329 mammalian samples used for the comparative study of gut metagenomes.

Sample name	Order	Family	Genus	Species	Host common name	Diet percentage of invertebrates	Captive/wild	Country of origin	Sample type	Sex	Age	Study	Study accession
X344_Indian_Rhinoceros	Perissodactyla	Rhinocerotidae	*Rhinoceros*	*Rhinoceros unicornis*	Indian Rhinoceros	0	Wild	Nepal	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078
X285_Short_beaked_Echidna	Monotremata	Tachyglossidae	*Tachyglossus*	*Tachyglossus aculeatus*	Short beaked Echidna	100	Wild	Australia	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078
X286_Short_beaked_Echidna	Monotremata	Tachyglossidae	*Tachyglossus*	*Tachyglossus aculeatus*	Short beaked Echidna	100	Wild	Australia	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078
X238_Hanuman_Langur	Primates	Cercopithecidae	*Trachypithecus*	*Trachypithecus hatinhensis*	Hanuman Langur	0	Wild	Viet_Nam	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078
X347_Brown_Bear	Carnivora	Ursidae	*Ursus*	*Ursus arctos*	Brown Bear	10	Captive	Austria	feces	NA	NA	Youngblut et al 2020 mSystems	PRJEB38078
Leopard Seal - Albino Rookery (LAR)	Carnivora	Phocidae	*Hydrurga*	*Hydrurga leptonyx*	Leopard seal	80	Wild	Antarctica	feces	NA	NA	NA	PRJNA481090
Weddell seal - Sion Island (WEDSI)	Carnivora	Phocidae	*Leptonychotes*	*Leptonychotes weddellii*	Weddell seal	40	Wild	Antarctica	feces	NA	NA	NA	PRJNA481090
CRM1B008FFO121	Rodentia	Muridae	*Mus*	*Mus musculus*	House Mouse	20	NA	China	feces	NA	NA	NA	PRJNA497852

Chapter references (excluding manuscripts' references)

Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., Rodriguez-Mueller, B., Zucker, J., Thiagarajan, M., Henrissat, B., White, O., Kelley, S.T., Methé, B., Schloss, P.D., Gevers, D., Mitreva, M., Huttenhower, C., 2012. Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comput Biol 8, e1002358.

Adrangi, S., Faramarzi, M.A., 2013. From bacteria to human: A journey into the world of chitinases. Biotechnology Advances 31, 1786–1795.

Alessandri, G., Rizzo, S.M., Ossiprandi, M.C., van Sinderen, D., Ventura, M., 2022. Creating an atlas to visualize the biodiversity of the mammalian gut microbiota. Current Opinion in Biotechnology 73, 28–33.

Allio, R., Teullet, S., Lutgen, D., Magdeleine, A., Koual, R., Tilak, M.-K., Thoisy, B. de, Emerling, C.A., Lefébure, T., Delsuc, F., 2023. Comparative transcriptomics reveals divergent paths of chitinase evolution underlying dietary convergence in ant-eating mammals. bioRxiv. https://doi.org/10.1101/2022.11.29.518312. In revision in Genome Biology and Evolution.

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., Quince, C., 2014. Binning metagenomic contigs by coverage and composition. Nature Methods 11, 1144–1146.

Amato, K., 2013. Co-evolution in context: The importance of studying gut microbiomes in wild animals. Microbiome Science and Medicine 1, 10–29.

Arumugam, K., Bessarab, I., Haryono, M.A.S., Liu, X., Zuniga–Montanez, R.E., Roy, S., Qiu, G., Drautz–Moses, D.I., Law, Y.Y., Wuertz, S., Lauro, F.M., Huson, D.H., Williams, R.B.H., 2021. Recovery of complete genomes and non-chromosomal replicons from activated sludge enrichment microbial communities with long read metagenome sequencing. npj Biofilms and Microbiomes 7, 1–13.

Asnicar, F., Thomas, A.M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M., Cumbo, F., May, U., Sanders, J.G., Zolfo, M., Kopylova, E., Pasolli, E., Knight, R., Mirarab, S., Huttenhower, C., Segata, N., 2020. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nature Communications 11, 2500.

Beier, S., Bertilsson, S., 2013. Bacterial chitin degradation—mechanisms and ecophysiological strategies. Front. Microbiol. 4, 149.

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C.C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G.H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J.A., Maguin, E., Mauchline, T., McClure, R., Mitter, B., Ryan, M., Sarand, I., Smidt, H., Schelkle, B., Roume, H., Kiran, G.S., Selvin, J., Souza, R.S.C. de, van Overbeek, L., Singh, B.K., Wagner, M., Walsh, A., Sessitsch, A., Schloter, M., 2020. Microbiome definition re-visited: old concepts and new challenges. Microbiome 8, 103.

Bharti, R., Grimm, D.G., 2021. Current challenges and best-practice protocols for microbiome analysis. Briefings in Bioinformatics 22, 178–193.

Bickhart, D.M., Kolmogorov, M., Tseng, E., Portik, D.M., Korobeynikov, A., Tolstoganov, I., Uritskiy, G., Liachko, I., Sullivan, S.T., Shin, S.B., Zorea, A., Andreu, V.P., Panke-Buisse, K., Medema, M.H., Mizrahi, I., Pevzner, P.A., Smith, T.P.L., 2022. Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. Nat Biotechnol 40, 711-719.

Blekhman, R., Tang, K., Archie, E.A., Barreiro, L.B., Johnson, Z.P., Wilson, M.E., Kohn, J., Yuan, M.L., Gesquiere, L., Grieneisen, L.E., Tung, J., 2016. Common methods for fecal sample storage in field studies yield consistent signatures of individual identity in microbiome sequencing data. Scientific Reports 6, 31519.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A.M., Chase, J., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibbons, S.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K.B., Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G.I., Lee, J., Ley, R., Liu, Y.-X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., McDonald, D., McIver, L.J., Melnik, A.V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., van der Hooft, J.J.J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C.H.D., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Zhu, Q., Knight, R., Caporaso, J.G., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37, 852–857.

Breitwieser, F.P., Lu, J., Salzberg, S.L., 2019. A review of methods and databases for metagenomic classification and assembly. Brief Bioinform 20, 1125–1136.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., Forterre, P., 2008. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nat Rev Microbiol 6, 245–252.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High resolution sample inference from Illumina amplicon data. Nat Methods 13, 581–583.

Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J 11, 2639–2643.

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.

Chen, Y.-H., Chiang, P.-W., Rogozin, D.Y., Degermendzhy, A.G., Chiu, H.-H., Tang, S.-L., 2021. Salvaging high-quality genomes of microbial species from a meromictic lake using a hybrid sequencing approach. Commun Biol 4, 996.

Cheng, S.-C., Liu, C.-B., Yao, X.-Q., Hu, J.-Y., Yin, T.-T., Lim, B.K., Chen, W., Wang, G.-D., Zhang, C.-L., Irwin, D.M., Zhang, Z.-G., Zhang, Y.-P., Yu, L., 2023. Hologenomic insights into mammalian adaptations to myrmecophagy. National Science Review 10, nwac174.

Chibani, C.M., Mahnert, A., Borrel, G., Almeida, A., Werner, A., Brugère, J.-F., Gribaldo, S., Finn, R.D., Schmitz, R.A., Moissl-Eichinger, C., 2021. A catalogue of 1,167 genomes from the human gut archaeome. Nat Microbiol 1–14.

Costea, P.I., Zeller, G., Sunagawa, S., Pelletier, E., Alberti, A., Levenez, F., Tramontano, M., Driessen, M., Hercog, R., Jung, F.-E., Kultima, J.R., Hayward, M.R., Coelho, L.P., Allen-Vercoe, E., Bertrand, L., Blaut, M., Brown, J.R.M., Carton, T., Cools-Portier, S., Daigneault, M., Derrien, M., Druesne, A., de Vos, W.M., Finlay, B.B., Flint, H.J., Guarner, F., Hattori, M., Heilig, H., Luna, R.A., van Hylckama Vlieg, J., Junick, J., Klymiuk, I., Langella, P., Le Chatelier, E., Mai, V., Manichanh, C., Martin, J.C., Mery, C., Morita, H., O'Toole, P.W., Orvain, C., Patil, K.R., Penders, J., Persson, S., Pons, N., Popova, M., Salonen, A., Saulnier, D., Scott, K.P., Singh, B., Slezak, K., Veiga, P., Versalovic, J., Zhao, L., Zoetendal, E.G., Ehrlich, S.D., Dore, J., Bork, P., 2017. Towards standards for human fecal sample processing in metagenomic studies. Nature Biotechnology 35, 1069–1076.

Cox, C.J., Foster, P.G., Hirt, R.P., Harris, S.R., Embley, T.M., 2008. The archaebacterial origin of eukaryotes. Proc Natl Acad Sci U S A 105, 20356–20361.

Cuscó, A., Pérez, D., Viñes, J., Fàbregas, N., Francino, O., 2021. Long-read metagenomics retrieves complete single-contig bacterial genomes from canine feces. BMC Genomics 22, 330.

Deeba, F., Shakir, H., Irfan, M., Qazi, J., 2016. Chitinase production in organisms: A review. Punjab University Journal of Zoology 31, 101–106.

Delsuc, F., Metcalf, J.L., Wegener Parfrey, L., Song, S.J., González, A., Knight, R., 2014. Convergence of gut microbiomes in myrmecophagous mammals. Molecular Ecology 23, 1301–1317.

Dilthey, A.T., Jain, C., Koren, S., Phillippy, A.M., 2019. Strain-level metagenomic assignment and compositional estimation for long reads with MetaMaps. Nat Commun 10, 3066.

Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., Terrapon, N., 2022. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Research 50, D571–D577.

El Kaoutari, A., Armougom, F., Gordon, J.I., Raoult, D., Henrissat, B., 2013. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Microbiol 11, 497–504.

El Knidri, H., Belaabed, R., Addaou, A., Laajeb, A., Lahsini, A., 2018. Extraction, chemical modification and characterization of chitin and chitosan. International Journal of Biological Macromolecules 120, 1181–1189.

Emerling, C.A., Delsuc, F., Nachman, M.W., 2018. Chitinase genes (*CHIAs*) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. Science Advances 4, eaar6478.

Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G., Sogin, M.L., 2013. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. Methods in Ecology and Evolution 4, 1111–1119.

Eren, A.M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S.E., Schechter, M.S., Fink, I., Pan, J.N., Yousef, M., Fogarty, E.C., Trigodet, F., Watson, A.R., Esen, Ö.C., Moore, R.M., Clayssen, Q., Lee, M.D., Kivenson, V., Graham, E.D., Merrill, B.D., Karkman, A., Blankenberg, D., Eppley, J.M., Sjödin, A., Scott, J.J., Vázquez-Campos, X., McKay, L.J., McDaniel, E.A., Stevens, S.L.R., Anderson, R.E., Fuessel, J., Fernandez-Guerra,

A., Maignien, L., Delmont, T.O., Willis, A.D., 2021. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 6, 3–6.

Gevers, D., Knight, R., Petrosino, J.F., Huang, K., McGuire, A.L., Birren, B.W., Nelson, K.E., White, O., Methé, B.A., Huttenhower, C., 2012. The Human Microbiome Project: a community resource for the healthy human microbiome. PLoS Biol 10, e1001377.

Ghurye, J.S., Cepeda-Espinoza, V., Pop, M., 2016. Metagenomic Assembly: Overview, Challenges and Applications. Yale J Biol Med 89, 353–362.

Gibson, K.M., Nguyen, B.N., Neumann, L.M., Miller, M., Buss, P., Daniels, S., Ahn, M.J., Crandall, K.A., Pukazhenthi, B., 2019. Gut microbiome differences between wild and captive black rhinoceros – implications for rhino health. Sci Rep 9, 7570.

Gilbert, J.A., Meyer, F., Jansson, J., Gordon, J., Pace, N., Tiedje, J., Ley, R., Fierer, N., Field, D., Kyrpides, N., Glöckner, F.-O., Klenk, H.-P., Wommack, K.E., Glass, E., Docherty, K., Gallery, R., Stevens, R., Knight, R., 2010. The Earth Microbiome Project: Meeting report of the "1st EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6th 2010. Stand in Genomic Sci 3, 249–253.

Gooday, G.W., 1990. The Ecology of Chitin Degradation, in: Marshall, K.C. (Ed.), Advances in Microbial Ecology, Advances in Microbial Ecology. Springer US, Boston, MA, pp. 387–430.

Goodrich, T.D., Morita, R.Y., 1977. Bacterial chitinase in the stomachs of marine fishes from Yaquina Bay, Oregon, USA. Mar. Biol. 41, 355–360.

Groussin, M., Mazel, F., Sanders, J.G., Smillie, C.S., Lavergne, S., Thuiller, W., Alm, E.J., 2017. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. Nature Communications 8, 14319.

Gruen, D.S., Wolfe, J.M., Fournier, G.P., 2019. Paleozoic diversification of terrestrial chitin-degrading bacterial lineages. BMC Evolutionary Biology 19, 34.

Guo, W., Mishra, S., Zhao, J., Tang, J., Zeng, B., Kong, F., Ning, R., Li, M., Zhang, Hengzhi, Zeng, Y., Tian, Y., Zhong, Y., Luo, H., Liu, Y., Yang, J., Yang, M., Zhang, M., Li, Yan, Ni, Q., Li, C., Wang, C., Li, D., Zhang, Hemin, Zuo, Z., Li, Ying, 2018. Metagenomic Study Suggests That the Gut Microbiota of the Giant Panda (*Ailuropoda melanoleuca*) May Not Be Specialized for Fiber Fermentation. Front. Microbiol. 9, 229.

Guy, L., Ettema, T.J.G., 2011. The archaeal "TACK" superphylum and the origin of eukaryotes. Trends Microbiol 19, 580–587.

Hamid, R., Khan, M.A., Ahmad, M., Ahmad, M.M., Abdin, M.Z., Musarrat, J., Javed, S., 2013. Chitinases: An update. J Pharm Bioallied Sci 5, 21–29.

Hird, S.M., 2017. Evolutionary Biology Needs Wild Microbiomes. Frontiers in Microbiology 8, 725.

Hirsch, A., 2004. Plant-microbe symbioses: A continuum from commensalism to parasitism. Symbiosis 37, 345–363.

Hu, Y., Fang, L., Nicholson, C., Wang, K., 2020. Implications of Error-Prone Long-Read Whole-Genome Shotgun Sequencing on Characterizing Reference Microbiomes. iScience 23, 101223.

Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119.

Irulan, A., Nathan, P.T., Priya, Y.S., Marimuthu, S., Elangovan, V., 2011. Isolation and characterization of chitinase producing gut microflora of insectivorous bats. Trends in Biosciences 4, 8–11.

Itoi, S., Okamura, T., Koyama, Y., Sugita, H., 2006. Chitinolytic bacteria in the intestinal tract of Japanese coastal fishes. Can. J. Microbiol. 52, 1158–1163.

Jeuniaux, Ch., 1950. Production D'une Exochitinase par des Bactéries Chitinolytiques Isolées du Contenu Intestinal de L'escargot. Archives Internationales de Physiologie 58, 352–353.

Kang, D.D., Froula, J., Egan, R., Wang, Z., 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3, e1165.

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z., 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7, e7359.

Kohl, K.D., 2020. Ecological and evolutionary mechanisms underlying patterns of phylosymbiosis in host-associated microbial communities. Philosophical Transactions of the Royal Society B: Biological Sciences 375, 20190251.

Kohl, K.D., Dieppa-Colón, E., Goyco-Blas, J., Peralta-Martínez, K., Scafidi, L., Shah, S., Zawacki, E., Barts, N., Ahn, Y., Hedayati, S., Secor, S.M., Rowe, M.P., 2022. Gut Microbial Ecology of Five Species of Sympatric Desert Rodents in Relation to Herbivorous and Insectivorous Feeding Strategies. Integrative and Comparative Biology icac045.

Kolmogorov, M., Bickhart, D.M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S.B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T.P.L., Pevzner, P.A., 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17, 1103–1110.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359.

Latorre-Pérez, A., Villalba-Bermell, P., Pascual, J., Vilanova, C., 2020. Assembly methods for nanopore-based metagenomic sequencing: a comparative study. Sci Rep 10, 13588.

Lavrinienko, A., Tukalenko, E., Mousseau, T.A., Thompson, L.R., Knight, R., Mappes, T., Watts, P.C., 2020. Two hundred and fifty-four metagenome-assembled bacterial genomes from the bank vole gut microbiota. Sci Data 7, 312.

Levin, D., Raab, N., Pinto, Y., Rothschild, D., Zanir, G., Godneva, A., Mellul, N., Futorian, D., Gal, D., Leviatan, S., Zeevi, D., Bachelet, I., Segal, E., 2021. Diversity and functional landscapes in the microbiota of animals in the wild. Science 372, eabb5352.

Leung, T.L.F., Poulin, R., 2008. Parasitism, Commensalism, And Mutualism: Exploring the many shades of symbioses. Vie Milieu 58, 107–115.

Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R., Gordon, J.I., 2008a. Worlds within worlds: evolution of the vertebrate gut microbiota. Nature Reviews Microbiology 6, 776–788.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., Gordon, J.I., 2008b. Evolution of Mammals and Their Gut Microbes. Science 320, 1647–1651.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676.

Liu, L., Yang, Y., Deng, Y., Zhang, T., 2022. Nanopore long-read-only metagenomics enables complete and high-quality genome reconstruction from mock and complex metagenomes. Microbiome 10, 209.

Ma, J.-E., Jiang, H.-Y., Li, L.-M., Zhang, X.-J., Li, G.-Y., Li, H.-M., Jin, X.-J., Chen, J.-P., 2018. The Fecal Metagenomics of Malayan Pangolins Identifies an Extensive Adaptation to Myrmecophagy. Front. Microbiol. 9, 2793.

Ma, J., Sheng, L., Hong, Y., Xi, C., Gu, Y., Zheng, N., Li, M., Chen, L., Wu, G., Li, Y., Yan, J., Han, R., Li, B., Qiu, H., Zhong, J., Jia, W., Li, H., 2020. Variations of Gut Microbiome Profile Under Different Storage Conditions and Preservation Periods: A Multi-Dimensional Evaluation. Front. Microbiol. 11, 972.

Macdonald, C., Barden, S., Foley, S., 2014. Isolation and characterization of chitin-degrading micro-organisms from the faeces of Goeldi's monkey, *Callimico goeldii*. Journal of Applied Microbiology 116, 52–59.

Magdeleine, A., Tilak, M.-K., Teullet, S., Delsuc, F., 2021. High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing. protocol.io. https://doi.org/dx.doi.org/10.17504/protocols.io.bvyin7ue.

Mallott, E.K., Amato, K.R., 2021. Host specificity of the gut microbiome. Nat Rev Microbiol 19, 639–653.

Marchesi, J.R., Ravel, J., 2015. The vocabulary of microbiome research: a proposal. Microbiome 3, 31.

Marić, J., Križanović, K., Riondet, S., Nagarajan, N., Šikić, M., 2023. Comprehensive benchmarking of metagenomic classification tools for long-read sequencing data. bioRxiv. https://doi.org/10.1101/2020.11.25.397729.

Marotz, C., Cavagnero, K.J., Song, S.J., McDonald, D., Wandro, S., Humphrey, G., Bryant, M., Ackermann, G., Diaz, E., Knight, R., 2021. Evaluation of the Effect of Storage Methods on Fecal, Saliva, and Skin Microbiome Composition. mSystems 6, e01329-20.

Mazel, F., Davis, K., Loudon, A., Kwong, W., Groussin, M., Wegener Parfrey, L., 2018. Is Host Filtering the Main Driver of Phylosymbiosis across the Tree of Life? mSystems 3, e00097-18.

Mazel, F., Guisan, A., Parfrey, L.W., 2023. Transmission mode and dispersal traits correlate with host specificity in mammalian gut microbes. Mol Ecol 0, 1–11.

McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S.F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A.H., Kremer, N., Mazmanian, S.K., Metcalf, J.L., Nealson, K., Pierce, N.E., Rawls, J.F., Reid, A., Ruby, E.G., Rumpho, M., Sanders, J.G., Tautz, D., Wernegreen, J.J., 2013. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A 110, 3229–3236.

Moeller, A.H., Suzuki, T.A., Lin, D., Lacey, E.A., Wasser, S.K., Nachman, M.W., 2017. Dispersal limitation promotes the diversification of the mammalian gut microbiota. PNAS 114, 13768–13773.

Moeller, A.H., Sanders, J.G., 2020. Roles of the gut microbiota in the adaptive evolution of mammalian species. Philosophical Transactions of the Royal Society B: Biological Sciences 375, 20190597.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R., Gordon, J.I., 2011. Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans. Science 332, 970–974.

Nawani, N.N., Kapadnis, B.P., 2003. Chitin degrading potential of bacteria from extreme and moderate environment. IJEB 41, 248-254.

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I.-M., Huntemann, M., Palaniappan, K., Ladau, J., Mukherjee, S., Reddy, T.B.K., Nielsen, T., Kirton, E., Faria, J.P., Edirisinghe, J.N., Henry, C.S., Jungbluth, S.P., Chivian, D., Dehal, P., Wood-Charlson, E.M., Arkin, A.P., Tringe, S.G., Visel, A., Woyke, T., Mouncey, N.J., Ivanova, N.N., Kyrpides, N.C., Eloe-Fadrosh, E.A., 2020. A genomic catalog of Earth's microbiomes. Nature Biotechnology 39, 499-509.

Nishida, A.H., Ochman, H., 2018. Rates of gut microbiome divergence in mammals. Molecular Ecology 27, 1884–1897.

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 27, 824–834.

Olm, M.R., Brown, C.T., Brooks, B., Banfield, J.F., 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. The ISME Journal 11, 2864–2868.

Olm, M.R., Crits-Christoph, A., Diamond, S., Lavy, A., Matheus Carnevali, P.B., Banfield, J.F., 2020. Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries. mSystems 5, e00731-19.

Olsen, M.A., Blix, A.S., Utsi, T.H.A., Sørmo, W., Mathiesen, S.D., 1996. Chitinolytic bacteria in the minke whale forestomach. Ann Zootech 45, 287.

Olsen, M.A., Mathiesen, S.D., 1997. Anatomy and microbiology of the digestive tract of the Adelie penguin. Report of the Norwegian Antarctic research expedition 1996/97, 35–41.

Orkin, J.D., Campos, F.A., Myers, M.S., Cheves Hernandez, S.E., Guadamuz, A., Melin, A.D., 2019. Seasonality of the gut microbiota of free-ranging white-faced capuchins in a tropical dry forest. ISME J 13, 183–196.

Pardosi, L., Suryanto, D., Siregar, A.Z., 2018. Isolation of Chitinolytic Bacteria From Two Lizard Digestive Tract and Characterization of Their Crude Chitinase. International Journal of ChemTech Research 11, 17–28.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055.

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N., Hugenholtz, P., Tyson, G.W., 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol 2, 1533–1542.

Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., Hugenholtz, P., 2021. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Research 50, D785–D794.

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M.C., Rice, B.L., DuLong, C., Morgan, X.C., Golden, C.D., Quince, C., Huttenhower, C., Segata, N., 2019. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell 176, 649-662.e20.

Perez-Lamarque, B., Sommeria-Klein, G., Duret, L., Morlon, H., 2023. Phylogenetic comparative approach reveals evolutionary conservatism, ancestral composition, and integration of vertebrate gut microbiota. Molecular Biology and Evolution 40, msad144.

Portik, D.M., Brown, C.T., Pierce-Ward, N.T., 2022. Evaluation of taxonomic classification and profiling methods for long-read shotgun metagenomic sequencing datasets. BMC Bioinformatics 23, 541.

Poulsen, C.S., Ekstrøm, C.T., Aarestrup, F.M., Pamp, S.J., 2022. Library Preparation and Sequencing Platform Introduce Bias in Metagenomic-Based Characterizations of Microbiomes. Microbiol Spectrum 10, e0009022.

Raimundo, I., Silva, R., Meunier, L., Valente, S.M., Lago-Lestón, A., Keller-Costa, T., Costa, R., 2021. Functional metagenomics reveals differential chitin degradation and utilization features across free-living and host-associated marine microbiomes. Microbiome 9, 43.

Rathore, A.S., Gupta, R.D., 2015. Chitinases from Bacteria to Human: Properties, Applications, and Future Perspectives. Enzyme Research 2015, 791907.

Raymann, K., Brochier-Armanet, C., Gribaldo, S., 2015. The two-domain tree of life is linked to a new root for the Archaea. Proc Natl Acad Sci U S A 112, 6670–6675.

Redford, K.H., 1987. Ants and Termites As Food, in: Genoways, H.H. (Ed.), Current Mammalogy. Springer US, Boston, MA, pp. 349–399.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., Dodsworth, J.A., Hedlund, B.P., Tsiamis, G., Sievert, S.M., Liu, W.-T., Eisen, J.A., Hallam, S.J., Kyrpides, N.C., Stepanauskas, R., Rubin, E.M., Hugenholtz, P., Woyke, T., 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499, 431–437.

Sachs, J.L., Essenberg, C.J., Turcotte, M.M., 2011. New paradigms for the evolution of beneficial infections. Trends Ecol Evol 26, 202–209.

Sanders, J.G., Beichman, A.C., Roman, J., Scott, J.J., Emerson, D., McCarthy, J.J., Girguis, P.R., 2015. Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. Nature Communications 6, 8285.

Sanders, J.G., Sprockett, D.D., Li, Y., Mjungu, D., Lonsdorf, E.V., Ndjango, J.-B.N., Georgiev, A.V., Hart, J.A., Sanz, C.M., Morgan, D.B., Peeters, M., Hahn, B.H., Moeller, A.H., 2023. Widespread extinctions of co-diversified primate gut bacterial symbionts from humans. Nat Microbiol 8, 1039–1050.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., Huttenhower, C., 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nature Methods 9, 811–814.

Segata, N., Börnigen, D., Morgan, X.C., Huttenhower, C., 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nature Communications 4, 2304.

Shaffer, J.P., Nothias, L.-F., Thompson, L.R., Sanders, J.G., Salido, R.A., Couvillion, S.P., Brejnrod, A.D., Lejzerowicz, F., Haiminen, N., Huang, S., Lutz, H.L., Zhu, Q., Martino, C., Morton, J.T., Karthikeyan, S., Nothias-Esposito, M., Dührkop, K., Böcker, S., Kim, H.W., Aksenov, A.A., Bittremieux, W., Minich, J.J., Marotz, C., Bryant, M.M., Sanders, K., Schwartz, T., Humphrey, G., Vásquez-Baeza, Y., Tripathi, A., Parida, L., Carrieri, A.P., Beck, K.L., Das, P., González, A., McDonald, D., Ladau, J., Karst, S.M., Albertsen, M., Ackermann, G., DeReus, J., Thomas, T., Petras, D., Shade, A., Stegen, J., Song, S.J., Metz, T.O., Swafford, A.D., Dorrestein, P.C., Jansson, J.K., Gilbert, J.A., Knight, R., 2022. Standardized multi-omics of Earth's microbiomes reveals microbial and metabolite diversity. Nat Microbiol 7, 2128–2150.

Sharma, A.K., Petrzelkova, K., Pafco, B., Jost Robinson, C.A., Fuh, T., Wilson, B.A., Stumpf, R.M., Torralba, M.G., Blekhman, R., White, B., Nelson, K.E., Leigh, S.R., Gomez, A., 2020. Traditional Human Populations and Nonhuman Primates Show Parallel Gut Microbiome Adaptations to Analogous Ecological Conditions. mSystems 5, e00815-20.

Sharpton, T.J., 2014. An introduction to the analysis of shotgun metagenomic data. Front. Plant Sci. 5, 209.

Šimůnek, J., Hodrová, B., Bartoňová, H., Kopečný, J., 2001. Chitinolytic bacteria of the mammal digestive tract. Folia Microbiol 46, 76–78.

Singleton, C.M., Petriglieri, F., Kristensen, J.M., Kirkegaard, R.H., Michaelsen, T.Y., Andersen, M.H., Kondrotaite, Z., Karst, S.M., Dueholm, M.S., Nielsen, P.H., Albertsen, M., 2021. Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. Nature Communications 12, 2009.

Song, S.J., Amir, A., Metcalf, J.L., Amato, K.R., Xu, Z.Z., Humphrey, G., Knight, R., 2016. Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies. mSystems 1, e00021.

Song, S.J., Sanders, J.G., Delsuc, F., Metcalf, J., Amato, K., Taylor, M.W., Mazel, F., Lutz, H.L., Winker, K., Graves, G.R., Humphrey, G., Gilbert, J.A., Hackett, S.J., White, K.P., Skeen, H.R., Kurtis, S.M., Withrow, J., Braile, T., Miller, M., McCracken, K.G., Maley, J.M., Ezenwa, V.O., Williams, A., Blanton, J.M., McKenzie, V.J., Knight, R., 2020. Comparative Analyses of Vertebrate Gut Microbiomes Reveal Convergence between Birds and Bats. mBio 11, e02901-19.

Souza, C.P., Almeida, B.C., Colwell, R.R., Rivera, I.N.G., 2011. The Importance of Chitin in the Marine Environment. Mar Biotechnol 13, 823–830.

Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., Ettema, T.J.G., 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521, 173–179.

Stevens, C.E., Hume, I.D., 1998. Contributions of Microbes in Vertebrate Gastrointestinal Tract to Production and Conservation of Nutrients. Physiological Reviews 78, 393–427.

Štrojsová, A., Dyhrman, S.T., 2008. Cell-specific β-N-acetylglucosaminidase activity in cultures and field populations of eukaryotic marine phytoplankton: Chitinolytic activity in marine phytoplankton. FEMS Microbiology Ecology 64, 351–361.

Štrojsová, M., Vrba, J., 2005. Direct detection of digestive enzymes in planktonic rotifers using enzyme-labelled fluorescence (ELF). Mar. Freshwater Res. 56, 189–195.

Suzuki, T.A., 2017. Links between Natural Variation in the Microbiome and Host Fitness in Wild Mammals. Integrative and Comparative Biology 57, 756–769.

Swiontek Brzezinska, M., Jankiewicz, U., Burkowska, A., Walczak, M., 2014. Chitinolytic Microorganisms and Their Possible Application in Environmental Protection. Curr Microbiol 68, 71–81.

Taş, N., de Jong, A.E., Li, Y., Trubl, G., Xue, Y., Dove, N.C., 2021. Metagenomic tools in microbial ecology research. Current Opinion in Biotechnology 67, 184–191.

Teullet, S., Tilak, M.-K., Magdeleine, A., Schaub, R., Weyer, N.M., Panaino, W., Fuller, A., Loughry, W.J., Avenant, N.L., Thoisy, B. de, Borrel, G., Delsuc, F., 2023. Metagenomics uncovers dietary adaptations for chitin digestion in the gut microbiota of convergent myrmecophagous mammals. mSystems 8, e00388-23.

Thomas, C.M., Desmond-Le Quéméner, E., Gribaldo, S., Borrel, G., 2022. Factors shaping the abundance and diversity of the gut archaeome across the animal kingdom. Nat Commun 13, 3358.

Trigodet, F., Lolans, K., Fogarty, E., Shaiber, A., Morrison, H.G., Barreiro, L., Jabri, B., Eren, A.M., 2022. High molecular weight DNA extraction strategies for long-read sequencing of complex metagenomes. Molecular Ecology Resources 22, 1786–1802.

Uritskiy, G.V., DiRuggiero, J., Taylor, J., 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6, 158.

Valentini, A., Pompanon, F., Taberlet, P., 2009. DNA barcoding for ecologists. Trends Ecol Evol 24, 110–117.

Vanni, C., Schechter, M.S., Acinas, S.G., Barberán, A., Buttigieg, P.L., Casamayor, E.O., Delmont, T.O., Duarte, C.M., Eren, A.M., Finn, R.D., Kottmann, R., Mitchell, A., Sánchez, P., Siren, K., Steinegger, M., Gloeckner, F.O., Fernàndez-Guerra, A., 2022. Unifying the known and unknown microbial coding sequence space. eLife 11, e67667.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE 9, e112963.

Wardman, J.F., Bains, R.K., Rahfeld, P., Withers, S.G., 2022. Carbohydrate-active enzymes (CAZymes) in the gut microbiome. Nat Rev Microbiol 20, 542–556.

Weissenbach, J., Sghir, A., 2016. Microbiotes et métagénomique. Med Sci (Paris) 32, 937–943.

Whitaker, J.O., Jr., Dannelly, H.K., Prentice, D.A., 2004. Chitinase in Insectivorous Bats. Journal of Mammalogy 85, 15–18.

Whitman, W.B., Coleman, D.C., Wiebe, W.J., 1998. Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A 95, 6578–6583.

Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E.Y. 2017, 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microbial Genomics 3, e000132.

Wu, Y.-W., Simmons, B.A., Singer, S.W., 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32, 605–607.

Xiao, L., Estellé, J., Kiilerich, P., Ramayo-Caldas, Y., Xia, Z., Feng, Q., Liang, S., Pedersen, A.Ø., Kjeldsen, N.J., Liu, C., Maguin, E., Doré, J., Pons, N., Le Chatelier, E., Prifti, E., Li, J., Jia, H., Liu, X., Xu, X., Ehrlich, S.D., Madsen, L., Kristiansen, K., Rogel-Gaillard, C., Wang, J., 2016. A reference gene catalogue of the pig gut microbiome. Nat Microbiol 1, 16161.

Yan, Q., Fong, S.S., 2015. Bacterial chitinase: nature and perspectives for sustainable bioproduction. Bioresources and Bioprocessing 2, 31.

Yan, Y., Ren, S., Duan, Y., Lu, C., Niu, Y., Wang, Z., Inglis, B., Ji, W., Zheng, Y., Si, W., 2021. Gut microbiota and metabolites of α-synuclein transgenic monkey models with early stage of Parkinson's disease. npj Biofilms Microbiomes 7, 69.

Yang, F., Sun, J., Luo, H., Ren, H., Zhou, H., Lin, Y., Han, M., Chen, B., Liao, H., Brix, S., Li, J., Yang, H., Kristiansen, K., Zhong, H., 2020. Assessment of fecal DNA extraction protocols for metagenomic studies. Gigascience 9, giaa071.

Youngblut, N.D., Reischer, G.H., Walters, W., Schuster, N., Walzer, C., Stalder, G., Ley, R.E., Farnleitner, A.H., 2019. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. Nature Communications 10, 2200.

Youngblut, N.D., de la Cuesta-Zuluaga, J., Reischer, G.H., Dauser, S., Schuster, N., Walzer, C., Stalder, G., Farnleitner, A.H., Ley, R.E., 2020. Large-Scale Metagenome Assembly Reveals Novel Animal-Associated Microbial Genomes, Biosynthetic Gene Clusters, and Other Genetic Diversity. mSystems 5, e01045-20.

Youngblut, N.D., Reischer, G.H., Dauser, S., Maisch, S., Walzer, C., Stalder, G., Farnleitner, A.H., Ley, R.E., 2021. Vertebrate host phylogeny influences gut archaeal diversity. Nat Microbiol 6, 1443–1454.

Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., Stott, M.B., Nunoura, T., Banfield, J.F., Schramm, A., Baker, B.J., Spang, A., Ettema, T.J.G., 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541, 353–358.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P.K., Xu, Y., Yin, Y., 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Research 46, W95–W101.

DISCUSSION AND PERSPECTIVES

1. Convergent dietary adaptations in the genomes and gut microbiomes of myrmecophagous mammals – summary of the results and main conclusions

Ant- and termite-eating (*i.e.*, myrmecophagous) mammals are a textbook example of convergent evolution. To understand how these species convergently adapted to this highly specialized diet and whether the same mechanisms were involved between the different species, my PhD project focused on the associated adaptations in their genomes and gut microbiomes. More specifically, this project aimed at understanding the adaptations related to prey perception and digestion, respectively through the study of taste receptor and chitinase genes; the latter being studied in both the host genome and its associated gut microbiota.

1.1. Genomic dietary adaptations to myrmecophagy: what have we learned from the study of two multigenic families?

Multigenic families often evolve following a birth-and-death process where gene gains and losses occur independently in different lineages resulting in variations of gene repertoire sizes between species. Events of gene duplications and pseudogenizations can be associated with changes in selective pressures and reflect adaptations toward new environmental conditions and/or relaxed selective pressures (Demuth and Hahn, 2009; Albalat and Cañestro, 2016).

Focusing on taste perception, we investigated the evolution of sweet, umami, sour, and bitter taste perception using comparative genomics. Because of their narrow diet and absence of mastication facilitated by a peculiar tongue morphology not suited for taste (*e.g.*, Doran and Allbrook, 1973; Casali *et al*, 2017), myrmecophagous placentals are thought to rely less on taste suggesting their taste receptor gene repertoires might have been reduced. Indeed, some myrmecophagous species such as the Malayan pangolin, the short-beaked echidna, and the numbat have lost some taste receptor genes (Liu *et al*, 2016; Zhou *et al*, 2021; Peel *et al*, 2021).

Accordingly, analyzing patterns of shared inactivating mutations in *Tas1r* genes and *PKD2L1* revealed several events of loss of function suggesting losses of sour, sweet, and/or umami perception in many myrmecophagous lineages (Emerling *et al*, *in prep* see Chapter I part 1.2.2). Additionally, preliminary analyses suggested that ant- and termite-eating species experienced a reduction of their *Tas2r* gene repertoires, and therefore alteration of bitter taste perception (Garland, 2018). These results should be confirmed by analyses started at the end of this PhD that will focus on reconstructing the evolutionary history of this gene family in placentals using reconciliation methods and investigating pseudogenization signatures. Overall, these studies highlight several cases of taste receptor gene losses in ant- and termite-eating mammals compared to non-myrmecophagous species, and thus suggest reduced taste perception abilities for these species. To fully decipher how taste perception evolved in these species, the functionality of other receptor genes could be studied such as *OTOP1*, which is also involved in sour perception or receptors associated with fat perception (*e.g*., CD36, GPR120, GRP40, FFAR4). Finally, because many taste receptors are pleiotropic and found expressed notably in several organs of the digestive tract (Wu *et al*, 2002; Taniguchi, 2004; Dyer *et al*, 2005), incorporating transcriptomic data of digestive organs, including the tongue, could help further decipher the molecular mechanisms underlying taste perception losses (see chapter I discussion of part 1.2.3).

To understand how myrmecophagous mammals convergently adapted to digest the chitinous exoskeleton of their prey, we studied the evolution of chitinase genes in these species using comparative genomics and transcriptomics. Previous work revealed that the last common ancestor of placental mammals carried at least five functional chitinase paralogues, *CHIA1-5* (active chitinases also known as Acidic Mammalian Chitinase) (Emerling *et al*, 2018). In non-insectivorous species, some, and sometimes even all, of these *CHIA* genes become non-functional because of relaxed selective pressures (Emerling *et al*, 2018). These genes are indeed involved in chitin digestion as they were found expressed in digestive organs and salivary glands of the Malayan pangolin and giant anteater (Ma *et al*, 2017, 2019; Cheng *et al*, 2023). Myrmecophagous species nevertheless differ in the size of their chitinase gene repertoires (between one and five functional genes) raising questions on how the different paralogues are used to digest chitin. Comparisons of chitinase gene expression profiles in

digestive and non-digestive organs of the southern tamandua and Malayan pangolin highlighted the role of these genes in digestion and revealed they are differentially expressed between species (Allio *et al*, in revision, see Chapter I part I.3.2). As a result of phylogenetic constraints, myrmecophagous species inherited different chitinase gene repertoires (Emerling *et al*, 2018). Molecular tinkering of these gene repertoires enabled these species to digest their prey by expressing their chitinase genes differently (Allio *et al*, in revision). These results therefore highlight the different molecular adaptations involved in prey digestion in convergent myrmecophagous species. To fully understand the adaptations to prey digestion in ant- and termite-eating species, other genes could be studied. For instance, the trehalase enzyme (TREH) degrades trehalose, a sugar found in insect blood. The evolution of the trehalase gene seems to have played a role in the adaptation to an insectivorous diet in mammals (Jiao *et al*, 2019). Besides, this gene was found overexpressed in the small intestine of pangolins compared to closely related non-myrmecophagous species suggesting a role in insect prey digestion (Cheng *et al*, 2023).

Moreover, our results further shed light on the influence of both historical contingency (*i.e*., phylogenetic constraints) as well as selective pressures to adopt a myrmecophagous diet in the convergent evolution of ant- and termite-eating species, consistent with previous comparative morphological (Ferreira-Cardoso *et al*, 2019, 2020, 2022) and genomic studies (Emerling *et al*, 2018). Such analyses highlighted differences in the degree of adaptation between anciently diverged myrmecophagous lineages (*e.g*., pangolins, anteaters: around ~80 Mya; Álvarez-Carretero *et al*, 2022) compared to recently ones (*e.g*., aardwolves: around ~10 Mya; Eizirik *et al*, 2010). Indeed, anteaters and pangolins are highly specialized ant-eating species characterized by extreme morphological and genomic adaptations with many phenotypic trait losses. Conversely, the aardwolf does not present such marked morphological myrmecophagous adaptations and genomic analyses (Westbury *et al*, 2021) did not reveal gene losses associated with the myrmecophagous diet raising questions on how this species specialized toward termite consumption. Analyzing transcriptomic data of digestive organs of the aardwolf will, for instance, help decipher whether its only functional chitinase gene (*CHIA5*) is used for digestion. Besides, studying its gut microbiota in more details might also help understand how the aardwolf digests its prey. These differences in the degree of specialization between myrmecophagous species due to their divergent evolutionary histories

further highlight the importance of combining several approaches to fully understand the phenomenon of convergence (see also parts 2 and 3).

1.2. Role of the gut microbiota in the adaptation to a chitin-rich diet in mammals

Gut microbial symbionts have played a major role in the dietary diversification of mammals as they are involved in digestive processes (*e.g.*, Ley *et al*, 2008; Muegge *et al*, 2011). Previous studies revealed convergences in gut microbiota composition among myrmecophagous species (Delsuc *et al*, 2014) and identified chitinolytic bacteria in the gut microbiota of the Malayan pangolin and giant anteater (Ma *et al*, 2018; Cheng *et al*, 2023). Yet, the potential for chitin digestion in the gut microbiota of all myrmecophagous species remained poorly investigated. The second chapter of my PhD therefore aimed at understanding the role their gut microbiota played in their convergent adaptation to myrmecophagy by i) identifying chitinolytic bacteria that might participate in prey digestion, ii) understanding whether the same bacteria have been convergently recruited among myrmecophagous species, and iii) demonstrating whether there is a pattern of convergence among these species compared to non-myrmecophagous closely related species.

Focusing on myrmecophagous species only, we investigated the presence of chitinolytic bacteria in their gut using shotgun metagenomic sequencing to reconstruct high-quality bacterial genomes and identify chitinase genes (Teullet *et al*, 2023, see Chapter II part II.2). These analyses revealed a potential role of the gut microbiota to digest prey as several chitinolytic bacteria were identified. Chitinase genes were identified in numerous bacterial genomes, mainly belonging to three bacterial families (*i.e.*, Lachnospiraceae, Actualibacteraceae, and Ruminococcaceae) notably in species already known to degrade polysaccharides and/or chitin. Different bacterial genomes carried several chitinase genes suggesting complex microbial chitin-degrading pathways are involved. At least one chitinolytic bacteria was identified in each of the nine focal species representatives of the five myrmecophagous orders. Xenarthran species seem to carry more chitinolytic bacteria than

pangolins, aardwolves, or aardvarks which might reflect phylogenetic constraints as well as environmental influences. Many bacterial genomes were shared between host species, mainly closely related ones emphasizing on the influence of phylosymbiosis whereas other were found in almost all species suggesting these bacterial taxa might have been important for the adaptation toward myrmecophagy. Overall, this comparative analysis of myrmecophagous gut metagenomes highlights the influence of both historical contingency (*i.e.*, constraints) and natural selection (*i.e.*, adaptation toward the same diet) in the convergent adaptation to myrmecophagy.

To better understand how the gut microbiota of these species contributes to prey digestion, such study would beneficiate from the investigation of other microbial genes. For instance, genes involved in trehalose degradation or formic acid detoxification could be investigated, as these functions have been identified in the Malayan pangolin and giant anteater gut microbiota (Cheng *et al*, 2023). Besides, as many mammalian species ingest chitin in their diet, whether specifically or opportunistically while foraging, chitinolytic bacteria can be found in species with diverse diets (*e.g.*, Olsen *et al*, 1996; Šimůnek *et al*, 2001; Whitaker *et al*, 2004; Irulan *et al*, 2011; Macdonald *et al*, 2014; Sanders *et al*, 2015). To understand whether patterns observed in myrmecophagous species are convergent and to decipher influences from phylogeny compared to diet, it is crucial to compare their gut microbiota with non-myrmecophagous species. Comparisons of chitin digestion abilities of placental gut microbiota started at the end of this PhD with the finalization of the dataset at the scale of mammals and metagenome-assembled-genome reconstruction. These analyses will be pursued with the aim of quantifying whether myrmecophagous species carry more and/or different chitinolytic bacteria than non-myrmecophagous species, and to assess whether there are similitudes in species with less specialized insectivorous diets. As shown by this PhD project, both the host and its gut microbial symbionts can participate in chitin digestion. Therefore, combining several approaches to consider genomic adaptations of the host and its microbiome should help us to better understand their relative contribution to chitin digestion. In the framework of the holobiont concept (see part 2), such an integrative approach should shed light, more broadly, on the underlying mechanisms involved in the convergent adaptation to myrmecophagy.

2. Role of the holobiont in the adaptation to myrmecophagy in mammals

The host and its associated symbionts both influence each other's fitness. Each partner evolves reciprocally in response to adaptive changes occurring in the other, a process called coevolution (Suzuki, 2017; Rosenberg and Zilber-Rosenberg, 2018; Berg *et al*, 2020). Microbes play important roles in host functions and health (*e.g*., digestion, immunity, behavior). In return, the host influences assembly and evolution of microbial communities through several mechanisms (*e.g*., immunity system, gut morphology, and physiology) (Suzuki, 2017; Moeller and Sanders, 2020; Mallot and Amato, 2021; Alessandri *et al*, 2022). Moreover, numerous and complex abiotic and biotic factors shape host-microbe interactions. To fully understand an organism's evolution and adaptation toward specific conditions, it is therefore crucial to study its microbiota because of their close relationships and adopt holistic approaches (Limborg and Heeb, 2018; Berg *et al*, 2020).

To better study such system, the hologenomic theory of evolution proposes a framework in which the holobiont (*i.e*., the host and all its symbiotic microorganisms) is the unit of selection. Changes occurring in the hologenome (*i.e*., the set of host and microbial genes), such as mutations, acquisition of new microbes, or horizontal gene transfers, contribute to the holobiont adaptation in changing environments (Zilber-Rosenberg and Rosenberg, 2008). Moreover, if changes in the hologenome are transmitted to the next generation, they can be subject to evolutionary forces such as natural selection or genetic drift (Zilber-Rosenberg and Rosenberg, 2008). Numerous evidences gathered through studies of multiple systems, from plants to animals, now support this theory (Rosenberg and Zilber-Rosenberg, 2018). In this framework rapid changes occurring in the microbiota, because of its dynamic nature, could help the holobiont to adapt to changing environments before host genomic changes occur (Zilber-Rosenberg and Rosenberg, 2008; Rosenberg and Zilber-Rosenberg, 2018). Supporting this hypothesis is the concept of a multilayered hologenome, which defines a continuum from the core microbiota that is adapted to the host, to the flexible pool of microbes influenced by the environment that might help the holobiont to adapt rapidly to changes in selective pressures (Shapira, 2016). Between these core and flexible microbiota

are a multitude of host-microbe interactions with various degrees of association and diverse factors influencing them (Shapira, 2016).

As opposed to approaches separating host-microbe interactions based on the specific effects of the microbes on the host (*i.e.*, antagonistic, mutualistic), holistic approaches, which focus on the holobiont as one entity, are now more often conducted to understand the evolution of the holobiont, and aim at considering the global effect of all these interactions on the host (Berg *et al*, 2020). As host-microbe interactions are strongly influenced by numerous abiotic (*e.g.*, soil, air, water properties) and biotic (*e.g.*, plant and animal interactions) external factors (Singh *et al*, 2020), some authors argue that more global approaches should be conducted to include these factors in eco-holobiont approaches (Singh *et al*, 2020). In the framework of the "One Health" concept (*i.e.*, integrative approaches to study and manage human, environmental, and animal health; onehealthinitiative.com) such approaches should have broad applications notably in medicine and agriculture (Berg *et al*, 2020).

To understand the underlying mechanisms involved in the evolution of the holobiont, the concept of hologenomic adaptations has recently emerged (Suárez and Triviño, 2020). Hologenomic adaptations occur at the level of the holobiont and their evolution can only be explained if the holobiont is considered as one entity (Suárez and Triviño, 2020). This concept is perhaps best illustrated with the example of the dietary adaptation to sanguivory in the common vampire bat (Mendoza *et al*, 2018). Combining comparative genomics and metagenomics on the vampire bat with non-sanguivorous bats enabled the identification of hologenomic adaptations related to the specific requirements of sanguivory, for instance, for taste perception, and carbohydrates, vitamins, and amino-acid metabolism (Mendoza *et al*, 2018). Such kind of study emphasizes on the importance to adopt integrative hologenomic approaches by combining several types of -omics data, from the host and its associated microbes, to understand how the holobiont, as one entity, responds to specific selective pressures (Alberdi *et al*, 2021). To promote hologenomic research, projects like the Earth Hologenome Initiative (www.earthhologenome.com) have been initiated. Indeed, the Earth Hologenome Initiative, aims at collecting paired genomic and metagenomic data from wild organisms to answer eco-evolutionary questions, for instance, linked with the adaptation to climate change, convergent evolution between placental and marsupial species, or adaptations to high altitudes in vertebrates.

In the case of myrmecophagous mammals, this PhD project shed light on how the holobiont might have responded to selective pressures for chitin digestion by combining genomics, transcriptomics, and metagenomics. More specifically, studying both endogenous and microbial chitinases highlighted the different and potentially complementary roles of the host and its gut microbiota in ensuring insect prey digestion. Such complementary approaches raised questions notably regarding the underlying adaptations involved in recently diverged species (*e.g.*, aardwolf) compared to more anciently lineages (*e.g.*, pangolins, anteaters). Indeed, as discussed in part 1, the role of endogenous chitinases for prey digestion in the aardwolf still needs further investigation and raises questions on the potential complementary role of its gut microbiota. It could be hypothesized that recently diverged myrmecophagous species might rely more on their gut microbiota to digest prey than more anciently diverged species as shifts in the gut microbiota composition could happen more easily and rapidly than genomic adaptations (*e.g.*, adaptive substitutions, gene duplications, changes in expression profile) (Rosenberg and Zilber-Rosenberg, 2018). Our analyses revealed that the aardwolf carries chitinolytic bacteria but they were not as diverse as in Xenarthra for instance. Therefore, questions remain regarding how the aardwolf digests its prey. Investigating expression profile of its only functional chitinase in digestive organs should help clarify whether endogenous chitinases participate in chitin digestion in the aardwolf. Additionally, our metagenomic analyses only highlighted the potential for chitin digestion of the gut microbiota of myrmecophagous species but to fully understand its implication in chitin digestion metatranscriptomic data will be necessary. Indeed, some species might have diverse chitinolytic bacteria expressing evenly their chitinase genes, whereas in other species having few chitinolytic bacteria producing abundant chitinases might have been selected. Combining genomic, metagenomic, transcriptomic, and metatranscriptomic data will therefore help to assess the relative contribution of both the host and its gut microbiota in ensuring chitin digestion along the digestive tract, and unravel the different mechanisms involved in the adaptation to myrmecophagy. Such a hologenomic approach has recently been conducted on four ant-eating species (*i.e.*, Malayan pangolin, short-beaked echidna, southern tamandua, and giant anteater) by combining genomics, metagenomics, transcriptomics, and functional assays of enzyme activity (Cheng *et al*, 2023). By studying the functionality and quantifying the expression of specific genes involved in chitin, trehalose, protein, and lipid digestion as well as formic acid detoxification, this study highlighted hologenomic adaptations regarding nutrient

metabolism and detoxification potentially linked to the specificities of the myrmecophagous diet (Cheng *et al*, 2023).

3. Prospects: combining genomics, metagenomics, and morphology

To study the underlying adaptive mechanisms involved in the adaptation toward myrmecophagy in placental mammals, three different and complementary approaches were developed in the ConvergeAnt project (www.convergeant-project.com): morphology, genomics, and metagenomics. As discussed above, combining comparative genomics, transcriptomics, and metagenomics have the potential to help us understand the evolution of convergent myrmecophagous phenotypes through the adaptation for prey digestion. Yet, several links remain to be made between the different parts of this project to get a global view, notably by integrating morphological information into genomic studies.

Morphological and genomic data can help understand the underlying molecular mechanisms involved in the regression of traits characterizing the myrmecophagous phenotypes. For instance, because of their absence of mastication, ant- and termite-eating mammals are characterized by teeth reduction (*e.g.*, enamel alteration in armadillos) or even complete loss (*e.g.*, pangolins, anteaters) (Reiss, 2001; Ferreira-Cardoso *et al*, 2019). Genomic studies shed light on the molecular bases of this regression by highlighting losses of genes involved in enamel and dentin development (Meredith *et al*, 2009; Springer *et al*, 2019; Mu *et al* 2021; Emerling *et al*, 2023; Emerling *et al*, *in prep* see Chapter I part 1.2.2). This absence of mastication was also accompanied by a reduction of their jaw muscles (Ferreira-Cardoso *et al*, 2020), which might correlate with the inactivation of myosin genes expressed in jaw muscles (Emerling *et al*, *in prep* see Chapter I part 1.2.2). Another phenotypic feature associated with their peculiar feeding habits is the modification of their tongue to catch prey making it less suited for taste perception (*e.g.*, Doran and Allbrook, 1973; Reiss, 2001; Casali *et al*, 2017). Studying patterns of taste receptor gene losses revealed that myrmecophagous species might

have lost the ability to perceive several tastes (Emerling *et al*, *in prep* see chapter I part 1.2.2). Combining morphological data linked to tongue gustatory abilities (*e.g*., number and type of papillae, number of taste buds) and genomic data on taste receptor gene repertoire sizes will further help understanding how taste perception abilities were lost in these species, and potentially reveal links between morphological features and genomic signatures.

Morphological and genomic adaptations for prey perception can also be investigated through the study of olfaction. Indeed, in animals, olfaction plays a major role in feeding behavior. The huge variety of odorant molecules is detected thanks to olfactory receptors (ORs), expressed in the nasal cavity (Buck, 1996; Buck, 2000; Hayden and Teeling, 2014). *OR* genes constitute the largest gene family in mammals (~ 1 000 functional copies for most taxa) and their evolution, characterized by duplications and losses specific to certain clades, seems to be linked to ecological characteristics leading to a huge variation in the size of *OR* gene repertoires between species (Niimura, 2009; Hayden *et al*, 2010; Hayden and Teeling, 2014; Hughes *et al*, 2018). Myrmecophagous species rely a lot on olfaction to detect ants and termites as they smell the substrate while foraging suggesting they might have a huge number of functional *OR* genes. The size of the *OR* gene repertoire and olfactory structures can be correlated as it has been shown in placentals in which a positive correlation between the number of functional *OR* genes and the relative size of the cribriform plate (*i.e*., a plate, part of the ethmoid bone, perforated by the passage of olfactory nerves) was found (Bird *et al*, 2018). The only partially myrmecophagous species included in this study was the nine-banded armadillo, which falls at the top of the correlation. Previous work done as part of the ConvergeAnt project has characterized the cribriform plate, olfactory turbinals (*i.e*., bony structures in the nasal cavity connected to the olfactory bulb and covered by the olfactory epithelium), and the olfactory bulb. Those morphological data could be compared with the size of *OR* gene repertoires to better understand olfactory adaptations in relation to the myrmecophagous diet.

Morphological adaptations can also be studied in the light of microbial adaptations. Indeed, both the gut morphology and physiology influence the composition and structure of microbial communities (Stevens and Hume, 1998; McKenzie *et al*, 2017; Reese and Dunn, 2018; Milani *et al*, 2020; Alessandri *et al*, 2022; McCallum and Tropini, 2023). Different

microbial communities carrying different functions can thus be present along the digestive tract of an animal. For instance, the stomach, small intestine, and large intestine each present unique microbiota (McCallum and Tropini, 2023). Moreover, major differences in microbial communities were evidenced when comparing species with simple or complex guts (Alessandri *et al*, 2022). Omnivorous and carnivorous species usually have simple guts and low food retention time whereas herbivorous species often have high retention food time needed for plant fermentation which occurs in enlarged modified parts of the digestive tract that is usually longer. Such morphological and physiological differences of the gut (*e.g.*, total length, pH, oxygen concentration, mucus type, presence and location of fermentation chambers) can have major influences on gut microbial communities (McKenzie *et al*, 2017; Alessandri *et al*, 2022; McCallum and Tropini, 2023). For instance, microbial diversity is strongly affected by the gut physiology in mammals when comparing fermenters and non-fermenters, with the latter having lower microbial diversities (Stevens and Hume, 1998; Reese and Dunn, 2018; Milani *et al*, 2020). Therefore, studying changes in microbial communities along the digestive tract can shed light on the various functions these symbionts ensure and help understand their role in digestion. In addition, comparisons of these changes between host species can reveal convergent compositional patterns among species sharing similar gut morphologies and diets (Milani *et al*, 2020). In the case of myrmecophagous species, their digestive tracts are usually simple, not strongly differentiated, with some variations between species and putative slower digestion compared to closely related species, probably due to the digestion of insect exoskeletons (De Oliveira Firmino *et al*, 2019; Ogunleye *et al*, 2022; personal observations during anteater and armadillo dissections). These morphological observations suggest that microbial communities might not differ a lot along the gut and therefore potentially digest chitin throughout the digestive tract. This hypothesis could be tested by investigating the presence of chitinolytic bacteria along the digestive tract of myrmecophagous species. Moreover, such an analysis could help us to better understand the relative contribution of the host and its microbial symbionts in digesting chitin by combining information of chitinolytic bacteria location within the digestive tract and expression of host chitinases in digestive organs.

4. Conclusion

To conclude this thesis, combining genomics, transcriptomics, and metagenomics approaches has contributed to our understanding of the underlying genomic and microbial mechanisms involved in the convergent adaptation to myrmecophagy in placental mammals regarding prey perception and digestion. Studying the evolutionary histories of two gene families in myrmecophagous species genomes highlighted molecular adaptations associated with reduced taste perception abilities. It also revealed how molecular tinkering enabled myrmecophagous species to use different chitinase repertoires, shaped by phylogenetic constraints, to digest their prey. Investigating the presence of chitinolytic bacteria in their gut microbiota revealed the potential role of their gut symbionts to participate in prey digestion, raising questions on the role of the holobiont in adapting to specific selective pressures linked to the diet. The results from this thesis highlighted the different adaptive mechanisms involved in the convergent adaptation to myrmecophagy. Together they emphasize on the importance to integrate several approaches to fully understand the phenomenon of evolutionary convergence. Finally, studies conducted during this PhD further confirm the influence of both historical contingency (*i.e.*, constraints) and natural selection (*i.e.*, adaptation) in shaping convergent phenotypes. Studying convergent evolution might thus reveal more predictability and repeatability in evolution than previously thought.

Discussion references

Albalat, R., Cañestro, C., 2016. Evolution by gene loss. Nature Reviews Genetics 17, 379–391.

Alberdi, A., Andersen, S.B., Limborg, M.T., Dunn, R.R., Gilbert, M.T.P., 2021. Disentangling host–microbiota complexity through hologenomics. Nat Rev Genet 23, 281–297.

Alessandri, G., Rizzo, S.M., Ossiprandi, M.C., van Sinderen, D., Ventura, M., 2022. Creating an atlas to visualize the biodiversity of the mammalian gut microbiota. Current Opinion in Biotechnology 73, 28–33.

Álvarez-Carretero, S., Tamuri, A.U., Battini, M., Nascimento, F.F., Carlisle, E., Asher, R.J., Yang, Z., Donoghue, P.C.J., dos Reis, M., 2022. A Species-Level Timeline of Mammal Evolution Integrating Phylogenomic Data. Nature 602, 263–267.

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C.C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G.H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J.A., Maguin, E., Mauchline, T., McClure, R., Mitter, B., Ryan, M., Sarand, I., Smidt, H., Schelkle, B., Roume, H., Kiran, G.S., Selvin, J., Souza, R.S.C. de, van Overbeek, L., Singh, B.K., Wagner, M., Walsh, A., Sessitsch, A., Schloter, M., 2020. Microbiome definition re-visited: old concepts and new challenges. Microbiome 8, 103.

Bird, D.J., Murphy, W.J., Fox-Rosales, L., Hamid, I., Eagle, R.A., Van Valkenburgh, B., 2018. Olfaction written in bone: cribriform plate size parallels olfactory receptor gene repertoires in Mammalia. Proc Biol Sci 285, 20180100.

Buck, L.B., 1996. Information Coding in the Vertebrate Olfactory System. Annu. Rev. Neurosci. 19, 517–544.

Buck, L.B., 2000. The molecular architecture of odor and pheromone sensing in mammals. Cell 100, 611–618.

Casali, D.M., Martins-Santos, E., Santos, A.L.Q., Miranda, F.R., Mahecha, G.A.B., Perini, F.A., 2017. Morphology of the tongue of Vermilingua (Xenarthra: Pilosa) and evolutionary considerations. Journal of Morphology 278, 1380–1399.

Cheng, S.-C., Liu, C.-B., Yao, X.-Q., Hu, J.-Y., Yin, T.-T., Lim, B.K., Chen, W., Wang, G.-D., Zhang, C.-L., Irwin, D.M., Zhang, Z.-G., Zhang, Y.-P., Yu, L., 2023. Hologenomic insights into mammalian adaptations to myrmecophagy. National Science Review 10, nwac174.

De Oliveira Firmino, M., Pereira, H.C. da S., Carvalho, L.R.R.A., Guerra, R.R., 2020. External and digestive system morphology of the *Tamandua tetradactyla*. Anatomia, Histologia, Embryologia 49, 97–104.

Delsuc, F., Metcalf, J.L., Wegener Parfrey, L., Song, S.J., González, A., Knight, R., 2014. Convergence of gut microbiomes in myrmecophagous mammals. Molecular Ecology 23, 1301–1317.

Demuth, J.P., Hahn, M.W., 2009. The life and death of gene families. BioEssays 31, 29–39.

Doran, G.A., Allbrook, D.B., 1973. The tongue and associated structures in two species of African pangolins, *Manis gigantea* and *Manis tricuspis*. J Mammal 54, 887–899.

Dyer, J., Salmon, K.S.H., Zibrik, L., Shirazi-Beechey, S.P., 2005. Expression of sweet taste receptors of the T1R family in the intestinal tract and enteroendocrine cells. Biochem Soc Trans 33, 302–305.

Emerling, C.A., Delsuc, F., Nachman, M.W., 2018. Chitinase genes (*CHIAs*) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. Science Advances 4, eaar6478.

Emerling, C.A., Gibb, G.C., Tilak, M.-K., Hughes, J.J., Kuch, M., Duggan, A.T., Poinar, H.N., Nachman, M.W., Delsuc, F., 2023. Genomic data suggest parallel dental vestigialization within the xenarthran radiation. Peer Community Journal 3, e75.

Eizirik, E., Murphy, W.J., Koepfli, K.-P., Johnson, W.E., Dragoo, J.W., Wayne, R.K., O'Brien, S.J., 2010. Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. Molecular Phylogenetics and Evolution 56, 49–63.

Ferreira-Cardoso, S., Delsuc, F., Hautier, L., 2019. Evolutionary Tinkering of the Mandibular Canal Linked to Convergent Regression of Teeth in Placental Mammals. Current Biology 29, 468-475.e3.

Ferreira-Cardoso, S., Fabre, P.-H., Thoisy, B. de, Delsuc, F., Hautier, L., 2020. Comparative masticatory myology in anteaters and its implications for interpreting morphological convergence in myrmecophagous placentals. PeerJ 8, e9690.

Ferreira-Cardoso, S., Claude, J., Goswami, A., Delsuc, F., Hautier, L., 2022. Flexible conservatism in the skull modularity of convergently evolved myrmecophagous placental mammals. BMC Ecology and Evolution 22, 87.

Garland, K. (2018). Evolution of bitter taste receptors in convergent myrmecophagous placentals. Master thesis. University of Montpellier.

Hayden, S., Bekaert, M., Crider, T.A., Mariani, S., Murphy, W.J., Teeling, E.C., 2010. Ecological adaptation determines functional mammalian olfactory subgenomes. Genome Res 20, 1–9.

Hayden, S., Teeling, E.C., 2014. The molecular biology of vertebrate olfaction. Anat Rec (Hoboken) 297, 2216–2226.

Hughes, G.M., Boston, E.S.M., Finarelli, J.A., Murphy, W.J., Higgins, D.G., Teeling, E.C., 2018. The Birth and Death of Olfactory Receptor Gene Families in Mammalian Niche Adaptation. Mol Biol Evol 35, 1390–1406.

Irulan, A., Nathan, P.T., Priya, Y.S., Marimuthu, S., Elangovan, V., 2011. Isolation and characterization of chitinase producing gut microflora of insectivorous bats. Trends in Biosciences 4, 8–11.

Jiao, H., Zhang, L., Xie, H.-W., Simmons, N.B., Liu, H., Zhao, H., 2019. Trehalase Gene as a Molecular Signature of Dietary Diversification in Mammals. Mol Biol Evol 36, 2171–2183.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., Gordon, J.I., 2008. Evolution of Mammals and Their Gut Microbes. Science 320, 1647–1651.

Limborg, M.T., Heeb, P., 2018. Special Issue: Coevolution of Hosts and Their Microbiome. Genes 9, 549.

Liu, Z., Liu, G., Hailer, F., Orozco-terWengel, P., Tan, X., Tian, J., Yan, Z., Zhang, B., Li, M., 2016. Dietary specialization drives multiple independent losses and gains in the bitter taste gene repertoire of Laurasiatherian Mammals. Frontiers in Zoology 13, 28.

Ma, J.-E., Li, L.-M., Jiang, H.-Y., Zhang, X.-J., Li, J., Li, G.-Y., Yuan, L.-H., Wu, J., Chen, J.-P., 2017. Transcriptomic analysis identifies genes and pathways related to myrmecophagy in the Malayan pangolin (*Manis javanica*). PeerJ 5, e4140.

Ma, J.-E., Jiang, H.-Y., Li, L.-M., Zhang, X.-J., Li, G.-Y., Li, H.-M., Jin, X.-J., Chen, J.-P., 2018. The Fecal Metagenomics of Malayan Pangolins Identifies an Extensive Adaptation to Myrmecophagy. Front. Microbiol. 9, 2793.

Ma, J.-E., Jiang, H.-Y., Li, L.-M., Zhang, X.-J., Li, H.-M., Li, G.-Y., Mo, D.-Y., Chen, J.-P., 2019. SMRT sequencing of the full-length transcriptome of the Sunda pangolin (*Manis javanica*). Gene 692, 208–216.

Macdonald, C., Barden, S., Foley, S., 2014. Isolation and characterization of chitin-degrading micro-organisms from the faeces of Goeldi's monkey, *Callimico goeldii*. Journal of Applied Microbiology 116, 52–59.

Mallott, E.K., Amato, K.R., 2021. Host specificity of the gut microbiome. Nat Rev Microbiol 19, 639–653.

McCallum, G., Tropini, C., 2023. The gut microbiota and its biogeography. Nat Rev Microbiol XX, 1–14.

McKenzie, V.J., Song, S.J., Delsuc, F., Prest, T.L., Oliverio, A.M., Korpita, T.M., Alexiev, A., Amato, K.R., Metcalf, J.L., Kowalewski, M., Avenant, N.L., Link, A., Di Fiore, A., Seguin-Orlando, A., Feh, C., Orlando, L., Mendelson, J.R., Sanders, J., Knight, R., 2017. The Effects of Captivity on the Mammalian Gut Microbiome. Integrative and Comparative Biology 57, 690–704.

Mendoza, M.L.Z., Xiong, Z., Escalera-Zamudio, M., Runge, A.K., Thézé, J., Streicker, D., Frank, H.K., Loza-Rubio, E., Liu, Shengmao, Ryder, O.A., Samaniego Castruita, J.A., Katzourakis, A., Pacheco, G., Taboada, B., Löber, U., Pybus, O.G., Li, Y., Rojas-Anaya, E., Bohmann, K., Carmona Baez, A., Arias, C.F., Liu, Shiping, Greenwood, A.D., Bertelsen, M.F., White, N.E., Bunce, M., Zhang, G., Sicheritz-Pontén, T., Gilbert, M.P.T., 2018. Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. Nat Ecol Evol 2, 659–668.

Meredith, R.W., Gatesy, J., Murphy, W.J., Ryder, O.A., Springer, M.S., 2009. Molecular Decay of the Tooth Gene Enamelin (*ENAM*) Mirrors the Loss of Enamel in the Fossil Record of Placental Mammals. PLOS Genetics 5, e1000634.

Milani, C., Alessandri, G., Mancabelli, L., Mangifesta, M., Lugli, G.A., Viappiani, A., Longhi, G., Anzalone, R., Duranti, S., Turroni, F., Ossiprandi, M.C., van Sinderen, D., Ventura, M., 2020. Multi-omics Approaches To Decipher the Impact of Diet and Host Physiology on the Mammalian Gut Microbiome. Appl Environ Microbiol 86, e01864-20.

Moeller, A.H., Sanders, J.G., 2020. Roles of the gut microbiota in the adaptive evolution of mammalian species. Philosophical Transactions of the Royal Society B: Biological Sciences 375, 20190597.

Mu, Y., Huang, X., Liu, R., Gai, Y., Liang, N., Yin, D., Shan, L., Xu, S., Yang, G., 2021. *ACPT* gene is inactivated in mammalian lineages that lack enamel or teeth. PeerJ 9, e10219.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R., Gordon, J.I., 2011. Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans. Science 332, 970–974.

Niimura, Y., 2009. Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents. Human Genomics 4, 107.

Ogunleye, A.T., Olatunji-Akioye, A.O., Emikpe, B.O., Jarikre, T.A., Omotosho, O.O., Olajumole, A.M., 2022. Contrast radiographic anatomy of the gastrointestinal tract of white-bellied pangolin. West Afr J Radiol 29, 8–14.

Olsen, M.A., Blix, A.S., Utsi, T.H.A., Sørmo, W., Mathiesen, S.D., 1996. Chitinolytic bacteria in the minke whale forestomach. Ann Zootech 45, 287.

Reese, A.T., Dunn, R.R., 2018. Drivers of Microbiome Biodiversity: A Review of General Rules, Feces, and Ignorance. mBio 9, e01294-18.

Reiss, K.Z., 2001. Using Phylogenies to Study Convergence: The Case of the Ant-Eating Mammals. Integr Comp Biol 41, 507–525.

Rosenberg, E., Zilber-Rosenberg, I., 2018. The hologenome concept of evolution after 10 years. Microbiome 6, 78.

Sanders, J.G., Beichman, A.C., Roman, J., Scott, J.J., Emerson, D., McCarthy, J.J., Girguis, P.R., 2015. Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. Nature Communications 6, 8285.

Shapira, M., 2016. Gut Microbiotas and Host Evolution: Scaling Up Symbiosis. Trends Ecol Evol 31, 539–549.

Šimůnek, J., Hodrová, B., Bartoňová, H., Kopečný, J., 2001. Chitinolytic bacteria of the mammal digestive tract. Folia Microbiol 46, 76–78.

Singh, B.K., Liu, H., Trivedi, P., 2020. Eco-holobiont: A new concept to identify drivers of host-associated microorganisms. Environmental Microbiology 22, 564–567.

Springer, M.S., Emerling, C.A., Gatesy, J., Randall, J., Collin, M.A., Hecker, N., Hiller, M., Delsuc, F., 2019. Odontogenic ameloblast-associated (*ODAM*) is inactivated in toothless/enamelless placental mammals and toothed whales. BMC Evolutionary Biology 19, 31.

Stevens, C.E., Hume, I.D., 1998. Contributions of Microbes in Vertebrate Gastrointestinal Tract to Production and Conservation of Nutrients. Physiological Reviews 78, 393–427.

Suárez, J., Triviño, V., 2020. What Is a Hologenomic Adaptation? Emergent Individuality and Inter-Identity in Multispecies Systems. Front. Psychol. 11, 187.

Suzuki, T.A., 2017. Links between Natural Variation in the Microbiome and Host Fitness in Wild Mammals. Integrative and Comparative Biology 57, 756–769.

Taniguchi, K., 2004. Expression of the sweet receptor protein, T1R3, in the human liver and pancreas. J Vet Med Sci 66, 1311–1314.

Teullet, S., Tilak, M.-K., Magdeleine, A., Schaub, R., Weyer, N.M., Panaino, W., Fuller, A., Loughry, W.J., Avenant, N.L., Thoisy, B. de, Borrel, G., Delsuc, F., 2023. Metagenomics uncovers dietary adaptations for chitin digestion in the gut microbiota of convergent myrmecophagous mammals. mSystems 8, e00388-23.

Peel, E., Silver, L., Brandies, P., Hayakawa, T., Belov, K., Hogg, C.J., Silver, L., Brandies, P., Hayakawa, T., Belov, K., Hogg, C.J., 2022. Genome assembly of the numbat (*Myrmecobius fasciatus*), the only termitivorous marsupial. Gigabyte 2022, 1–17.

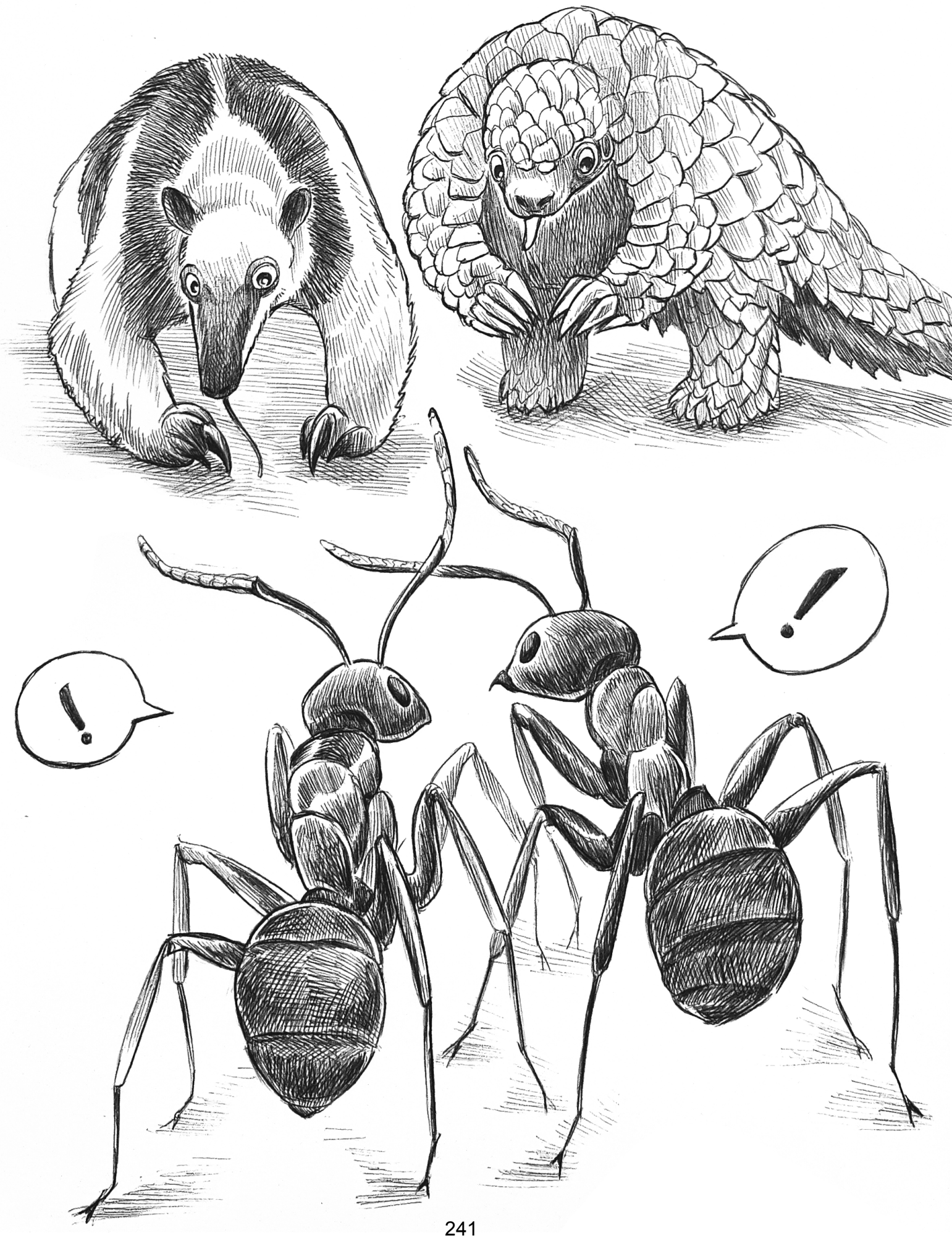
Westbury, M.V., Le Duc, D., Duchêne, D.A., Krishnan, A., Prost, S., Rutschmann, S., Grau, J.H., Dalén, L., Weyrich, A., Norén, K., Werdelin, L., Dalerum, F., Schöneberg, T., Hofreiter, M., 2021. Ecological Specialization and Evolutionary Reticulation in Extant Hyaenidae. Molecular Biology and Evolution 38, 3884–3897.

Whitaker, J.O., Jr., Dannelly, H.K., Prentice, D.A., 2004. Chitinase in Insectivorous Bats. Journal of Mammalogy 85, 15–18.

Wu, V., Rozengurt, N., Yang, M., Young, S., Sinnett-Smith, J., Rozengurt, E., 2002. Expression of bitter taste receptors of the T2R family in gastrointestinal tract and enteroendocrine STC-1 cells. Proceedings of the National Academy of Sciences of the United States of America 99, 2392–7.

Zhou, Y., Shearwin-Whyatt, L., Li, J., Song, Z., Hayakawa, T., Stevens, D., Fenelon, J.C., Peel, E., Cheng, Y., Pajpach, F., Bradley, N., Suzuki, H., Nikaido, M., Damas, J., Daish, T., Perry, T., Zhu, Z., Geng, Y., Rhie, A., Sims, Y., Wood, J., Haase, B., Mountcastle, J., Fedrigo, O., Li, Q., Yang, H., Wang, J., Johnston, S.D., Phillippy, A.M., Howe, K., Jarvis, E.D., Ryder, O.A., Kaessmann, H., Donnelly, P., Korlach, J., Lewin, H.A., Graves, J., Belov, K., Renfree, M.B., Grutzner, F., Zhou, Q., Zhang, G., 2021. Platypus and echidna genomes reveal mammalian biology and evolution. Nature 592, 756–762.

Zilber-Rosenberg, I., Rosenberg, E., 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiology Reviews 32, 723–735.



ANNEX

Annex. Growing and maintaining a network for early career researchers through the Animal Microbiome Research Group

The following paper gives a summary of the Animal Microbiome Research Group (AMRG; https://amrg.mystrikingly.com/) 2nd Virtual Meeting that took place in November 2021. This article also presents news about the AMRG which aims to connect early career microbiome scientists from diverse backgrounds who study a variety of model systems with a broad range of approaches to address various fundamental eco-evolutionary questions.

After attending the AMRG 2021 Virtual Meeting, I had the chance to participate in this paper to promote this research group. Being an early career researcher (ECR) can be quite difficult, especially during covid times, as we need to build our network at the same time we learn how to conduct proper research on our own, and initiatives like the AMRG should be highlighted. Indeed, the AMRG offers a place for ECRs, as well as permanent researchers, to exchange on many topics (*i.e.*, literature, methods) through the Slack channel and through virtual workshops, and promote the work of ECRs. Many members of the AMRG are also part of the organizing team of the Wild Animal Microbiome Evolution (WAME; https://www.wamestn.com/) special topic network (funded by the European Society for Evolutionary Biology, ESEB). This represents a valuable resource for ECRs as it can support them for example through research exchanges. This is also a place for discussions between microbiome researchers, for instance via the organization of virtual symposiums as well as congress symposiums such as the "Microbiomes in the wild" symposium organized during ESEB 2022.

Personal contribution

I contributed, as a second author, as follow: conceptualization of the manuscript structure, writing, reading, and editing.

This manuscript has been **published in Evolutionary Anthropology** and can be found here:
Webb, S.E., **Teullet, S.**, Stothart, M.R., 2022. Growing and maintaining a network for early career researchers through the Animal Microbiome Research Group. Evolutionary Anthropology: Issues, News, and Reviews 31, 108–111. https://doi.org/10.1002/evan.21941.

Received: 10 January 2022 | Accepted: 14 January 2022

DOI: 10.1002/evan.21941

NEWS

Evolutionary Anthropology ISSUES, NEWS, AND REVIEWS WILEY

Growing and maintaining a network for early career researchers through the Animal Microbiome Research Group

1 | INTRODUCTION

On November 10th and 11th 2021, the Animal Microbiome Research Group (AMRG; formerly separate mammalian and avian microbiome research groups) hosted its second annual research meeting. This group was founded by early career researchers (ECRs) out of a desire to connect with other early career microbiome scientists—individuals who might otherwise be the only scholar to study wildlife microbiomes within their lab group, department, or institution. Connections among researchers are established during an annual virtual multi-day meeting. During these meetings, ECRs from around the world present new findings and discuss microbiome-research related topics. Connections made are then reinforced using a Slack workspace, which allows for the maintenance of a lasting network. The group's inaugural meeting and objectives have been summarized previously.[1] Here, we provide group updates, summarize the 2021 meeting, and offer a one-year retrospective on the creation of an ECR virtual network from the participants' viewpoint.

2 | WHAT'S NEW?

The 2021 meeting spanned 2 days and included research talks and group discussions. In contrast to the 2020 meeting, the 2021 symposium included a virtual poster session and keynote lectures from early career professors who are leaders in the field: Dr. Kevin Kohl (University of Pittsburgh, USA), Dr. Qiyun Zhu (Arizona State University, USA), and Dr. Sarah Knowles (University of Oxford, UK). Attendees had the opportunity to meet with keynote speakers in small, informal groups. As of the 2021 meeting, the formerly separate mammalian and avian microbiome research groups merged to form the more inclusive AMRG. This merger will enrich this nascent virtual network by including researchers who study a greater breadth of host systems.

3 | TALKS & POSTERS

Host and environmental factors shape the microbiome's composition and structure, which can affect host phenotype and fitness. Understanding these bidirectional effects is at the core of microbiome studies as highlighted by the presentations given this year.

Among host factors, host genetic variation can be linked to the gut microbiome composition as shown by Charli Davies (University of East Anglia, UK) working on the Seychelles warbler. She showed that certain alleles of the Major Histocompatibility Complex can be associated with gut microbiome compositional shifts.[2] Gestation in females, which has important consequences on their physiology, is another important factor shaping the gut microbiome as illustrated by Chelsea Southworth's poster (University of Notre Dame, USA) on the effects of pregnancy in baboons.

Host diet also strongly influences the gut microbiome and is often studied alongside other factors like host evolutionary history. Dr. Elin Videvall's work (Brown University, USA) on the link between diet and gut microbiome composition in three giraffe species revealed differences among species. At a larger taxonomic scale, Sophie Teullet (University of Montpellier, France) studies the convergent adaptation to myrmecophagy in placentals and tries to reconstruct bacterial genomes and identify microbial genes that might play a role in prey digestion. Dr. Kevin Kohl presented results from experimental approaches demonstrating that the gut microbiome contributes to diet selection in mice.[3] Dr. Sarah Knowles found seasonal shifts in the gut microbiome composition of wild wood mice,[4] likely driven by diet. She also showed that social interactions are a strong predictor of the gut microbiota.[5]

Microbiomes are partly shaped by the dispersal of microbiota between hosts, a process which can be strongly structured by host social behavior. Hanna Bensch (Linnaeus University, Sweden) demonstrated this in the form of a clear "colony" effect on Damaraland mole rat gut microbiomes. Jonas Torfs (University of Antwerp, Belgium) presented his PhD project on the microbiome-gut-brain axis in captive bonobos where he will investigate effects of grooming. Dr. Sally Bornbusch (Smithsonian's National Zoo & Conservation Biology Institute, USA) emphasized that microbes can also be acquired from the environment by showing that wild ring-tailed lemurs harbored a greater relative abundance of soil-associated microbes than captive-environment conspecifics.[6] Thus, both social and environmental transmission can influence gut microbiome composition. This was reinforced by Dr. Aura Raulo (University of Turku, Finland), who reported that among wild wood mice, anaerobic taxa in the microbiome were linked to social patterns, while aerotolerant taxa underlay spatial patterns.

[Correction added after first online publication on February 21, 2022. Sophie Teullet's affiliation has been amended.]

External factors also shape an individual's microbiome with consequences on host fitness, especially in the face of anthropogenic perturbations. Dr. Kevin Kohl showed that microbial communities contributed to host thermal tolerance in tadpoles subjected to temperature increase.[7] In wild *Anolis* lizards, short-term climate warming does not have a significant effect on the gut microbiome whereas sustained drought does, as explained in Claire E. Williams's poster (University of Nevada, USA). In addition to climate change, pollutants and radiation exposure can also impact wild animals gut microbiome as illustrated by Dr. Diana Lafferty (Northern Michigan University, USA) working on wild boars in the Fukushima difficult-to-return zone, and Sameli Piirto (University of Jyväskylä, Finland) whose PhD project will investigate the effects of radionuclides exposure on great tits in Chernobyl (Ukraine).

Gut microbiomes vary over time. Empirical Dynamic Modeling has revealed a non-linear and highly chaotic dynamic of the microbiome, as summarized in Dr. Amy R. Sweeny's poster (University of Edinburgh, UK). When studying changes in the microbiome and how different factors influence this dynamic, the analysis[8] and visualization[9] of longitudinal data are important, as explained by Dr. Laura Grieneisen (University of Minnesota, USA). For instance, it is possible to study the gut microbiome's heritability as illustrated by her work on wild baboons.[10] Shasta Webb (University of Calgary, Canada) studies the effect of drought on the gut microbiome of wild white-faced capuchin monkeys and showed, in particular, lower microbiome richness during drought periods with microbial phyla shifts associated with rain level. Mason Stothart (University of Calgary, Canada) presented his work linking environmental variation and gut microbiome variation in feral horses,[11] and reported correlates between the microbiome and host survival. Dr. Sarah Worsley (University of East Anglia, UK) demonstrated a link between host survival and the gut microbiome composition of the Seychelles warblers.[12] Lifetime data also offer the possibility to investigate the link between host aging and the gut microbiome. Despite being a dynamic system, the gut microbiome might stabilize with host age as suggested by Dr. Alice Risely's results (Ulm University, Germany) on the stability and repeatability of wild meerkats' gut microbiomes.

While 16S sequencing remains the prevailing methodology, metagenomics were more common this year. These methods require different bioinformatic tools and analytical approaches than 16S data. Dr. Qiyun Zhu showcased methods created by his research group which allow for phylogeny-informed analysis of metagenomic data using the Web Of Life tree (https://biocore.github.io/wol/[13]) and Woltka (https://github.com/qiyunzhu/woltka[14]), an approach which advocates for a newly proposed unit, the Operational Genomic Unit (OGU).

For metagenomics approaches to be feasible for a wide array of host systems, we must expand the diversity of available microbial reference genomes. Additionally, we must begin to consider other types of 'omics data (e.g., transcriptomics, metabolomics) and microbiomes (e.g., oral, skin). The Earth Hologenomic Initiative (www.earthhologenome.org)—introduced by Dr. Aoife Leonard (University of Copenhagen, Denmark)—offers an ambitious solution to this problem, by seeking to collaborate with researchers from around the world to simultaneously generate a diversity of host and microbiome 'omics data.

4 | DISCUSSIONS

In addition to presentations, discussions with specific prompts are a key component of the annual AMRG meeting. Prior to the start of each discussion, conference organizers posted relevant resources in Slack. On Day 1, participants were invited to discuss microbiome data visualization. An emergent theme from these discussions was the limitation and over-reliance on visualizing complex microbiome beta diversity data in 2-dimensional ordination plots. By contrast, some attendees argued that visualizations of microbiome data would benefit from presentation of raw beta-diversity distances, and that more schematic representations of microbial relationships to the host are needed. Discussants agreed that there is not one "correct" way to visualize complex microbial systems.

On Day 2, participants were encouraged to discuss diversity, equity, and inclusion in microbiome research, namely: (1) racism in microbiome research; (2) ethics of indigenous microbiome research; (3) inclusivity in microbiome research; and (4) microbes, social equity, and human health. Initial discussions started on the equitable attribution of co-authorship and whether co-authorship is the best and most equitable reward for people who contribute to a project, particularly in cases when researchers are collaborating with people outside of academia. For many outside of academia, authorship is meaningless, and more tangible efforts such as training or funding are more valuable. Another major point of discussion centered around positionality of microbiome researchers, namely how biases can stem from ethnic, racial, socioeconomic, and language backgrounds. Actionable items that resulted from this discussion included that (1) researchers/PIs need to actively recruit students from underrepresented groups and maintain contact with them; and (2) researchers should translate their work into other languages, particularly when they work with people in different countries who speak different languages. We would also like to highlight the Microbes and Social Equity (MSE) Working Group founded by Dr. Sue Ishaq, which is composed of microbiome researchers aiming to meaningfully affect social change through microbiome research.[15] Many members of the AMRG are also active members of the MSE group. Importantly, issues of structural inequity and discrimination are not unique to microbiome research. Actions that arose in this discussion are applicable to a wide variety of academic fields.

5 | A GROWING NETWORK

The AMRG has grown over its first year. The Slack workspace, which now boasts 244 community members, is structured around 6 channels in which members can (1) introduce themselves to the AMRG, (2) circulate recently published articles, (3) advertise talks, funding opportunities, job/student postings, or journal special issues,

(4) exchange advice on experimental design, lab techniques, or analytical approaches and troubleshoot problems, and (5) launch collaborative projects. Aside from public channels, the Slack workspace also allows for direct messaging, enabling researchers to connect privately.

The periods of fastest membership growth and greatest Slack engagement occurs immediately before and after annual meetings, with plateaus between meetings. An exception coincides with the publication of the first meeting summary[1] and subsequent re-circulation of the Slack group invitation on social media (Figure 1a). Social media appears critical for growing membership, but notably, this membership growth event did not cause an increase in the number of Slack posts, suggesting that new members do not immediately create posts in the Slack group (Figure 1b). However, this event did appear to trigger a sustained increase in member engagement (Figure 1c).

Maintaining member engagement throughout the year may require concerted posting efforts by community members to keep the workspace contents new (i.e., engagement begets engagement). For example, we see a positive correlation between the number of posts per week and the proportion of active but non-posting members (Figure 1d). Member-led workshops have recently been proposed as a way to maintain engagement in Slack. A qualitative analysis of the Slack aggregate data suggests that additional events could help sustain member engagement between annual meetings.

6 | CONCLUSION

The AMRG is a growing network of ECRs studying host-microbiome relationships in the wild. The virtual home of the AMRG annual meetings and community has been valuable in an era where travel and in-person meetings are curtailed in the interest of public health. Despite a successful first year, there remains room for a growth in the diversity of methods (16S amplicon sequencing most prevalent), host systems (mammalian and avian hosts predominate),

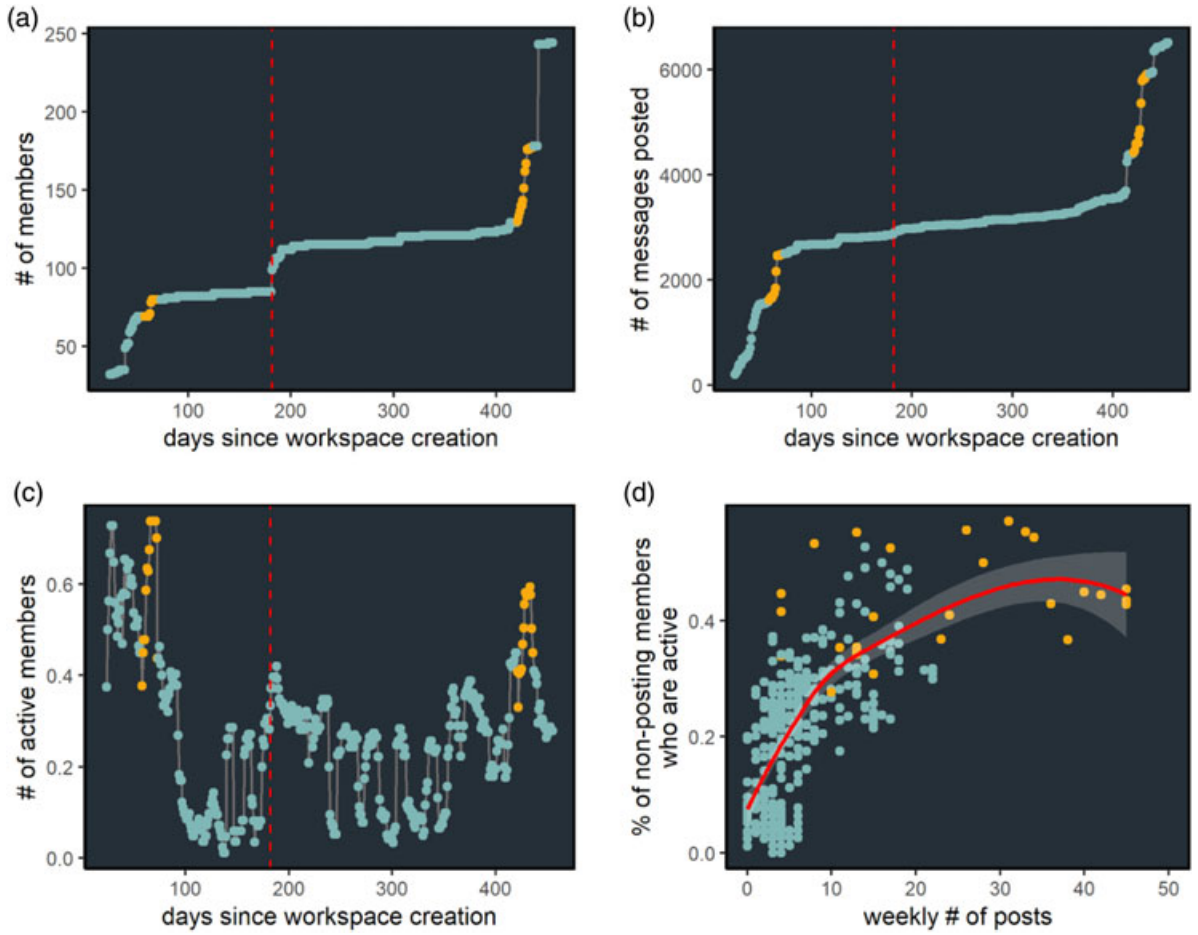


FIGURE 1 Days since the creation of the Slack workspace versus (a) number of group members, (b) number of messages posted, and (c) percent of AMRG members active on Slack in a 7-day sliding window. (d) Number of posts made in a sliding 7-day window versus percent of non-posting members who are active in the workspace. Active members are those who read posts or follow links (does not necessitate posting). Points colored based on whether they are (yellow), or are not (light blue), within 7 days of an Animal Microbiome Research Group meeting. Red vertical dotted line corresponds to a period of AMRG information circulation on social media. Solid red line is the best-fit locally weighted smoothing regression with standard error shading

microbiome types (focus is often on gut microbiomes), and global membership seen within the AMRG group (currently heavily biased to the global north). Those interested in joining and helping to grow this vibrant virtual community can do so by visiting https://amrg.mystrikingly.com/#join-us or sending a direct message on Twitter (@AnimMicrobioRes).

ACKNOWLEDGMENTS

We thank A. Sweeny, L. Petrullo, and A. Baniel—the creators of the Animal Microbiome Research Group—for their feedback and guidance. This is contribution ISEM 2022-014 of the Institut des Sciences de l'Evolution de Montpellier. Sophie Teullet has been supported by a grant from the European Research Council (ConvergeAnt project: ERC-2015-CoG-683257).

DATA AVAILABILITY STATEMENT

Not applicable.

Shasta E. Webb[1], Sophie Teullet[2], Mason R. Stothart[3]

[1]*Department of Anthropology and Archaeology, University of Calgary, Calgary, Alberta, Canada*

[2]*ISEM, Univ Montpellier, CNRS, IRD, Montpellier, France*

[3]*Department of Ecosystem and Public Health, University of Calgary, Calgary, Alberta, Canada*

Correspondence

Shasta E. Webb, Department of Anthropology and Archaeology, University of Calgary, Calgary, Alberta, Canada.

Email: shasta.webb@ucalgary.ca

ORCID

Shasta E. Webb https://orcid.org/0000-0002-9329-2553

Sophie Teullet https://orcid.org/0000-0003-2693-1797

Mason R. Stothart https://orcid.org/0000-0002-2863-908X

REFERENCES

[1] Petrullo L, Baniel A, Sweeny AR. 2021. Establishing a virtual network in mammalian microbiome research. Evol Anthropol 30(2):105–107.

[2] Davies CS., Worsley SF, Maher KH, et al. 2021. Immunogenetic variation shapes the gut microbiome in a natural vertebrate population. (In Review). https://doi.org/10.21203/rs.3.rs-703361/v1

[3] Trevelline BK, Kohl KD. 2020. Microbial control over host diet selection. BioRxiv 2020.07.02.184382.

[4] Marsh KJ, Raulo AM, Brouard M, et al. 2021. Synchronous seasonality in the gut microbiota of wild wood mouse populations. BioRxiv 2021.10.15.464528.

[5] Raulo A, Allen BE, Troitsky T, et al. 2021. Social networks strongly predict the gut microbiota of wild mice. ISME J 15:2601–2613.

[6] Bornbusch SL, Greene LK, Rahobilalaina S, et al. 2021. Gut microbiota of ring-tailed lemurs (*Lemur catta*) vary across natural and captive populations and correlate with environmental microbiota. BioRxiv 2021.06.27.450077.

[7] Fontaine SS, Mineo PM, Kohl KD. 2021. Experimental depletion of gut microbiota diversity reduces host thermal tolerance and fitness under heat stress in a vertebrate ectotherm. BioRxiv 2021.06.04.447101.

[8] Björk JR, Dasari M, Grieneisen L, et al. 2019. Primate microbiomes over time: longitudinal answers to standing questions in microbiome research. Am J Primatol 81:e22970.

[9] Fink I, Abdill RJ, Blekhman R, et al. 2021. BiomeHorizon: visualizing microbiome time series data in R. BioRxiv 2021.08.29.458140.

[10] Grieneisen L, Dasari M, Gould TJ, et al. 2021. Gut microbiome heritability is nearly universal but environmentally contingent. Science 373:181–186.

[11] Stothart MR, Greuel RJ, Gavriliuc S, et al. 2021. Bacterial dispersal and drift drive microbiome diversity patterns within a population of feral hindgut fermenters. Mol Ecol 30(2):555–571.

[12] Worsley SF, Davies CS, Mannarelli M-E, et al. 2021. Gut microbiome composition, not alpha diversity, is associated with survival in a natural vertebrate population. Anim Microbiome 3:84.

[13] Zhu Q, Mai U, Pfeiffer W, et al. 2019. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains bacteria and Archaea. Nat Commun 10:5477.

[14] Zhu Q, Huang S, Gonzalez A, et al. 2021. OGUs enable effective, phylogeny-aware analysis of even shallow metagenome community structures. BioRxiv. 2021.04.04.438427.

[15] Ishaq SL, Parada FJ, Wolf PG, et al. 2021. Introducing the microbes and social equity working group: considering the microbial components of social, environmental, and health justice. MSystems 6:e00471-21.

FRENCH SUMMARY

La convergence évolutive constitue un phénomène fascinant pour étudier la répétabilité et la prédictibilité de l'évolution. La convergence correspond à l'évolution répétée et indépendante de phénotypes similaires dans plusieurs lignées (Arendt et Reznick, 2008). La convergence évolutive peut être le résultat d'une adaptation à des pressions sélectives similaires (*i.e.*, sélection naturelle), mais aussi de contraintes (*i.e.*, contingence historique) (Losos, 2011). De nombreux exemples de convergence ont été rapportés chez divers organismes (*e.g.*, Blount *et al*, 2018). Parmi eux, la myrmécophagie (*i.e.*, consommation de fourmis et de termites) représente un exemple classique de convergence liée au régime alimentaire chez les mammifères.

Une espèce est considérée comme strictement myrmécophage lorsque son régime alimentaire est composé de plus de 90 % de fourmis et/ou de termites (Redford, 1987). Les espèces myrmécophages ont évolué indépendamment au moins cinq fois chez les mammifères placentaires au sein desquels elles sont retrouvées dans cinq ordres : Tubulidentata (oryctérope), Pilosa (fourmiliers), Cingulata (tatous), Pholidota (pangolins) et Carnivora (protèles). Ces espèces sont caractérisées par des adaptations morphologiques remarquables qui ont évolué de manière convergente (Redford, 1987 ; Reiss, 2001) : des membres antérieurs robustes et des griffes utilisés pour creuser dans les fourmilières et termitières, un crâne allongé avec une réduction ou même une perte complète des dents, des muscles de la mâchoire modifiés, des glandes salivaires hypertrophiées produisant des quantités importantes de salive visqueuse, une langue longue, fine et extensible utilisée pour attraper et ingérer rapidement des proies sans mastication (compensée par des adaptations anatomiques de leur tube digestif), et un faible taux métabolique en raison de leur régime alimentaire énergétiquement pauvre et de l'ingestion de matériel non digestible (Griffiths, 1968 ; Patterson, 1975 ; McNab, 1984 ; Reiss, 2001 ; Casali *et al*, 2017 ; Ferreira-Cardoso *et al*, 2019, 2020, 2022). Toutes les espèces myrmécophages ne possèdent pas ces caractéristiques anatomiques marquées. Les pangolins et les fourmiliers présentent les adaptations morphologiques les plus extrêmes (*i.e.*, crâne allongé et absence totale de dents), alors que les tatous représentent des morphologies intermédiaires et que le protèle ne présente que peu d'adaptations typiques du phénotype myrmécophage (Patterson, 1975 ; Reiss, 2001).

L'évolution convergente de ces espèces vers le même régime alimentaire soulève des questions sur les mécanismes impliqués. Comment les espèces myrmécophages se sont-elles adaptées de manière convergente à ce régime alimentaire ? Les mêmes mécanismes évolutifs

ont-ils été impliqués entre les différentes espèces ? Le projet ConvergeAnt (www.convergeant-project.com/) propose une approche intégrative combinant morphologie, génomique, et métagénomique pour comprendre les mécanismes adaptatifs impliqués dans l'évolution convergente vers la myrmécophagie et comment ces trois niveaux d'étude interagissent. Dans le cadre de ce projet, mon sujet de thèse visait à étudier les adaptations liées à ce régime alimentaire à la fois dans les génomes et les microbiomes intestinaux des espèces myrmécophages en utilisant des approches de génomique et métagénomique comparatives. Plus spécifiquement, pour comprendre comment ces espèces perçoivent et digèrent leurs proies, mon projet de thèse s'est intéressé respectivement à l'étude des gènes des récepteurs gustatifs et de chitinases ; ces derniers étant étudiés à la fois dans les génomes des espèces hôtes et de leurs microbiotes intestinaux.

Apport de l'étude de deux familles multigéniques pour comprendre les adaptations génomiques en lien avec la myrmécophagie chez les mammifères

Les familles multigéniques sont composées d'un ensemble de gènes homologues (*i.e.*, partageant une origine évolutive commune), qui ont évolué à partir d'un gène ancestral par duplications successives (Koonin, 2005 ; Nei et Rooney, 2005). Au cours de l'évolution d'un gène à partir d'un gène ancestral, on distingue i) les gènes paralogues ayant évolué à partir d'un gène ancestral commun par duplications successives, et ii) les gènes orthologues ayant évolué à partir d'un gène ancestral commun par spéciation au cours de laquelle chaque copie se retrouve dans chacune des deux nouvelles espèces (Koonin, 2005). Plusieurs modèles d'évolution des familles multigéniques existent pour décrire le devenir des différentes copies (*i.e.*, duplication, pseudogénisation, néo-fonctionnalisation, sous-fonctionnalisation) (Koonin, 2005 ; Nei et Rooney, 2005). Les familles de gènes évoluent souvent selon un processus de naissance-mort où les gains et les pertes de gènes se produisent indépendamment dans différentes lignées, entraînant des variations de la taille des répertoires de gènes entre espèces.

Les événements de duplication de gènes et de pseudogénisation se produisant au sein d'une famille de gènes peuvent être associés à des changements dans les pressions sélectives (Demuth et Hahn, 2009). Les pertes de gènes peuvent être adaptatives (en réponse à un changement de pression sélective) ou neutres si la fonction est dispensable dans le nouvel environnement (Albalat et Cañestro, 2016 ; Sharma *et al*, 2018). L'évolution des familles de

gènes peut donc être étudiée pour comprendre les adaptations moléculaires à des pressions sélectives spécifiques et, plus généralement, pour comprendre des processus évolutifs tels que la convergence. Pour comprendre le phénomène de convergence chez les mammifères myrmécophages et plus précisément les mécanismes moléculaires sous-tendant la perception et digestion des proies chez ces espèces, le premier chapitre de ma thèse s'est focalisé sur l'étude de deux familles de gènes, les récepteurs gustatifs (*Tasrs*) et les chitinases (*CHIAs*).

Le goût permet à un organisme d'identifier les sources de nourriture et d'éviter les toxines. Cinq goûts sont généralement reconnus : acide, salé, sucré, umami et amer (Chandrashekar *et al*, 2006 ; Chaudhari et Roper, 2010). Ces cinq goûts sont reconnus par différents récepteurs gustatifs. Les goûts salé et acide sont détectés grâce à des canaux ioniques mais les mécanismes de détection restent encore peu connus. PKD2L1 a cependant été proposé comme un récepteur intervenant dans la perception du goût acide (*e.g*., Chandrashekar *et al*, 2006 ; Huang *et al*, 2006 ; Kataoka *et al*, 2008). Les goûts sucré, umami et amer sont perçus grâce à des récepteurs gustatifs transmembranaires (TASRs) classés en deux types : les TAS1Rs impliqués dans la perception des goûts sucré et umami et les TAS2Rs pour le goût amer (Chaudhari et Roper, 2010) avec respectivement comme gènes codants les *Tas1rs* (*Tas1r1* et *Tas1r3* pour l'umami et *Tas1r2* et *Tas1r3* pour le sucré) et les *Tas2rs* (~ 30 copies) (Hoon *et al*, 1999 ; Adler *et al*, 2000 ; Chandrashekar *et al*, 2000). L'évolution des gènes *Tas1rs* et *Tas2rs* chez les placentaires est caractérisée par des duplications et des pertes spécifiques à certains clades qui semblent être liées à des adaptations en lien avec le régime alimentaires (*e.g*., Hu et Shi, 2013 ; Hong et Zhao, 2014 ; Shan *et al*, 2018).

Dans le cas des mammifères myrmécophages, il est supposé que leurs capacités gustatives soient réduites. En effet, leur langue, avec peu de papilles gustatives et l'absence de certains types de papilles, n'est pas adaptée à la perception du goût, en particulier chez les fourmiliers et les pangolins (Doran et Allbrook, 1973 ; Casali *et al*, 2017). De plus, les espèces myrmécophages ont un régime alimentaire hautement spécialisé et ne mastiquent pas leurs proies. Ces caractéristiques ont été suggérées comme liées à la perte de récepteurs gustatifs chez les pangolins et les mammifères marins (Feng *et al*, 2014 ; Liu *et al*, 2016). Par conséquent, une réduction des répertoires de gènes des récepteurs gustatifs chez les espèces myrmécophages par rapport aux autres placentaires est prédite, avec des pertes de gènes potentiellement partagées. Une autre hypothèse non mutuellement exclusive réside, dans le

cas des gènes *Tas2rs*, dans le fait que certaines copies pourraient être conservées chez les mammifères myrmécophages, leur permettant la détection des toxines présentes dans leurs proies.

Le premier chapitre de ma thèse s'est donc intéressé à l'évolution de la perception des goûts sucré, umami, acide et amer par des approches de génomique comparative. La recherche et analyse de mutations inactivatrices dans les gènes *Tas1rs* et *PKD2L1* a révélé plusieurs événements de pseudogénisation suggérant des pertes indépendantes de la perception de l'acide, du sucré et/ou de l'umami dans de nombreuses lignées myrmécophages. De plus, il a déjà été montré que les espèces myrmécophages ont connu une réduction de leurs répertoires de gènes *Tas2rs* et donc de la perception du goût amer (Garland, 2018). Ces résultats devront être confirmés par des analyses commencées en fin de thèse et se concentrant sur la reconstruction de l'histoire évolutive de cette famille de gènes chez les placentaires. Pour mieux comprendre comment la perception du goût a évolué chez ces espèces, la fonctionnalité d'autres gènes de récepteurs pourrait être étudiée, comme *OTOP1*, qui est également impliqué dans la perception du goût acide, ou les récepteurs associés à la perception du gras (*e.g.*, CD36, GPR120, GRP40, FFAR4). Enfin, de nombreux récepteurs gustatifs étant pléiotropes et s'exprimant notamment dans plusieurs organes du tube digestif (Wu *et al*, 2002 ; Dyer *et al*, 2005), l'intégration des données transcriptomiques pour ces organes digestifs, y compris la langue, pourrait aider à mieux décrypter les mécanismes moléculaires qui sous-tendent la perte de certains goûts.

Les mammifères myrmécophages ingèrent de grandes quantités d'insectes eusociaux dont ils doivent digérer l'exosquelette chitineux pour accéder aux nutriments. Pour comprendre comment les espèces myrmécophages se sont adaptées de manière convergente à la digestion de leurs proies et si les mêmes mécanismes sous-jacents sont impliqués, le premier chapitre de ma thèse a également porté sur l'étude de l'évolution des gènes de chitinases, des enzymes permettant de dégrader la chitine.

Différents types de chitinases hydrolysent les liaisons glycosidiques de la chitine qui est un polymère de β-N-acétylglucosamine. Les chitinases de mammifères sont généralement classées en deux catégories en fonction de leur activité : i) les chitinases (ou chitinases actives) ayant une fonction catalytique capable d'hydrolyser la chitine et ii) les chitinases inactives qui ne peuvent pas dégrader la chitine mais peuvent s'y lier grâce au domaine de liaison à la chitine

et sont qualifiées de protéines *chitinases-like*. Tous ces gènes de chitinases sont homologues et forment une famille multigénique à l'histoire évolutive complexe chez les vertébrés, au cours de laquelle plusieurs événements de duplication de gènes ont conduit à différents paralogues (Bussink *et al*, 2007 ; Funkhouser et Aronson, 2007 ; Hussain et Wilson, 2013).

Des chitinases ont été identifiées chez de nombreuses espèces de vertébrés. Jeuniaux (1971) a émis l'hypothèse que l'ancêtre commun des vertébrés avait un régime alimentaire composé de chitine et qu'à partir de cet ancêtre les espèces n'ayant pas besoin de chitinases pour la digestion, suite à un changement dans leur régime alimentaire, ne devraient plus en produire. Emerling *et al* (2018) ont testé cette hypothèse en reconstruisant l'histoire évolutive des gènes *CHIA* chez les placentaires. Ils ont démontré que l'ancêtre commun des placentaires aurait pu avoir au moins cinq copies fonctionnelles des gènes *CHIA* (*CHIA1-5*) suggérant qu'il était insectivore et que certaines copies ont été perdues dans plusieurs lignées indépendantes, notamment carnivores et herbivores. Ceci a conduit à une corrélation positive entre le nombre de gènes *CHIA* fonctionnels portés par une espèce et le pourcentage d'invertébrés dans son régime alimentaire (Emerling *et al*, 2018). Les espèces myrmécophages, présentent entre un et cinq gènes *CHIA* fonctionnels (Emerling *et al*, 2018). L'oryctérope porte cinq gènes *CHIA* fonctionnels dans son génome. Le fourmilier à collier (*Tamandua tetradactyla*) ne possède que quatre gènes *CHIA* fonctionnels, *CHIA5* étant pseudogénisé. Comme la plupart des carnivores, le protèle (*Proteles cristatus*) ne possède qu'un seul gène fonctionnel, *CHIA5*. Les pangolins tels que le pangolin malais (*Manis javanica*) ou le pangolin chinois (*M. pentadactyla*) ne possèdent également que *CHIA5* et ont perdu *CHIA1-4* du fait de leur histoire évolutive commune avec les Carnivores (Emerling *et al*, 2018). Dans ce contexte, des questions se posaient sur la manière dont ces paralogues ont été impliqués dans l'adaptation à la myrmécophagie chez les mammifères et, plus spécifiquement, sur la manière dont les espèces myrmécophages utilisent ces différents répertoires de *CHIAs*. Des gènes de chitinases ont été trouvés exprimés dans les glandes salivaires hypertrophiées et d'autres organes digestifs de *M. javanica* et du fourmilier géant (*Myrmecophaga tridactyla*), avec un niveau d'expression particulièrement élevé du seul gène *CHIA5* fonctionnel chez *M. javanica* (Ma *et al*, 2017 ; 2019 ; Cheng *et al*, 2023).

Pour comprendre comment les mammifères myrmécophages se sont adaptés de manière convergente à la digestion de l'exosquelette chitineux de leurs proies, nous avons étudié l'évolution des gènes de chitinases chez ces espèces en utilisant des approches de

génomique et de transcriptomique comparatives (Allio *et al*, 2023). L'évolution des neuf gènes paralogues de chitinases (*CHIT1, OVGP1, CHI3L1, CHI3L2, CHIA1-5*) a révélé une évolution de ces paralogues vers des fonctions différentes, les *CHIAs* étant impliqués dans la dégradation de la chitine car leur site chitinolytique est intact. Des analyses de transcriptomique comparative de plusieurs organes digestifs et non digestifs de *T. tetradactyla* et *M. javanica* ont mis en évidence l'importance des glandes salivaires dans l'adaptation à un régime insectivore et le fait que les gènes de chitinases sont exprimés différemment chez les espèces myrmécophages. Le seul gène *CHIA5* fonctionnel de *M. javanica* est fortement exprimé dans ses glandes salivaires et ses autres organes digestifs, alors que seuls *CHIA3* et *CHIA4* sont fortement exprimés dans les glandes salivaires et les tissus digestifs de *T. tetradactyla*, tandis que *CHIA1* et *CHIA2* sont exprimés dans le pancréas uniquement. Le bricolage moléculaire de ces répertoires de gènes a donc permis à ces espèces de digérer leurs proies riches en chitine en exprimant différemment leurs gènes de chitinases.

Rôle du microbiote intestinal dans l'adaptation à la myrmécophagie chez les mammifères

Le rôle des symbiontes microbiens dans l'évolution de leur hôte est de plus en plus reconnu car ils coévoluent depuis longtemps. Plus spécifiquement, le microbiote intestinal est impliqué dans de nombreuses fonctions de l'hôte telles que la digestion, l'immunité, ou le comportement (Suzuki, 2017). Plusieurs facteurs tels que le régime alimentaire, la physiologie et la phylogénie de l'hôte, mais aussi des facteurs environnementaux façonnent la composition taxonomique et la structure fonctionnelle du microbiote intestinal (Alessandri *et al*, 2022). Chez les mammifères, le microbiote intestinal a joué un rôle majeur dans la diversification de leurs régimes alimentaires, permettant des transitions vers de nouvelles sources de carbone (Ley *et al*, 2008a ; Moeller et Sanders, 2020). Par conséquent, de nombreuses études se sont concentrées sur les adaptations alimentaires convergentes et les impacts respectifs du régime alimentaire et de la phylogénie de l'hôte sur le microbiote intestinal (*e.g.*, Ley *et al*, 2008b ; Muegge *et al*, 2011 ; Youngblut *et al*, 2019).

De nombreux mammifères ingèrent des proies contenant une grande quantité de chitine, comme les insectes avec leur exosquelette chitineux (Gooday, 1990 ; Rathore et Gupta, 2015). Ces structures chitineuses dures doivent être digérées pour accéder aux nutriments. Des bactéries chitinolytiques ont été identifiées dans le tube digestif de plusieurs espèces,

confirmant leur rôle potentiel dans la digestion (Stevens et Hume, 1998), chez plusieurs espèces de mammifères ayant des régimes alimentaires variés (Šimůnek *et al*, 2001), ainsi que des mammifères ayant des régimes alimentaires riches en chitine (Irulan *et al*, 2011 ; Macdonald *et al*, 2014 ; Sanders *et al*, 2015). Le microbiote intestinal des espèces myrmécophages pourrait ainsi avoir joué un rôle important dans leur adaptation convergente vers ce régime hautement spécialisé. Des analyses taxonomiques basées sur des séquences d'ARNr 16S ont ainsi mis en évidence des convergences dans la composition du microbiote intestinal des espèces myrmécophages (Delsuc *et al*, 2014). Des questions subsistent cependant quant au rôle que jouent ces symbiontes dans la digestion de l'exosquelette chitineux des proies. Des études de métagénomique du microbiote intestinal du pangolin malais et du fourmilier géant ont par ailleurs identifié des chitinases microbiennes associées à des bactéries chitinolytiques spécifiques (Ma *et al*, 2018 ; Cheng *et al*, 2023). Les fonctions assurées par les symbiontes intestinaux chez d'autres espèces myrmécophages n'avaient par contre pas été étudiées jusqu'ici.

Le deuxième chapitre de ma thèse a donc porté sur le rôle du microbiote intestinal dans l'adaptation à la myrmécophagie chez les placentaires. En nous concentrant tout d'abord sur les espèces myrmécophages, nous voulions i) déterminer si des bactéries chitinolytiques sont bien présentes dans leur microbiote intestinal et, si oui, ii) comparer leur répartition entre les espèces myrmécophages pour comprendre si les mêmes espèces bactériennes ont été recrutées de façon convergente chez les différentes espèces hôtes. Vingt-neuf métagénomes intestinaux d'espèces myrmécophages ont été séquencés en combinant un séquençage métagénomique *short*- et *long-reads* pour reconstruire des génomes bactériens de haute qualité dans le but d'identifier directement les chitinases bactériennes portées par ces bactéries (Teullet *et al*, 2023). Afin de pouvoir travailler à partir des échantillons fécaux collectés sur le terrain, un protocole d'extraction d'ADN a été optimisé pour extraire de l'ADN bactérien à haut poids moléculaire à partir de ces échantillons fécaux conservés dans de l'éthanol (96 %) et stockés à -20 °C, en vue d'un séquençage *long-reads* (Magdeleine *et al*, 2021).

Plus de 300 génomes bactériens ont pu être reconstruits et près de 400 séquences de chitinases été identifiées dans 132 de ces génomes, ce qui démontre que des bactéries chitinolytiques sont bien présentes dans le microbiote intestinal des espèces myrmécophages et ont donc le potentiel de digérer leurs proies (Teullet *et al*, 2023). Ces gènes de chitinase

bactériens appartiennent principalement à trois familles bactériennes (*i.e.*, Lachnospiraceae, Actualibacteraceae et Ruminococcaceae). Plusieurs gènes de chitinase ont pu être identifiés dans le même génome bactérien, suggérant que des voies microbiennes complexes de dégradation de la chitine sont impliquées. Les espèces de Xénarthres semblent avoir plus de bactéries chitinolytiques que les pangolins, le protèle ou l'oryctérope, ce qui pourrait refléter des contraintes phylogénétiques ainsi que des influences environnementales. De nombreux génomes bactériens sont partagés entre certaines espèces hôtes, principalement entre espèces proches, soulignant un effet phylogénétique sur la composition du microbiote intestinal, tandis que d'autres sont retrouvés chez presque toutes les espèces, suggérant que ces taxons bactériens pourraient être clés pour l'adaptation à la myrmécophagie (Teullet *et al*, 2023).

Enfin, pour mieux comprendre si ces adaptations au niveau du microbiote intestinal sont convergentes chez les espèces myrmécophages et pour distinguer l'influence de la phylogénie par rapport au régime alimentaire, il est crucial de comparer leur microbiote intestinal avec celui d'espèces proches non myrmécophages. Cette analyse a commencé à la fin de ma thèse avec la finalisation du jeu de données et la reconstruction des génomes bactériens à partir de métagénomes intestinaux de placentaires. Ces analyses seront poursuivies dans le but de quantifier si les espèces myrmécophages présentent plus et/ou différentes bactéries chitinolytiques que les espèces non myrmécophages et d'évaluer s'il existe aussi des similitudes avec les espèces ayant un régime alimentaire insectivore moins spécialisé.

Perspectives

Dans l'ensemble, les résultats de cette thèse montrent l'influence à la fois de la contingence historique ainsi que des pressions sélectives pour adopter un régime myrmécophage dans l'évolution convergente des espèces myrmécophages. Des différences dans le degré d'adaptation entre les lignées myrmécophages ayant divergé anciennement (pangolins, fourmiliers : divergence autour de ~80 Mya ; Álvarez-Carretero *et al*, 2022) par rapport aux lignées plus récentes (protèles : divergence autour de ~10 Mya ; Eizirik *et al*, 2010) ont été mis en évidence. Les fourmiliers et les pangolins sont des espèces myrmécophages caractérisées par des adaptations morphologiques et génomiques extrêmes avec de nombreuses pertes de traits phénotypiques. À l'inverse, le protèle présente des adaptations morphologiques au

régime myrmécophage moins marquées et les analyses génomiques n'ont pas révélé de pertes de gènes spécifiquement liées à ce régime alimentaire, ce qui soulève des questions sur le degré de spécialisation de cette espèce vers la consommation quasi-exclusive de termites. L'analyse des données transcriptomiques des organes digestifs du protèle permettrait de déterminer si son seul gène fonctionnel de chitinase (*CHIA5*) est utilisé pour la digestion. De plus, le protèle pourrait potentiellement compter davantage sur son microbiote intestinal pour digérer ses proies que d'autres espèces myrmécophages plus anciennes, car des changements dans la composition du microbiote intestinal peuvent se produire plus facilement et plus rapidement que des adaptations génomiques. Cependant, nos analyses métagénomiques ont uniquement mis en évidence le potentiel de digestion de la chitine par le microbiote intestinal des espèces myrmécophages et n'ont pas montré d'adaptations particulières chez le protèle. Pour bien comprendre le rôle du microbiote intestinal dans la digestion de la chitine, des données métatranscriptomiques seraient nécessaires. En effet, certaines espèces pourraient avoir diverses bactéries chitinolytiques exprimant uniformément leurs gènes de chitinase, alors que chez d'autres espèces avoir peu de bactéries chitinolytiques produisant beaucoup de chitinases aurait pu être sélectionné. La combinaison de données génomiques, métagénomiques, transcriptomiques et métatranscriptomiques permettrait d'évaluer la contribution relative de l'hôte et de son microbiote intestinal (*i.e.*, de l'holobionte) pour la digestion de la chitine.

Ces résultats soulignent donc l'importance de combiner plusieurs approches pour bien comprendre le phénomène de convergence. Pour étudier les mécanismes adaptatifs impliqués dans l'adaptation à la myrmécophagie chez les mammifères placentaires, trois approches différentes et complémentaires ont été développées dans le projet ConvergeAnt : morphologie, génomique et métagénomique. Si le lien génomique-métagénomique a été exploré pendant cette thèse, plusieurs liens restent à explorer entre les différentes parties de ce projet, notamment en intégrant des données morphologiques dans les études génomiques. En effet, les données morphologiques et génomiques peuvent aider à comprendre les mécanismes moléculaires impliqués dans la régression de traits caractéristiques du phénotype myrmécophage tels que la réduction des dents ou des capacités de perception du goût. Ces adaptations peuvent également être explorées grâce à l'étude de l'olfaction. En effet, la taille du répertoire de gènes des récepteurs olfactifs et les structures olfactives peuvent être

corrélées positivement comme cela a été montré chez les placentaires, la seule espèce partiellement myrmécophage (le tatou à neuf bandes) incluse dans cette étude se plaçant en haut de la corrélation (Bird *et al*, 2018). Enfin, les adaptations morphologiques peuvent également être étudiées en lien avec le microbiote, notamment en étudiant les communautés microbiennes le long du tube digestif car la morphologie et physiologie du tube digestif peuvent jouer un rôle dans leur composition et structuration. Une telle analyse pourrait nous aider à mieux comprendre les contributions relatives de l'hôte et de ses symbiontes microbiens à la digestion de la chitine en combinant les informations sur la localisation des bactéries chitinolytiques dans le tube digestif et l'expression des chitinases de l'hôte dans les organes digestifs.

En conclusion de cette thèse, la combinaison des approches de génomique, transcriptomique et métagénomique a permis de mettre en lumière les adaptations génomiques et microbiennes impliquées dans l'adaptation convergente à la myrmécophagie chez les mammifères placentaires en ce qui concerne la perception et la digestion de leurs proies. L'étude de l'histoire évolutive de deux familles de gènes chez les hôtes a mis en évidence les adaptations moléculaires associées aux capacités réduites de perception du goût. Elle a également révélé comment le bricolage moléculaire a permis aux espèces myrmécophages d'utiliser différents répertoires de chitinases, façonnés par des contraintes phylogénétiques, pour digérer leurs proies. L'étude de la présence de bactéries chitinolytiques dans leur microbiote intestinal a mis en évidence le rôle potentiel de leurs symbiontes intestinaux dans la digestion des proies, soulevant des questions sur le rôle de l'holobionte dans l'adaptation à des pressions sélectives spécifiques. Les résultats de cette thèse ont mis en évidence les différents mécanismes adaptatifs impliqués dans l'adaptation convergente à la myrmécophagie. Ensemble, ils soulignent l'importance d’adopter des approches intégratives pour comprendre le phénomène de convergence. Enfin, les études menées au cours de cette thèse soulignent l'influence de la contingence historique (*i.e.*, des contraintes) et de la sélection naturelle (*i.e.*, de l’adaptation) dans l’évolution de phénotypes convergents. L'étude de l'évolution convergente pourrait donc révéler une évolution plus répétable et prédictible qu'on ne le pensait auparavant.

Références

Adler, E., Hoon, M.A., Mueller, K.L., Chandrashekar, J., Ryba, N.J., Zuker, C.S., 2000. A novel family of mammalian taste receptors. Cell 100, 693–702.

Albalat, R., Cañestro, C., 2016. Evolution by gene loss. Nature Reviews Genetics 17, 379–391.

Alessandri, G., Rizzo, S.M., Ossiprandi, M.C., van Sinderen, D., Ventura, M., 2022. Creating an atlas to visualize the biodiversity of the mammalian gut microbiota. Current Opinion in Biotechnology 73, 28–33.

Allio, R., Teullet, S., Lutgen, D., Magdeleine, A., Koual, R., Tilak, M.-K., Thoisy, B. de, Emerling, C.A., Lefébure, T., Delsuc, F., 2023. Comparative transcriptomics reveals divergent paths of chitinase evolution underlying dietary convergence in ant-eating mammals. bioRxiv. https://www.biorxiv.org/content/10.1101/2022.11.29.518312v2. In revision in Genome Biology and Evolution.

Álvarez-Carretero, S., Tamuri, A.U., Battini, M., Nascimento, F.F., Carlisle, E., Asher, R.J., Yang, Z., Donoghue, P.C.J., dos Reis, M., 2022. A Species-Level Timeline of Mammal Evolution Integrating Phylogenomic Data. Nature 602, 263–267.

Arendt, J., Reznick, D., 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends in Ecology & Evolution 23, 26–32.

Bird, D.J., Murphy, W.J., Fox-Rosales, L., Hamid, I., Eagle, R.A., Van Valkenburgh, B., 2018. Olfaction written in bone: cribriform plate size parallels olfactory receptor gene repertoires in Mammalia. Proc Biol Sci 285, 20180100.

Blount, Z.D., Lenski, R.E., Losos, J.B., 2018. Contingency and determinism in evolution: Replaying life's tape. Science 362.

Bussink, A.P., Speijer, D., Aerts, J.M.F.G., Boot, R.G., 2007. Evolution of Mammalian Chitinase(-Like) Members of Family 18 Glycosyl Hydrolases. Genetics 177, 959–970.

Casali, D.M., Martins-Santos, E., Santos, A.L.Q., Miranda, F.R., Mahecha, G.A.B., Perini, F.A., 2017. Morphology of the tongue of Vermilingua (Xenarthra: Pilosa) and evolutionary considerations. Journal of Morphology 278, 1380–1399.

Chandrashekar, J., Mueller, K.L., Hoon, M.A., Adler, E., Feng, L., Guo, W., Zuker, C.S., Ryba, N.J.P., 2000. T2Rs Function as Bitter Taste Receptors. Cell 100, 703–711.

Chandrashekar, J., Hoon, M.A., Ryba, N.J.P., Zuker, C.S., 2006. The receptors and cells for mammalian taste. Nature 444, 288–294.

Chaudhari, N., Roper, S.D., 2010. The cell biology of taste. J Cell Biol 190, 285–296.

Cheng, S.-C., Liu, C.-B., Yao, X.-Q., Hu, J.-Y., Yin, T.-T., Lim, B.K., Chen, W., Wang, G.-D., Zhang, C.-L., Irwin, D.M., Zhang, Z.-G., Zhang, Y.-P., Yu, L., 2023. Hologenomic insights into mammalian adaptations to myrmecophagy. National Science Review 10, nwac174.

Delsuc, F., Metcalf, J.L., Wegener Parfrey, L., Song, S.J., González, A., Knight, R., 2014. Convergence of gut microbiomes in myrmecophagous mammals. Molecular Ecology 23, 1301–1317.

Demuth, J.P., Hahn, M.W., 2009. The life and death of gene families. BioEssays 31, 29–39.

Doran, G.A., Allbrook, D.B., 1973. The tongue and associated structures in two species of African pangolins, *Manis gigantea* and *Manis tricuspis*. J Mammal 54, 887–899.

Dyer, J., Salmon, K.S.H., Zibrik, L., Shirazi-Beechey, S.P., 2005. Expression of sweet taste receptors of the T1R family in the intestinal tract and enteroendocrine cells. Biochem Soc Trans 33, 302–305.

Emerling, C.A., Delsuc, F., Nachman, M.W., 2018. Chitinase genes (*CHIAs*) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. Science Advances 4, eaar6478.

Eizirik, E., Murphy, W.J., Koepfli, K.-P., Johnson, W.E., Dragoo, J.W., Wayne, R.K., O'Brien, S.J., 2010. Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. Molecular Phylogenetics and Evolution 56, 49–63.

Feng, P., Zheng, J., Rossiter, S.J., Wang, D., Zhao, H., 2014. Massive Losses of Taste Receptor Genes in Toothed and Baleen Whales. Genome Biology and Evolution 6, 1254–1265.

Ferreira-Cardoso, S., Delsuc, F., Hautier, L., 2019. Evolutionary Tinkering of the Mandibular Canal Linked to Convergent Regression of Teeth in Placental Mammals. Current Biology 29, 468-475.e3.

Ferreira-Cardoso, S., Fabre, P.-H., Thoisy, B. de, Delsuc, F., Hautier, L., 2020. Comparative masticatory myology in anteaters and its implications for interpreting morphological convergence in myrmecophagous placentals. PeerJ 8, e9690.

Ferreira-Cardoso, S., Claude, J., Goswami, A., Delsuc, F., Hautier, L., 2022. Flexible conservatism in the skull modularity of convergently evolved myrmecophagous placental mammals. BMC Ecology and Evolution 22, 87.

Funkhouser, J.D., Aronson, N.N., 2007. Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. BMC Evol Biol 7, 96.

Garland, K. (2018). Evolution of bitter taste receptors in convergent myrmecophagous placentals. Master thesis. University of Montpellier.

Gooday, G.W., 1990. The Ecology of Chitin Degradation, in: Marshall, K.C. (Ed.), Advances in Microbial Ecology, Advances in Microbial Ecology. Springer US, Boston, MA, pp. 387–430.

Griffiths, M., 1968. Echidnas. Pergamon Press, Oxford; New York.

Hong, W., Zhao, H., 2014. Vampire bats exhibit evolutionary reduction of bitter taste receptor genes common to other bats. Proceedings of the Royal Society B: Biological Sciences 281, 20141079.

Hoon, M.A., Adler, E., Lindemeier, J., Battey, J.F., Ryba, N.J.P., Zuker, C.S., 1999. Putative Mammalian Taste Receptors: A Class of Taste-Specific GPCRs with Distinct Topographic Selectivity. Cell 96, 541–551.

Hu, L.-L., Shi, P., 2013. Smallest bitter taste receptor (T2Rs) gene repertoire in carnivores. Zoological Research 34, 13.

Huang, A.L., Chen, X., Hoon, M.A., Chandrashekar, J., Guo, W., Tränkner, D., Ryba, N.J.P., Zuker, C.S., 2006. The cells and logic for mammalian sour taste detection. Nature 442, 934–938.

Hussain, M., Wilson, J.B., 2013. New Paralogues and Revised Time Line in the Expansion of the Vertebrate GH18 Family. J Mol Evol 76, 240–260.

Irulan, A., Nathan, P.T., Priya, Y.S., Marimuthu, S., Elangovan, V., 2011. Isolation and characterization of chitinase producing gut microflora of insectivorous bats. Trends in Biosciences 4, 8–11.

Jeuniaux, C., 1971. On some biochemical aspects of regressive evolution in animals. E. Schoffeniels. North-Holland Publishing Company.

Kataoka, S., Yang, R., Ishimaru, Y., Matsunami, H., Sévigny, J., Kinnamon, J.C., Finger, T.E., 2008. The Candidate Sour Taste Receptor, *PKD2L1*, Is Expressed by Type III Taste Cells in the Mouse. Chemical Senses 33, 243–254.

Koonin, E.V., 2005. Orthologs, Paralogs, and Evolutionary Genomics. Annu. Rev. Genet. 39, 309–338.

Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R., Gordon, J.I., 2008a. Worlds within worlds: evolution of the vertebrate gut microbiota. Nature Reviews Microbiology 6, 776–788.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., Gordon, J.I., 2008b. Evolution of Mammals and Their Gut Microbes. Science 320, 1647–1651.

Liu, Z., Liu, G., Hailer, F., Orozco-terWengel, P., Tan, X., Tian, J., Yan, Z., Zhang, B., Li, M., 2016. Dietary specialization drives multiple independent losses and gains in the bitter taste gene repertoire of Laurasiatherian Mammals. Frontiers in Zoology 13, 28.

Losos, J.B., 2011. Convergence, adaptation, and constraint. Evolution 65, 1827–1840.

Ma, J.-E., Li, L.-M., Jiang, H.-Y., Zhang, X.-J., Li, J., Li, G.-Y., Yuan, L.-H., Wu, J., Chen, J.-P., 2017. Transcriptomic analysis identifies genes and pathways related to myrmecophagy in the Malayan pangolin (*Manis javanica*). PeerJ 5, e4140.

Ma, J.-E., Jiang, H.-Y., Li, L.-M., Zhang, X.-J., Li, G.-Y., Li, H.-M., Jin, X.-J., Chen, J.-P., 2018. The Fecal Metagenomics of Malayan Pangolins Identifies an Extensive Adaptation to Myrmecophagy. Front. Microbiol. 9, 2793.

Ma, J.-E., Jiang, H.-Y., Li, L.-M., Zhang, X.-J., Li, H.-M., Li, G.-Y., Mo, D.-Y., Chen, J.-P., 2019. SMRT sequencing of the full-length transcriptome of the Sunda pangolin (*Manis javanica*). Gene 692, 208–216.

Macdonald, C., Barden, S., Foley, S., 2014. Isolation and characterization of chitin-degrading micro-organisms from the faeces of Goeldi's monkey, *Callimico goeldii*. Journal of Applied Microbiology 116, 52–59.

Magdeleine, A., Tilak, M.-K., Teullet, S., Delsuc, F., 2021. High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing. protocol.io. dx.doi.org/10.17504/protocols.io.bvyin7ue.

McNab, B.K., 1984. Physiological convergence amongst ant-eating and termite-eating mammals. Journal of Zoology 203, 485–510.

Moeller, A.H., Sanders, J.G., 2020. Roles of the gut microbiota in the adaptive evolution of mammalian species. Philosophical Transactions of the Royal Society B: Biological Sciences 375, 20190597.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R., Gordon, J.I., 2011. Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans. Science 332, 970–974.

Nei, M., Rooney, A.P., 2005. Concerted and Birth-and-Death Evolution of Multigene Families. Annu. Rev. Genet. 39, 121–152.

Patterson, B., 1975. The Fossil Aardvarks (Mammalia: Tubulidentata). Bull. Mus. Comp. Zool. 147, 185–237.

Rathore, A.S., Gupta, R.D., 2015. Chitinases from Bacteria to Human: Properties, Applications, and Future Perspectives. Enzyme Research 2015, 791907.

Redford, K.H., 1987. Ants and Termites As Food, in: Genoways, H.H. (Ed.), Current Mammalogy. Springer US, Boston, MA, pp. 349–399.

Reiss, K.Z., 2001. Using Phylogenies to Study Convergence: The Case of the Ant-Eating Mammals. Integr Comp Biol 41, 507–525.

Sanders, J.G., Beichman, A.C., Roman, J., Scott, J.J., Emerson, D., McCarthy, J.J., Girguis, P.R., 2015. Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. Nature Communications 6, 8285.

Shan, L., Wu, Q., Wang, L., Zhang, L., Wei, F., 2018. Lineage-specific evolution of bitter taste receptor genes in the giant and red pandas implies dietary adaptation. Integrative Zoology 13, 152–159.

Sharma, V., Hecker, N., Roscito, J.G., Foerster, L., Langer, B.E., Hiller, M., 2018. A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. Nat Commun 9, 1–9.

Šimůnek, J., Hodrová, B., Bartoňová, H., Kopečný, J., 2001. Chitinolytic bacteria of the mammal digestive tract. Folia Microbiol 46, 76–78.

Stevens, C.E., Hume, I.D., 1998. Contributions of Microbes in Vertebrate Gastrointestinal Tract to Production and Conservation of Nutrients. Physiological Reviews 78, 393–427.

Suzuki, T.A., 2017. Links between Natural Variation in the Microbiome and Host Fitness in Wild Mammals. Integrative and Comparative Biology 57, 756–769.

Teullet, S., Tilak, M.-K., Magdeleine, A., Schaub, R., Weyer, N.M., Panaino, W., Fuller, A., Loughry, W.J., Avenant, N.L., Thoisy, B. de, Borrel, G., Delsuc, F., 2023. Metagenomics uncovers dietary adaptations for chitin digestion in the gut microbiota of convergent myrmecophagous mammals. mSystems 8, e00388-23.

Wu, V., Rozengurt, N., Yang, M., Young, S., Sinnett-Smith, J., Rozengurt, E., 2002. Expression of bitter taste receptors of the T2R family in gastrointestinal tract and enteroendocrine STC-1 cells. Proceedings of the National Academy of Sciences of the United States of America 99, 2392–7.

Youngblut, N.D., Reischer, G.H., Walters, W., Schuster, N., Walzer, C., Stalder, G., Ley, R.E., Farnleitner, A.H., 2019. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. Nature Communications 10, 2200.

Abstract

Adaptation to myrmecophagy (ant and/or termite consumption) in placental mammals represents a textbook example of convergent evolution, which is the independent evolution of similar phenotypes in different lineages. Myrmecophagous species evolved at least five times independently in placentals and include aardvarks (Tubulidentata), armadillos (Cingulata), anteaters (Pilosa), pangolins (Pholidota), and aardwolves (Carnivora). To understand how these species convergently adapted to this diet and whether the same mechanisms were involved, this project relied on comparative genomics, and metagenomics. More precisely, to decipher the underlying adaptive mechanisms involved in prey perception and digestion, two gene families (respectively taste receptors and chitinases) were studied, the latter being analyzed both in the host and in its associated gut microbiota. Investigating the evolution of taste receptor genes revealed several losses of umami, sweet and/or sour taste perception in myrmecophagous lineages. Comparative transcriptomics highlighted different uses of chitinase gene repertoires for prey digestion in these species reflecting the influence of phylogenetic constraints. Finally, reconstructing chitinolytic bacterial genomes from the gut microbiota of nine myrmecophagous species uncovered its role in the adaptation to this specialized diet raising questions on the role of the holobiont in adapting to myrmecophagy. Altogether these results emphasize the importance to integrate several approaches to better understand the phenomenon of convergence and further shed light on the influence of both historical contingency and natural selection in shaping convergent phenotypes.

Résumé

L'adaptation à la myrmécophagie (consommation de fourmis et/ou de termites) chez les mammifères placentaires représente un exemple classique d'évolution convergente qui correspond à l'évolution indépendante de phénotypes similaires dans différentes lignées. Les espèces myrmécophages ont évolué au moins cinq fois indépendamment chez les placentaires et comprennent les oryctéropes (Tubulidentata), les tatous (Cingulata), les fourmiliers (Pilosa), les pangolins (Pholidota) et les protèles (Carnivora). Pour comprendre comment ces espèces se sont adaptées de manière convergente à ce régime alimentaire et si les mêmes mécanismes sont impliqués, ce projet s'est appuyé sur des approches de génomique et métagénomique comparatives. Plus précisément, pour comprendre les mécanismes adaptatifs sous-jacents impliqués dans la perception et la digestion des proies, deux familles de gènes (respectivement les récepteurs gustatifs et les chitinases) ont été étudiées, ces dernières étant analysées à la fois chez l'hôte et dans son microbiote intestinal. L'étude de l'évolution des gènes des récepteurs gustatifs a révélé plusieurs pertes de la perception des goûts umami, sucré et/ou acide dans les lignées myrmécophages. Des analyses de transcriptomique comparative ont révélé différentes utilisations des répertoires de gènes de chitinase pour la digestion de la chitine chez ces espèces, reflétant l'influence des contraintes phylogénétiques. Enfin, la reconstruction de génomes de bactéries chitinolytiques à partir du microbiote intestinal de neuf espèces myrmécophages a mis en évidence son rôle dans l'adaptation à ce régime alimentaire, soulevant des questions sur le rôle de l'holobionte dans l'adaptation à la myrmécophagie. L'ensemble de ces résultats soulignent l'importance de combiner plusieurs approches pour mieux comprendre le phénomène de convergence et soulignent encore un peu plus l'influence de la contingence historique et de la sélection naturelle dans l'évolution de phénotypes convergents.