



Université de Montpellier

Habilitation à Diriger des Recherches

Ecole doctorale n°584: GAIA

Génomique évolutive et stratégies d'histoires de vie des animaux

présentée par

Jonathan Romiguier

« Je déclare avoir respecté, dans la conception et la rédaction de ce mémoire d'HDR, les valeurs et principes d'intégrité scientifique destinés à garantir le caractère honnête et scientifiquement rigoureux de tout travail de recherche, visés à l'article L.211-2 du Code de la recherche et énoncés par la Charte nationale de déontologie des métiers de la recherche et la Charte d'intégrité scientifique de l'Université de Montpellier. Je m'engage à les promouvoir dans le cadre de mes activités futures d'encadrement de recherche. »

| | |
|--|----|
| <u>Introduction</u> | 5 |
| <u>Chapitre 1 Activité de recherche</u> | 7 |
| <u>1.1. Liens entre évolution moléculaire et stratégie d'histoire de vie</u> | 8 |
| <u>1.1.1.Taux de GC et traits d'histoire de vie</u> | 8 |
| <u>1.1.2. Efficacité de la sélection purifiante et traits d'histoire de vie</u> | 9 |
| <u>1.1.3. Reconstruction ancestrale de traits d'histoire de vie</u> | 12 |
| <u>1.1.4. Efficacité de la sélection purifiante et eusocialité</u> | 14 |
| <u>1.1.5.Diversité génétique des animaux et traits d'histoire de vie</u> | 19 |
| <u>1.2 Phylogénomique et évolution de nouvelles stratégies d'histoire de vie</u> | 23 |
| <u>1.2.1 Phylogénomique et racine des mammifères placentaires</u> | 23 |
| <u>1.2.2. Phylogénomique et origine de l'eusocialité chez les abeilles corbiculées</u> | 24 |
| <u>1.2.3. Phylogénomique du genre <i>Formica</i> et évolution de l'esclavagisme</u> | 26 |
| <u>1.2.4. Phylogénomique et évolution moléculaire des Formicidae</u> | 27 |
| <u>1.3. Evolution vers une stratégie reproductive atypique : l'hybridogénèse sociale</u> | 31 |
| <u>1.3.1. Déterminisme de caste : contexte</u> | 31 |
| <u>1.3.2. Découverte de l'hybridogénèse sociale chez les fourmis du genre <i>Messor</i></u> | 32 |
| <u>1.3.3. Modèle théorique pour l'origine de l'hybridogénèse sociale</u> | 34 |
| <u>1.3.4. Méthode de détection d'hybride F1 pour scans d'hybridogénèse sociale à large échelle</u> | 37 |
| <u>Chapitre 2 Projet de recherche</u> | 40 |
| <u>2.1. Avant-propos</u> | 41 |
| <u>2.2. Projet RoyalMess : Résumé</u> | 42 |
| <u>2.3. Contexte</u> | 43 |
| <u>2.4. Objectifs</u> | 45 |
| <u>2.5. Méthodes</u> | 47 |
| <u>2.5.1. Objectif 1 - Prévalence de l'hybridogenèse sociale</u> | 47 |
| <u>2.5.2. Histoire des lignées hybridogénétiques</u> | 51 |
| <u>2.5.3. Objectif 3 - Identifier les gènes déterminant la caste</u> | 55 |
| <u>2.5.3.1 Objectif 3a : Expression différentielle embryonnaire</u> | 56 |
| <u>2.5.3.2. Objectif 3b : accouplement des mâles recombinés inter-lignées</u> | 58 |
| <u>2.5.3.3. Objectif 3c : Edition de gènes par CRISPR-Cas9</u> | 60 |
| <u>Chapitre 3 Curriculum Vitae</u> | 63 |
| <u>Chapitre 4 Bibliographie</u> | 71 |
| <u>Annexe Sélection d'articles</u> | 81 |

Introduction

Les premiers axes de recherche de ma carrière scientifique ont débuté lors de ma thèse de doctorat, en 2009. Cette dernière a commencé à éveiller mon intérêt pour une meilleure compréhension des processus reliant évolution des génomes, évolution des espèces et évolution des traits d'histoire de vie. Plus particulièrement, je me suis focalisé sur la classe des Mammifères. Afin d'aborder ces questions, j'ai utilisé au cours de ce travail l'outil de la génomique comparative pour analyser les variations d'efficacité de la sélection naturelle et les variations de compositions nucléotidiques. Cette analyse de près d'une quarantaine de génomes a mis en évidence une série de résultats quant à l'origine et l'évolution moléculaire, phénotypique et phylogénétique des mammifères. C'est à ce moment que j'ai commencé à initier deux des axes de recherches que j'ai ensuite développé tout au long de ma carrière :

1) Liens entre évolution moléculaire et stratégie d'histoire de vie.

Au cours de mes travaux que je détaille plus bas, j'ai démontré à de nombreuses reprises que l'évolution des génomes (mesure de taux de sélection naturelle, composition nucléotidique ou encore diversité génétique) est contrainte par la stratégie d'histoire de vie des espèces (taille, longévité, fécondité, taille de propagule...). Ces liens sont des preuves empiriques de plusieurs attendus de la théorie de la génétique des population en lien avec la taille efficace de population, et peuvent donc nous permettre de mieux comprendre les fluctuations phénotypiques et démographiques des espèces tout au long de plusieurs échelles de temps.

2) Phylogénomique et évolution de nouvelles stratégies d'histoire de vie.

Pour mieux comprendre l'évolution de nouvelles stratégies d'histoire de vie telle que le mode de vie placentaire ou l'eusocialité, reconstruire précisément les relations de parentés entre espèce est une étape incontournable. Étant donné la grande quantité de données génomiques de plus en plus disponible, je me suis intéressé à leur capacité pour résoudre des questions évolutives encore en débat. Si la quantité d'information contenu dans un génome offre une grande puissance statistique quand elle est associée aux dernières méthodes probabilistes de phylogénie moléculaire, cela ne va pas sans biais associés qu'il convient de considérer. Tout en m'intéressant à ces biais, je me suis intéressé à plusieurs relations de parentés controversées qui permettent de jeter un regard nouveau sur l'évolution de nouvelles stratégies d'histoire de vie.

Au delà des mammifères étudiés au cours de ma thèse, j'ai développé ces deux axes de recherche sur d'autres modèles et échelles évolutives lors de mes post-doctorats :

métazoaires, oiseaux, abeilles, téléostéens et plus particulièrement fourmis. Depuis ma prise de poste en tant que chargé de recherche CNRS en 2018, j'ai continué à travailler sur l'eusocialité en me focalisant plus particulièrement sur la fourmis moissonneuse barbare *Messor barbarus*, ce qui m'a permis d'initier un troisième axe de recherche :

3) Evolution vers une stratégie reproductive atypique : l'hybridogénèse sociale.

L'eusocialité est l'une des stratégies d'histoire de vie les plus particulières du monde vivant. Le lien entre génotype et traits d'histoire de vie y est en effet très intrigant, avec un même génotype pouvant donner naissance à deux phénotypes aux traits d'histoire de vie radicalement différents : l'ouvrière stérile à la durée de vie qui excède rarement les 3 ans, ou la reine hyper-fertile à la longévité extrême pouvant atteindre les 30 ans. Si ce déterminisme de caste a longtemps été pensé uniquement environnemental, des systèmes de reproduction atypiques démontrent que ce déterminisme peut être largement génétique. J'ai utilisé plusieurs approches théoriques et empiriques via des outils de génétique des populations, phylogénie et génomique comparative pour mieux comprendre l'ampleur et l'évolution de ces systèmes de reproduction fascinants.

C'est principalement dans la continuité de ce 3ème axe de recherche que s'articule mon projet de recherche principal, centré sur la **génomique évolutive de la royauté chez les fourmis hybridogénétiques du genre *Messor*.**

Dans un premier chapitre, je présente plus en détail ces trois axes de recherches. Le deuxième chapitre sera consacré à mon projet de recherche, où je présenterai en détail le projet ERC *Starting Grant* dont je suis actuellement porteur, et bien plus brièvement le résumé de trois ANR sur lesquelles je participe en tant que collaborateur. Enfin, le 3ème chapitre sera consacré à mon CV détaillé, avec la listage de mes publications et activités d'encadrement d'étudiants, doctorants et post-doctorants.

Chapitre 1

Activité de recherche

1.1. Liens entre évolution moléculaire et stratégie d'histoire de vie

Séquencer un génome n'a jamais été aussi simple. Pourtant, et malgré la facilité croissante avec laquelle on peut obtenir le génome de presque n'importe quelle espèce vivante, comprendre et donner un sens à ces données toujours plus complexes reste un défi majeur. Pour donner un sens aux millions de bases d'un génome d'espèce non-modèle, l'un des plus grands obstacles est le manque de contexte évolutif et écologique. À travers des approches de génomique comparative reposant sur des méthodes de phylogénie moléculaire et de génétique des populations, j'essaie de mieux comprendre comment et pourquoi les génomes évoluent de manière différente entre espèces. Plus particulièrement, je m'intéresse au lien entre évolution des génomes et biologie des espèces. Si l'évolution du phénotype est intrinsèquement codée par des mutations au niveau moléculaire, l'évolution des génomes est elle aussi conditionnée par les caractéristiques phénotypiques et écologiques d'une espèce. L'évolution moléculaire est en effet gouvernée par plusieurs paramètres issus de la théorie de la génétique des populations (mutation, sélection, dérive, recombinaison), paramètres qui dépendent à leur tour de variables telles que la masse corporelle, la longévité, la maturité sexuelle ou d'autres traits d'histoire de vie.

1.1.1. Taux de GC et traits d'histoire de vie

Que ce soit entre régions génomiques d'un même individu ou entre espèces, la composition nucléotique d'un génome peut varier. Chez les mammifères, cette composition en base est particulièrement hétérogène entre espèces, une hétérogénéité due à certaines régions plus riches qu'attendu en nucléotides G et C (Duret *et al.* 2002). Nous avons mis en évidence une relation entre composition nucléotidique et biologie des espèces. Dans un article publié dans *Genome Research*, nous avons en effet démontré que le taux de GC moyen des mammifères est inversement corrélé avec divers traits d'histoire de vie tels que la longévité ou la masse corporelle (**Romiguier *et al.* 2010**).

La raison de cet énigmatique lien est expliquée par un distorsion majeur de notre évolution moléculaire : la **conversion génique biaisée** (Duret et Galtier, 2009). Ce biais de réparation, intervenant au cours de la recombinaison méiotique, avantage les nucléotides G et C au détriment des nucléotides A et T (Fig. 1). En truquant ainsi la loterie génétique de la méiose, les allèles portant un G ou un C se retrouvent sur-représentés parmi les gamètes produits,

favorisant la dispersion et la fixation de bases GC au sein des points chauds de recombinaison de nos génomes.

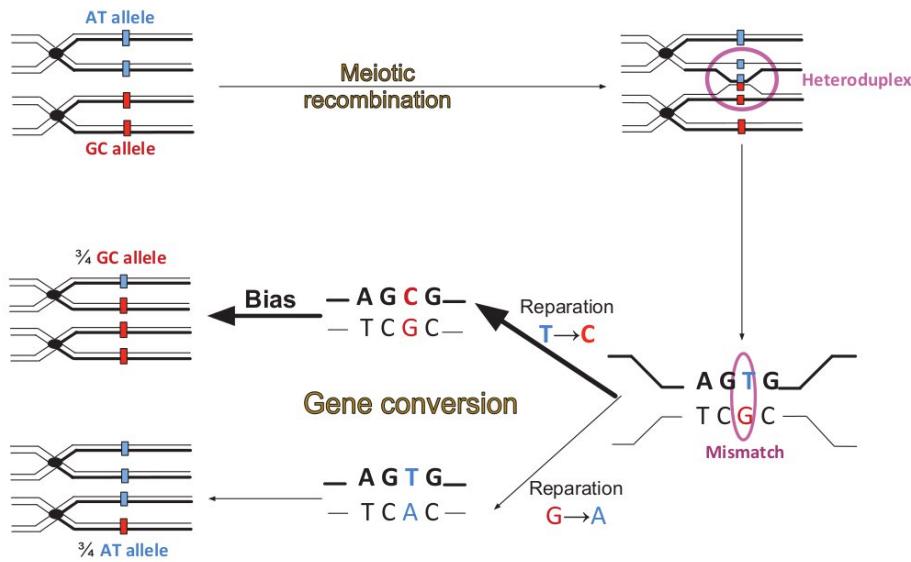


Figure 1 : La conversion génique biaisée

Parce que les espèces peu longévives ou de petites tailles ont des **temps de génération** plus courts, elles expérimentent plus d'événements de recombinaison méiotique pour un même pas de temps donné. Ces événements de recombinaison plus fréquents entraînent plus d'événements de conversion génique biaisée et de hausse de taux de GC. En moyenne, les espèces de petite taille sont donc plus riches en GC.

En mettant en lien mécanismes moléculaires et variables macroscopiques, cette étude souligne le rôle majeur de la conversion génique biaisée dans la composition nucléotidique de nos génomes. C'est lors de plusieurs collaborations intervenant après ma thèse de doctorat que j'ai aidé à généraliser ce constat pionnier en la matière au sein d'autres groupes que celui des mammifères, comme celui des oiseaux, reptiles ou reste des vertébrés (**Weber et al. 2014b; Figuet et al. 2015**).

1.1.2. Efficacité de la sélection purifiante et traits d'histoire de vie

L'immense majorité de nos gènes sont sous sélection purifiante, une forme de sélection qui élimine les mutations délétères et maintient l'intégrité fonctionnelle des protéines (Ohta 1992). L'efficacité de la sélection purifiante est cependant variable. Comme pour toutes les formes de sélection, elle est largement influencée par la **taille efficace de population** : en faible taille efficace, les effets de la sélection sont contrebalancés par les effets stochastiques

de la **dérive génétique**. Ainsi, l'efficacité globale de la sélection purifiante dépend de la taille efficace (Ne) de l'espèce considérée, taille efficace elle même influencée par des traits d'histoire de vie intimement liés à l'abondance (taille, fécondité, longévité, temps de génération...). Chez les mammifères, les espèces longévives ont de faibles tailles efficaces de population et de fait une efficacité globale de la sélection purifiante plus faible (Popadin *et al.* 2007; Nikolaev *et al.* 2007; Romiguier *et al.* 2012).

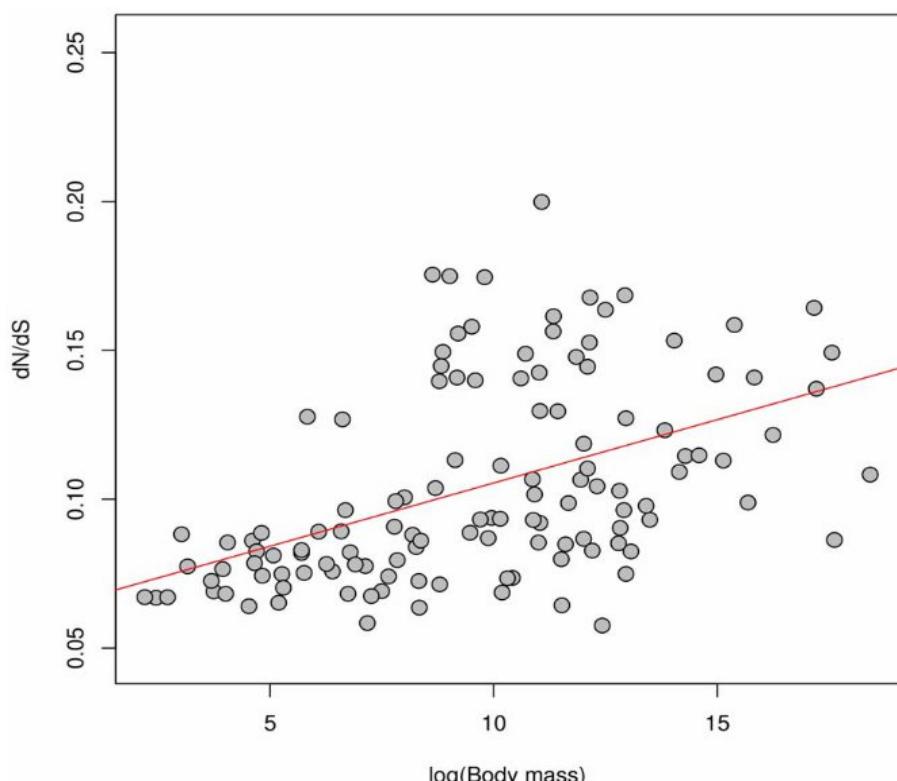


Figure 2 : Corrélation entre ratios de dN/dS des branches terminales et masse corporelle de 139 mammifères (Romiguier *et al.* 2012)

L'efficacité de la sélection purifiante d'une espèce est classiquement estimée via le ratio de dN/dS (taux de substitutions non-synonymes sur taux de substitutions synonymes) moyenné sur l'ensemble des gènes. En faible taille efficace de population, le dN/dS est élevé parce que plus de mutations faiblement délétères (= la plupart des mutations non-synonymes) se sont fixées par dérive dans la population, signe d'une sélection purifiante moins efficace. En pratique, les méthodes traditionnelles pour estimer le ratio de dN/dS de plusieurs espèces sont très coûteuses en temps de calcul. Ces méthodes sont dites probabilistes, et sont similaires à celles utilisées pour reconstruire un arbre phylogénétique dans un cadre statistique de maximum de vraisemblance ou bayésien. La méthode la plus utilisée consiste à estimer une valeur de paramètre correspondant au dN/dS pour chaque branche d'un arbre

phylogénétique : on parle de modèles non-homogènes - différentes valeurs de paramètres selon les branches (Yang *et al.* 1998).

Parce qu'ils sont très riches en paramètres, les modèles non-homogènes sont très coûteux en temps de calcul et en ressources informatiques. Ce coût est une limitation technique qui devient vite insoluble lorsqu'on cherche à utiliser ces méthodes sur des jeux de données génomiques riches en espèces. Adapter les méthodes d'estimation du dN/dS à l'ère de la génomique comparative a ainsi constitué une partie nécessaire de ma thèse de doctorat. Pour cela, j'ai testé la robustesse d'une méthode alternative qui accélère considérablement les estimations (**Romiguier *et al.* 2012**). Cette méthode dite de "cartographie des substitutions" (*substitution mapping*) se base sur l'optimisation d'un modèle homogène (une seule valeur de dN/dS pour toutes les branches de l'arbre) pour reconstruire l'histoire évolutive de chaque site d'un alignement, puis fait le décompte des substitutions de type synonyme et non-synonyme le long de chaque branche (**Romiguier *et al.* 2012**). Bien que simplificatrice, cette méthode évite les problèmes de sur-paramétrisation, et a l'avantage d'être à la fois plus précise et jusqu'à 25 000 fois plus rapide que les méthodes à modèle non-homogène.

Mieux adaptée aux défis de l'ère génomique, cette approche a permis de confirmer et renforcer les relations entre efficacité de la sélection purifiante (mesurée via le ratio de dN/dS) et taille efficace de population des mammifères (approximée via la masse corporelle ou la longévité d'une espèce). Par la suite, j'ai utilisé la même approche dans le cadre d'une collaboration pour explorer des questions similaires chez les oiseaux (**Weber *et al.* 2014a**).

J'ai également plus tard supervisé un travail similaire chez les télostéen afin de tester si ce type de corrélation était valable au sein d'un groupe d'animaux aquatique (**Rolland *et al.* 2020**). À l'aide d'un vaste jeu de données génomique de 76 espèces de poissons téléostéens, nous montrons que les espèces dont les traits d'histoire de vie sont associés à une forte vulnérabilité à la pêche (grande taille corporelle, longévité élevée, fécondité peu élevée) présentent un taux accru d'accumulation de mutations délétères (dN/dS élevé). Nos résultats, qui portent sur un grand clade d'espèces aquatiques, généralisent les modèles précédents trouvés jusqu'à présent dans quelques clades de vertébrés terrestres. Ces résultats montrent également que les espèces vulnérables à la pêche accumulent intrinsèquement plus de substitutions délétères que les espèces non menacées, ce qui illustre les liens potentiels entre la génétique des populations, l'écologie et les politiques de pêche pour prévenir l'extinction des espèces.

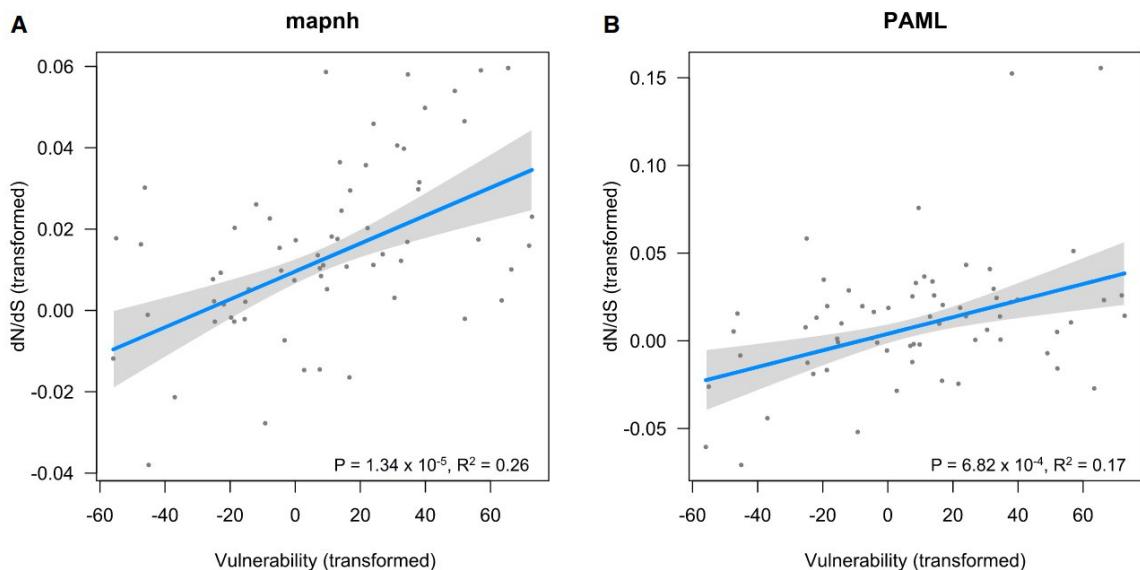


Figure 3 : Le ratio dN/dS est associé à la vulnérabilité. Le ratio de dN/dS a été calculé en utilisant soit (A) mapnh ou (B) PAML. Les lignes de régression, les p-values et le R^2 des régressions après contrôle pour l'inertie phylogénétique sont représentés sur chaque panneau pour toutes les branches terminales de plus de 10 000 substitutions synonymes.

La partie grisée représente les intervalles de confiance à 95 %. Les axes x et y ont été transformés afin de tenir compte de la parenté phylogénétique entre les variables. (Rolland et al. 2020).

1.1.3. Reconstruction ancestrale de traits d'histoire de vie

Il est communément admis que nos ancêtres du Crétacé étaient de minuscules animaux semblables aux rongeurs actuels, vivant dans l'ombre des dinosaures et leur ayant survécu après la 6ème extinction de masse, il y a de cela 65 millions d'années. Bien que profondément ancré dans l'imaginaire collectif, cette image d'épinal n'est supportée par aucune donnée paléontologique claire : les fossiles de mammifères du Crétacé découverts jusqu'à présent sont externes aux lignées à l'origine des mammifères placentaires actuels.

Comme j'ai notamment pu le montrer lors de mes premiers articles (Romiguier et al. 2010, 2012), les variations de traits d'histoire de vie laissent des marques visibles dans l'évolution des génomes mammaliens. Via des approches phylogénétiques (notamment développées au cours du travail décrit dans la partie 1.2), j'ai reconstruit l'état ancestral de deux aspects de la dynamique des génomes mammaliens : évolution du contenu en GC (Fig. 2) et ratio dN/dS (Fig. 3). En liant ces processus d'évolution moléculaire aux traits d'histoire de vie des espèces modernes, j'ai estimé que les ancêtres des mammifères placentaires avaient une durée de vie de plus de 25 ans et une masse corporelle de plus d'1 kg (Romiguier et al.

2013b). Ces chiffres sont similaires à ceux des Primates ou Carnivores actuels, mais très supérieurs à la masse et la durée de vie des Rongeurs ou des Insectivores. Ces estimations ont remis en question la vision traditionnelle que l'on avait de nos ancêtres directs vivant aux côtés des dinosaures, suggérant que leurs traits d'histoire de vie étaient en réalité bien éloignés de ceux d'une souris ou d'une musaraigne.

L'essentiel de ce travail a été réalisé grâce à la base de donnée OrthoMaM développée au sein de l'équipe et dans laquelle j'ai contribué au développement (Douzery *et al.* 2014). J'ai également participé à la poursuite du développement de ce type d'approche dans le groupe de Cétartiodactyles (Figuet *et al.* 2014).

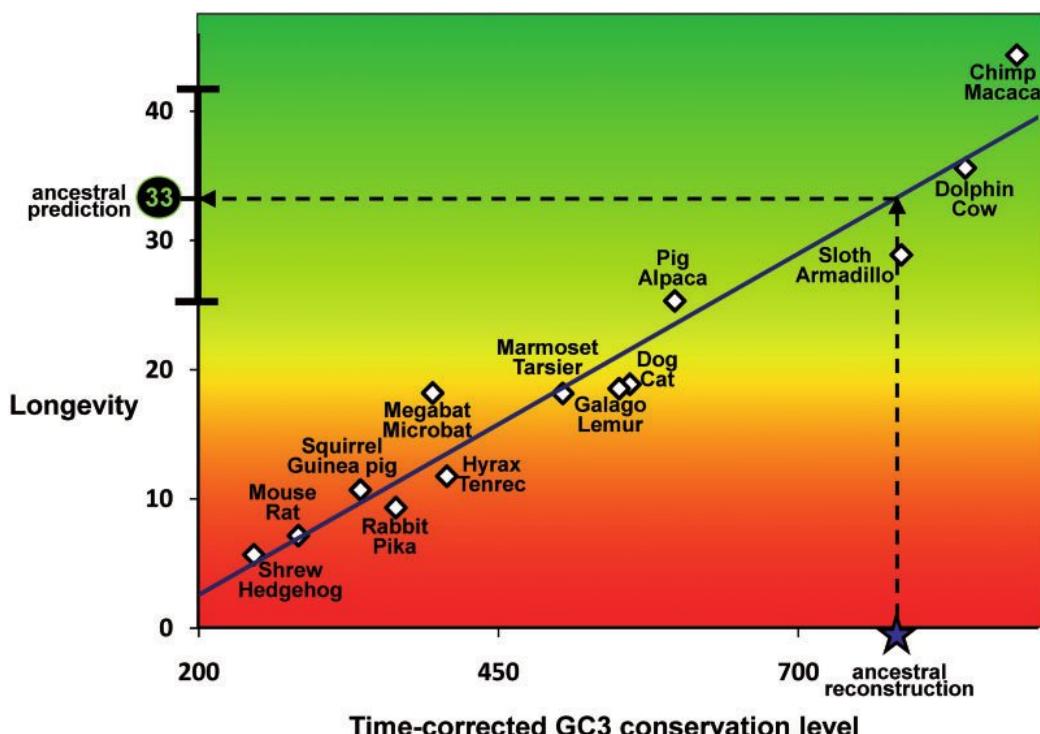


Figure 4 : Conservation des taux de GC à la 3ème position de codons en fonction de la longévité moyenne de 13 couples d'espèces proches (Romiguier *et al.* 2013b).

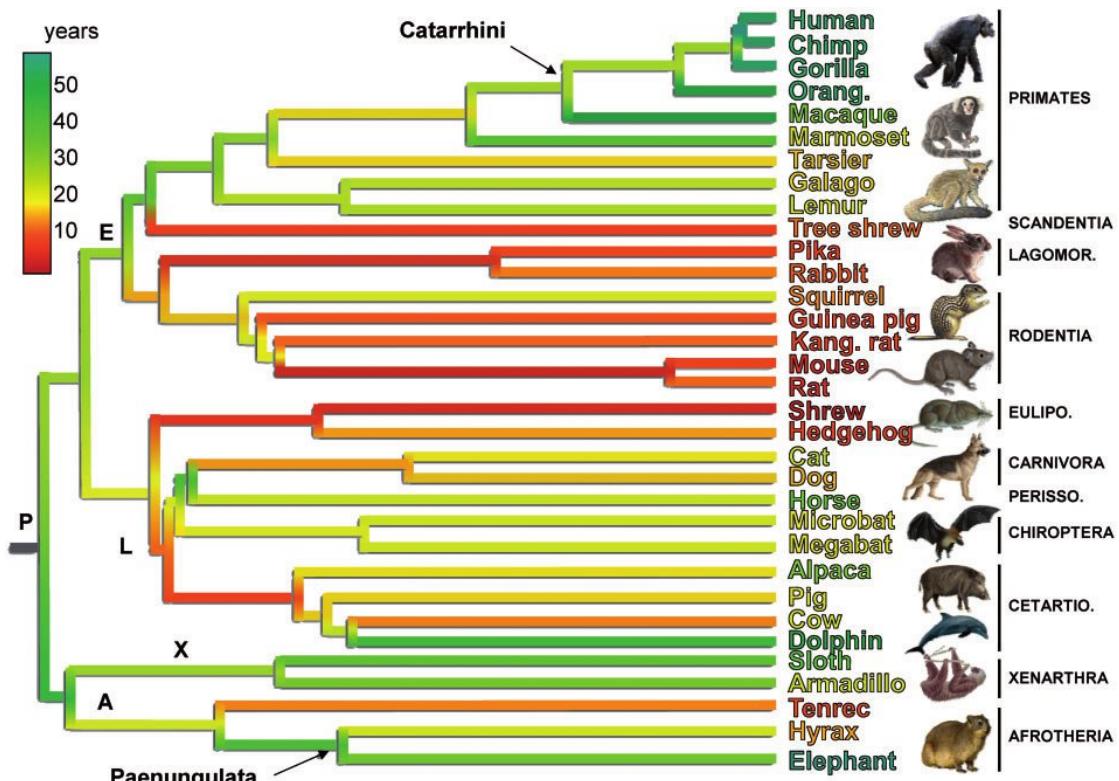


Figure 5 : Reconstructions de longévités ancestrales inférées via le ratio de dN/dS (Romiguier et al. 2013b)

1.1.4. Efficacité de la sélection purifiante et eusocialité

L'évolution de l'eusocialité chez les insectes s'est accompagnée par l'émergence de plusieurs traits d'histoire de vie typiques des grands vertébrés : considérée comme un super-organisme, une colonie peut atteindre une biomasse de plusieurs kilogrammes, commencer à produire des sexués après plusieurs années de maturation et peut avoir une longévité de plusieurs décennies. Ces caractéristiques et le fait que quelques individus monopolisent la reproduction (reines et mâles) réduisent considérablement la taille efficace de population d'un insecte eusocial, une réduction supposée avoir des conséquences et des coûts visibles au niveau moléculaire.

Dans une étude parue dans *Journal of Evolutionary Biology*, nous démontrons que les niveaux de polymorphismes et l'efficacité de la sélection purifiante (que cela soit au niveau intra ou inter-spécifique) des insectes eusociaux (fourmis, abeilles et termites) est plus comparable à ce que l'on peut observer chez les vertébrés (oiseaux, mammifères) que chez les insectes non-sociaux (papillons, moustiques, drosophiles) (Romiguier et al. 2014b). Le mode de vie eusocial des fourmis, abeilles et termites s'accompagne donc d'une taille

efficace de population et d'un fardeau génétique similaire à celui retrouvé chez des vertébrés disposant d'une faible efficacité de sélection purifiante.

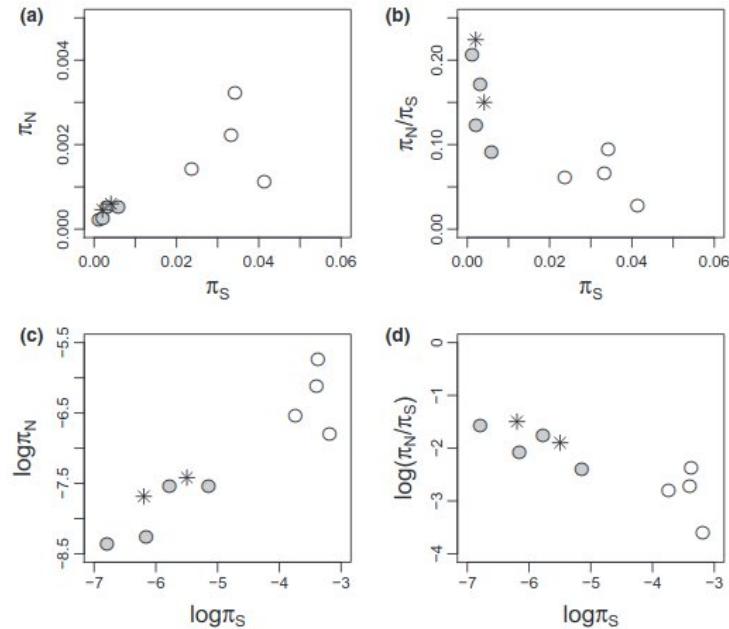


Figure 6 : Patron de génomique des populations d'insectes eusociale vs solitaire. π_S et π_N mesurent respectivement le niveau de polymorphisme synonyme et non-synonyme. Un fort ratio π_N et π_S est ainsi interprété comme mesurant une faible intensité de sélection purifiante et une accumulation accrue de mutations faiblement délétères ségrégant dans la population. Les cercles sombres représentent les insectes sociales (fourmis, termites, abeilles) tandis que les cercles clairs représentent les insectes solitaires (papillon, moustique, drosophile). Les étoiles correspondent à deux espèces de mammifères (lièvre et chimpanzé) et montrent des patrons de diversité génomiques proche des espèces eusociales. (Romiguier et al 2014a)

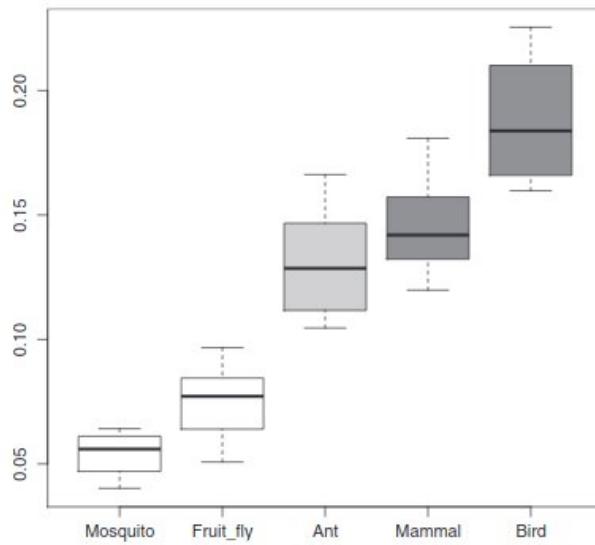


Figure 7 : Distribution de dN/dS dans 5 groupes d'animaux. Les fourmis présentent des niveaux de dN/dS plus comparables à ceux des vertébrés que d'autres insectes.
(Romiguier et al 2014a)

Ces résultats démontrent pour la première fois de manière claire que l'évolution de la vie en société a d'importantes conséquences à la fois au niveau génomique et populationnel. Dans le cadre d'une collaboration, nous avons utilisés des approches populationnelles similaires démontrant l'effet du syndrome insulaire sur l'évolution moléculaire de la tortue de Galapágos, une espèce iconique en danger d'extinction (**Loire et al. 2013**).

Lors de la supervision du travail de thèse d'Arthur Weyna, nous avons également montré que des traces de relâchements de sélection purifiante suggèrent une faible taille efficace de population chez les hyménoptères eusociaux et les abeilles pollinisatrices solitaires. (**Weyna et Romiguier 2021**).

Avec l'un des plus grands nombres d'espèces parasites, eusociales et pollinisatrices parmi tous les ordres d'insectes, les Hyménoptères présentent une grande diversité de modes de vie spécifiques, avec notamment l'évolution répétée de l'eusocialité. Au niveau génétique des populations, on s'attend à ce que de telles stratégies d'histoire de vie diminuent la taille efficace des populations et l'efficacité de la sélection purificatrice. En effet, de part la division du travail reproducteur inhérente à l'eusocialité, une faible proportion des individus peut se reproduire (les reines et mâles) alors que l'immense majorité des individus d'une colonie est stérile ou quasi-stérile (ouvrières). Dans cette étude, nous avons testé cette hypothèse en estimant le taux relatif de substitution non-synonyme chez 169 espèces afin d'étudier la variation de l'efficacité de la sélection naturelle au sein des hyménoptères. Nous n'avons trouvé aucun effet du parasitisme ou de la taille corporelle, mais nous avons détecté que la sélection relâchée est associée à l'eusocialité, ce qui suggère que la division du travail

reproductif diminue la taille efficace de la population chez les fourmis, les abeilles et les guêpes. De manière inattendue, l'effet de l'eusocialité est marginal comparé à un très fort relâchement de la sélection purifiante chez toutes les abeilles, qu'elles soient sociales ou solitaires, ce qui suggère que ces espèces pollinisatrices clés se caractérisent généralement par une faible taille de population efficace. Ce patron suggère des contraintes spécifiques chez les abeilles pollinisatrices, potentiellement liées à des ressources limitées et à un investissement parental élevé. Le fardeau génétique particulièrement élevé de mutations délétères que nous signalons dans le génome de ces espèces ingénieries des écosystèmes soulève également de nouvelles inquiétudes quant au déclin continu de leurs populations.

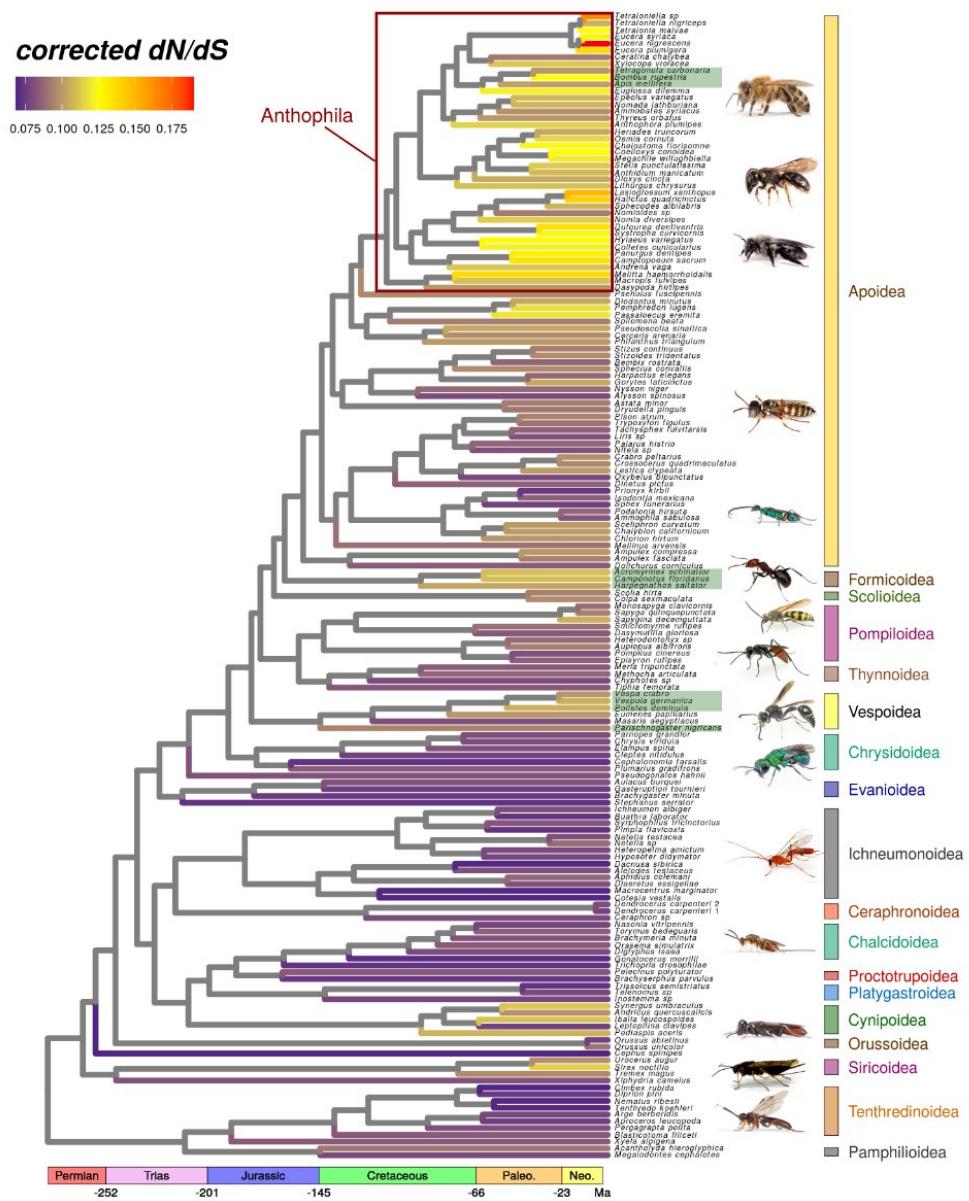


Figure 8 : Ratios de dN/dS génomiques pour 169 espèces d'Hyménoptères. Les ratios de dN/dS sont estimés sur les branches terminales en utilisant 3241 gènes et sont représentés sur un chronogramme déduit d'après Peters et al. (2017). Ils sont corrigés pour éviter de potentiels biais induits par la conversion génique biaisée (substitution AT→GC et GC→AT non prises en compte) Les rectangles verts autour des noms d'espèces indiquent les taxons eusociaux. (Weyna et Romiguier 2021).

1.1.5.Diversité génétique des animaux et traits d'histoire de vie

La diversité génétique est la quantité de variation observée entre les séquences ADN d'individus de la même espèce. Ce concept clé joue un rôle central en biologie des populations et dans les politiques de préservation des espèces. Si la diversité génétique semble grandement varier entre espèces, les déterminants de cette variation étaient jusque là en grande partie méconnus : même les méta-analyses les plus récentes n'étaient pas parvenues à prédire la diversité génétique d'une espèce à partir de quelque variable que ce soit (Leffler *et al.* 2012). Plusieurs facteurs confondants à l'influence relative et entremêlée étaient alors suspectés d'affecter le polymorphisme génétique d'une espèce : sa stratégie d'histoire de vie, ses taux de mutation, sa structure de population ou sa démographie ponctuée par des perturbations écologiques plus ou moins aléatoires dans son histoire récente.

Nous avons démontré que la diversité génétique d'une espèce est pourtant prévisible et largement déterminée par sa stratégie écologique (**Romiguier *et al.* 2014b**). Cette démonstration empirique jusque là inexistante dans la littérature est expliquée par l'ampleur du jeu de données analysé : alors que les études s'intéressant à la question étaient basées sur une couverture génomique très faible (de l'ordre de quelques gènes) ou sur un échantillonnage taxonomique biaisé (sur-représentation des primates), j'ai travaillé sur un jeu de données nouvellement généré regroupant 374 transcriptomes chez 76 espèces animales non-modèles. La distribution du polymorphisme génomique le long de l'arbre des métazoaires (Figure 9) a ainsi permis de mettre en lumière le rôle clé de l'investissement parental : plus de la moitié de la variance en diversité génétique est prédite par la taille du propagule, le stade qui quitte la mère et se disperse (oeuf, larve ou juvénile selon les cas).

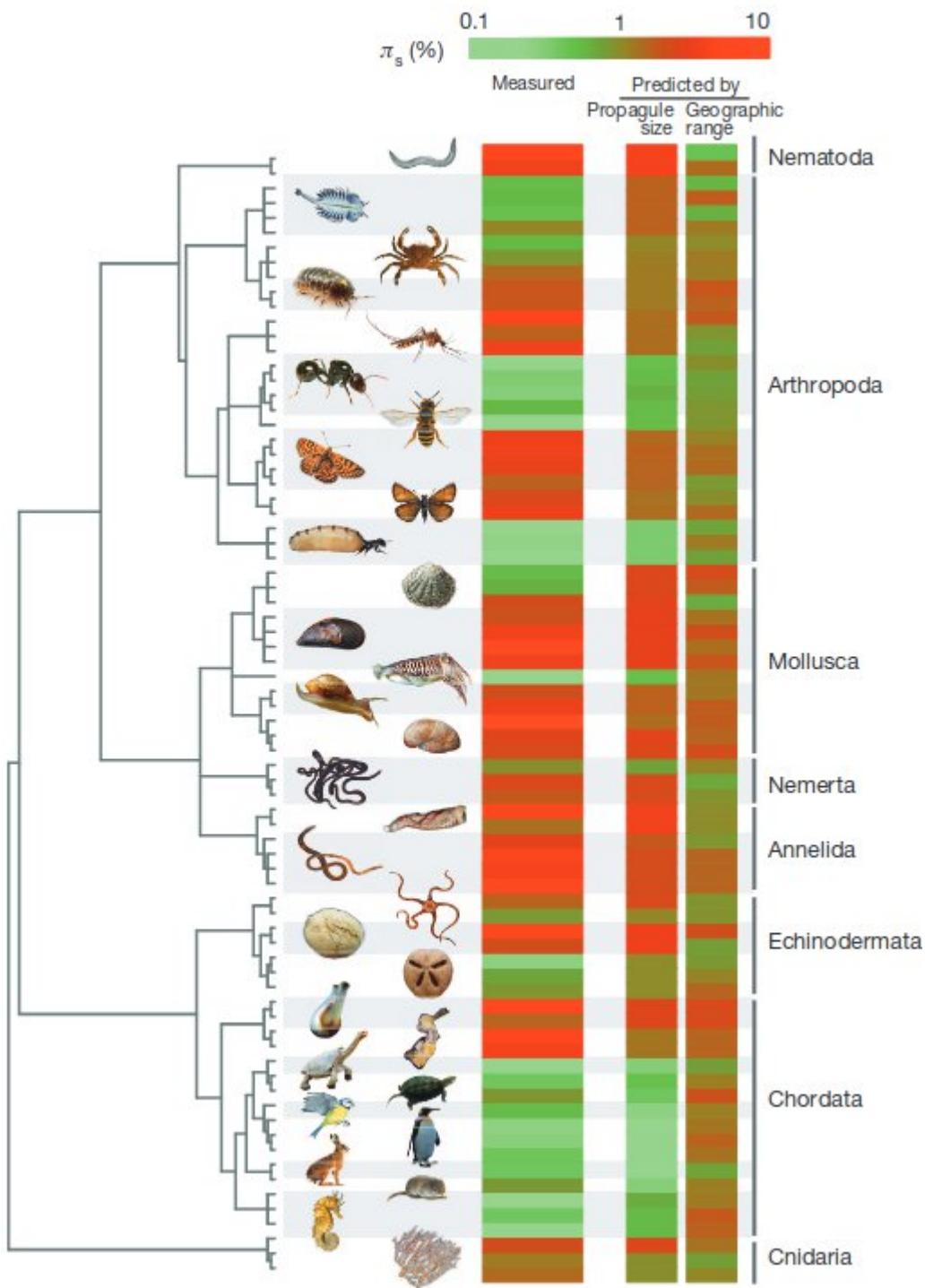


Figure 9 : Diversité génétique à l'échelle du génome dans l'arbre de la vie des métazoaires. Chaque branche de l'arbre représente une espèce ($n = 76$). La barre verticale colorée la plus à gauche est la diversité génétique estimée à l'échelle du génome (π_S), la barre centrale est la prédiction du π_S basée sur un modèle linéaire avec la taille des propagules comme variable explicative ($P < 10^{-14}$, $r^2 = 0.56$), et la barre la plus à droite est

la prédition du π_S basée sur un modèle linéaire incluant diverses métriques de distribution géographique et statut invasif comme variables explicatives ($P = 0.16$). Chaque vignette correspond à une famille de métazoaires. (Romiguier et al. 2014b).

Pour se reproduire, certaines espèces, dites à « stratégie r », misent sur le nombre ; c'est le cas des moules et des oursins, qui libèrent dans la colonne d'eau des dizaines de milliers de gamètes microscopiques. D'autres, au contraire, parient sur la qualité. Les parents investissent plus dans chacun de leurs descendants, qui sont donc moins nombreux mais ont chacun une probabilité plus élevée de survie jusqu'à l'état adulte – c'est la « stratégie K », employée notamment par les manchots, les tortues ou les insectes eusociaux. Comme on peut le voir illustré sur la Figure 10, ces espèces à fort investissement parental sont clairement peu polymorphes comparées aux espèces à stratégie r , probablement en raison de leur robustesse plus élevée aux changements environnementaux qui leur permettrait de tolérer de très forts goulots d'étranglement de population sans pour autant s'éteindre. Cette analyse démontre ainsi l'influence majeure des stratégies écologiques sur la taille effective de population de long-terme des métazoaires, un résultat aux implications immédiates pour les programmes de conservation des espèces. Les invertébrés à fort investissement parental (insectes eusociaux ou céphalopodes) sont probablement aussi vulnérables que les grands vertébrés aux risques d'extinction dû à une diversité génétique trop basse. Inversement, les espèces à stratégie r ont des diversités génétiques élevées quelque soit leur démographie actuelle, et pourraient donc encourir des risques d'extinction sans qu'aucun signal génétique ne soit visible au préalable.

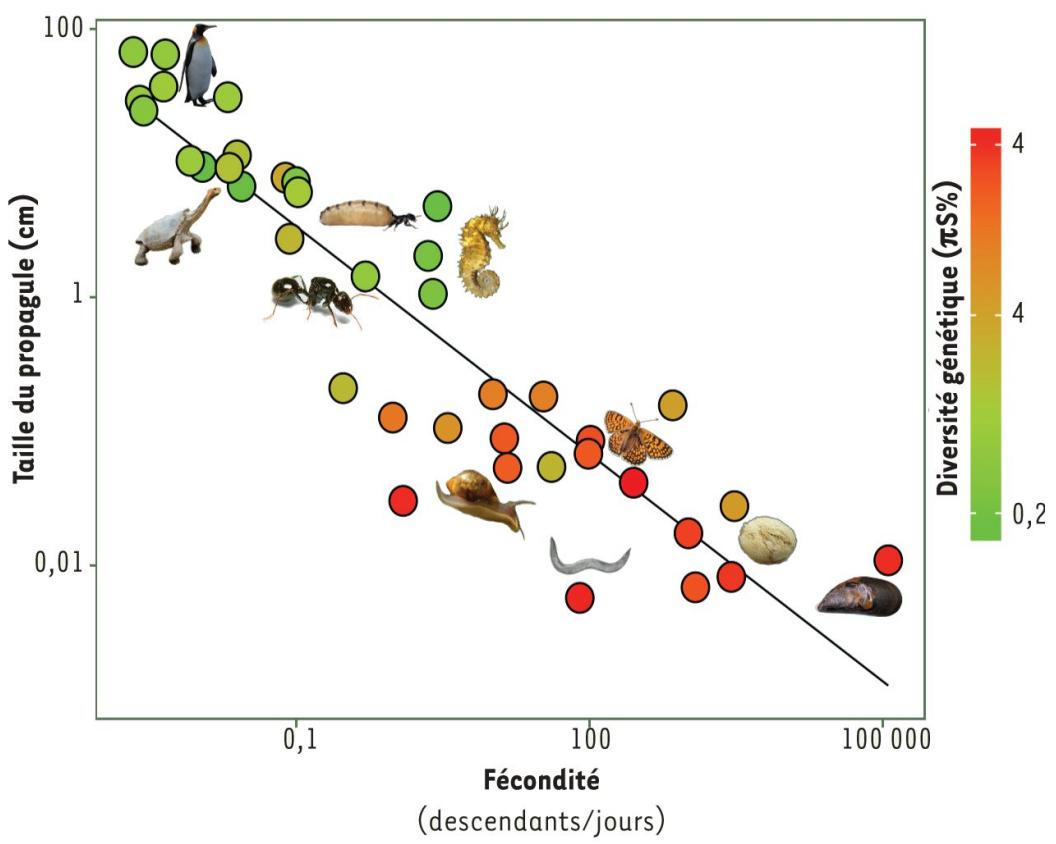


Figure 10 : Investissement parental et diversité génétique chez les animaux. (**Romiguier 2014b**)

1.2 Phylogénomique et évolution de nouvelles stratégies d'histoire de vie

1.2.1 Phylogénomique et racine des mammifères placentaires

Les relations de parentés les plus basales de l'arbre des mammifères placentaires ont toujours été le sujet de vives controverses. Depuis l'avènement de la phylogénie moléculaire, plusieurs scénarios de diversification biogéographique s'affrontent, une incertitude qui questionne notamment les hypothèses de spéciations suivant la dérive des continents. La difficulté majeure qui a toujours empêché de résoudre ces relations de parentés est la nature conflictuelle du signal phylogénétique entre marqueurs génétiques. Chaque gène peut en effet livrer sa propre histoire évolutive, celle-ci ne correspondant pas forcément à l'histoire des espèces (on parle alors de tri de lignée incomplet). En comparant les performances de plus de 13 000 gènes pour résoudre des relations de parenté unanimement soutenues dans l'arbre phylogénétique des mammifères, j'ai démontré que les gènes riches en nucléotides GC avaient tendance à moins bien refléter l'histoire des espèces (**Romiguier et al. 2013a**). Cet effet peut s'expliquer par les forts taux de recombinaison et de conversion génique biaisée subis par les gènes GC-riches, deux phénomènes qui peuvent entraîner d'importants biais pour les méthodes de reconstruction d'arbre (Boussau & Gouy 2006; Hobolth et al. 2011). En se focalisant sur la fraction la plus AT-riche des génomes mammaliens, nous avons mis en lumière une forte diminution du niveau de conflit entre gènes, résolvant par la même occasion l'enracinement de l'arbre des placentaires : les meilleurs marqueurs phylogénétiques soutiennent les Afrothériens (espèces à l'origine biogéographique essentiellement africaine) en tant que groupe-frère de tous les autres super-ordres (Fig. 4). Ce travail a démontré que la résolution de cet important noeud était jusque-là gênée par les parties les plus recombinantes et GC-riches des génomes, un résultat aux conséquences importantes pour la résolution future des noeuds les plus controversés de l'Arbre du Vivant.

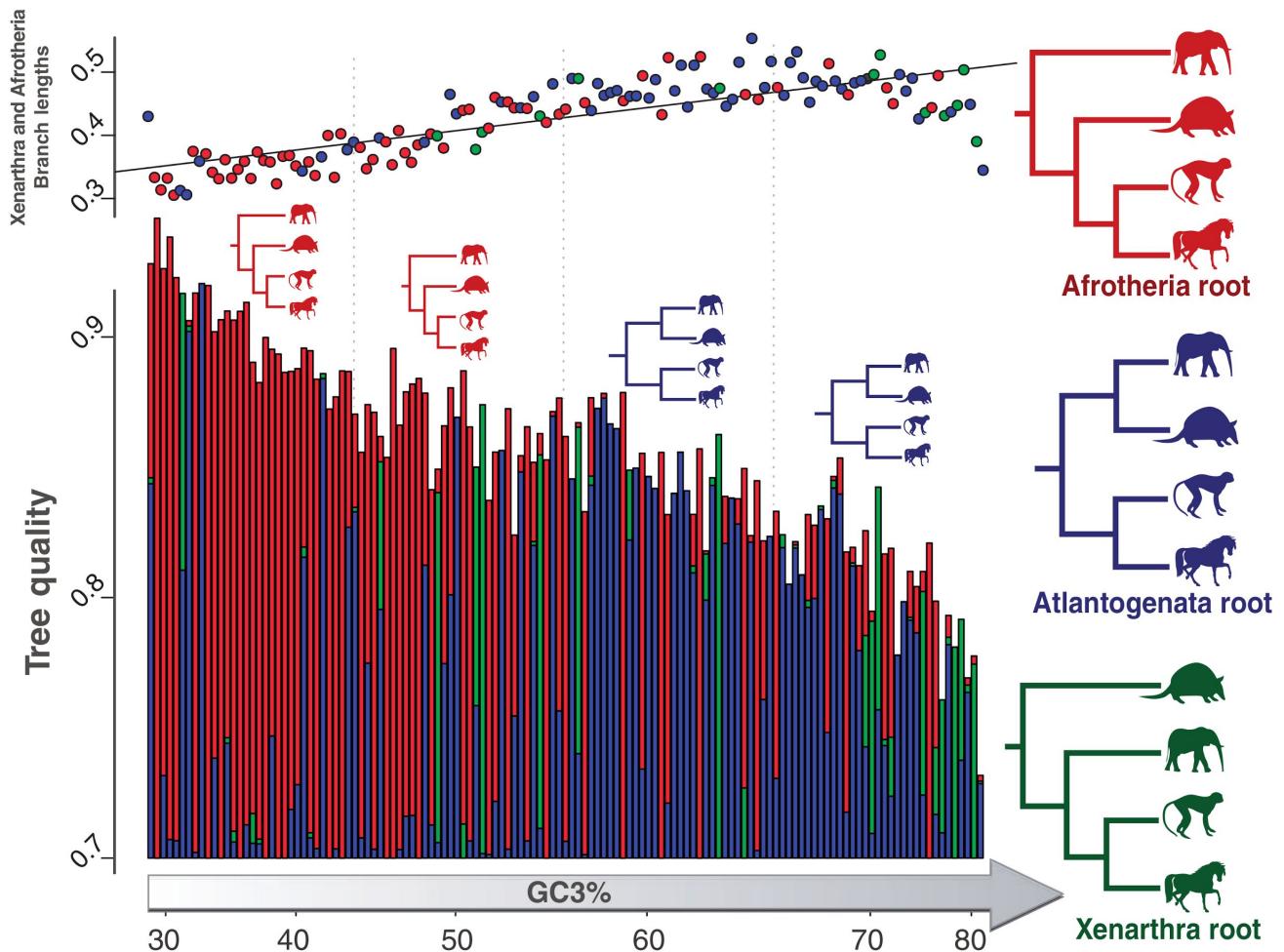


Figure 11 : Support pour la racine des mammifères placentaires en fonction du contenu en GC et de la qualité d'arbres (**Romiguier et al. 2013a**)

1.2.2. Phylogénomique et origine de l'eusocialité chez les abeilles corbiculées

L'abeille domestique *Apis mellifera* est une espèce à l'importance écologique et économique majeure. La position phylogénétique de sa tribu au sein des abeilles corbiculées a cependant été très débattue. Cette information est pourtant essentielle pour situer l'origine de l'eusocialité au sein de ce groupe, sujet central de nombreuses études et débats (Woodard *et al.* 2011; Danforth *et al.* 2013; Almeida & Porto 2014; Kapheim *et al.* 2015). La littérature la plus récente tendait cependant à favoriser une phylogénie suggérant un scénario évolutif peu parcimonieux qui impliquerait deux apparitions indépendantes de l'eusocialité.

Confronté à cette question largement débattue chez les spécialistes des abeilles, j'ai noté que l'hétérogénéité des taux de GC des génomes d'hyménoptères étaient susceptible d'entraîner d'importants biais de reconstruction d'arbres similaires à ceux que j'avais déjà observés chez les mammifères (**Romiguier et al. 2013a**). Les génomes des hyménoptères eusociaux (et tout particulièrement celui de l'abeille domestique) présentent en effet les taux de recombinaison parmi les plus extrêmes du monde vivant (Wilfert *et al.* 2007), un problème secondé par le fait que l'abeille domestique est la seule espèce d'insecte chez qui la conversion génique biaisée ait été démontrée comme étant particulièrement active (Kent *et al.* 2012; Wallberg *et al.* 2015).

En analysant les transcriptomes de 10 espèces d'abeilles corbiculées, j'ai démontré que le contenu en GC des gènes affecte considérablement les analyses phylogénétiques de ce groupe (**Romiguier et al. 2015**). Ces biais ne peuvent pas être corrigés par une simple traduction de données nucléotidiques en données protéiques, mais peuvent l'être dans une certaine mesure en utilisant des modèles d'évolution des séquences non-homogènes (Boussau & Gouy 2006). En se focalisant sur les marqueurs phylogénétiques les plus fiables ou en utilisant des modèles non-homogènes, j'ai retrouvé un fort support pour une phylogénie suggérant une seule origine à l'eusocialité chez les abeilles corbiculées (Fig. 8). Cet exemple démontre une nouvelle fois les biais qui peuvent être associés à la composition nucléotidique en évolution moléculaire (**Romiguier et Roux 2017**).

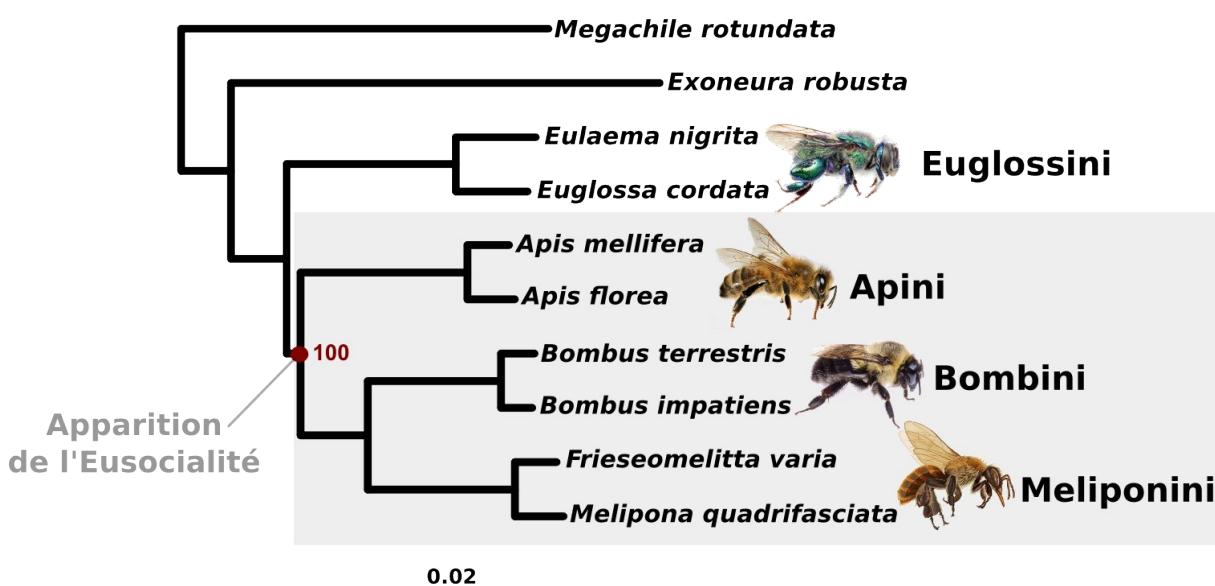


Figure 12 : Phylogénie des abeilles corbiculées (**Romiguier et al. 2015**)

1.2.3. Phylogénomique du genre *Formica* et évolution de l'esclavagisme

Le genre *Formica* est un genre de fourmis particulièrement étudié. En particulier, C. Darwin s'y intéressa tout particulièrement et les évoque longuement dans "L'Origine des Espèces" pour leur comportement de parasitisme social (Darwin 1871). Un groupe d'espèce de *Formica* à la monophylie incertaine, les "Serviformica", est activement exploité par plusieurs autres espèces du genre *Formica* ou *Polyergus*. Ainsi, les jeunes reines des sous-genres *Coptoformica* et *Formica sensu stricto* peuvent fonder une nouvelle colonie en tuant et prenant la place d'une reine de colonie de *Serviformica*, les ouvrières orphelines finissant par travailler pour le bourreau de leur mère. Ce parasitisme va encore plus loin au sein du sous-genre *Raptiformica*, avec un esclavagisme qui peut se poursuivre tout au long de la vie de la colonie. Les ouvrières *Raptiformica* peuvent ainsi effectuer des raids saisonniers afin de piller des cocons de *Serviformica* destinés à devenir de nouveaux esclaves. Au sein du genre *Polyergus*, on assiste à un hyper-spécialisation du comportement d'esclavagiste avec des ouvrières aux mandibules si spécialisées pour le combat qu'elles ne peuvent même plus entretenir le nid ou nourrir les larves de leurs propres soeurs. L'esclavagisme y est donc obligatoire et *Polyergus* est entièrement dépendant de ses esclaves du groupe *Serviformica* (Buschinger & Alfred 1986).

Dans ce projet, j'ai échantillonné, séquencé, assemblé et analysé le transcriptome entier de 16 espèces du genre *Formica* et *Polyergus*. Ces données m'ont permis d'assembler le plus gros jeu de données phylogénomique pour ces genres (**Romiguier et al. 2018**), ce qui m'a permis de clarifier les relations incertaines entre sous-genres (Goropashnaya et al. 2012). Ces analyses phylogénomiques suggèrent notamment que le style de vie "libre" est ancestral au genre *Formica*, tandis que le comportement de parasitisme temporaire lors de la fondation d'une nouvelle colonie n'aurait évolué qu'une seule fois (Fig. 13). Ces résultats supportent l'hypothèse selon laquelle la fondation parasitaire de colonies serait une étape intermédiaire nécessaire à l'évolution de l'esclavagisme, une hypothèse déjà émises par Bushinger et Alfred (1986). Plus récemment, un jeu de données plus complet en terme de nombre d'espèces est venu confirmer ce résultat (Borowiec et al. 2021).

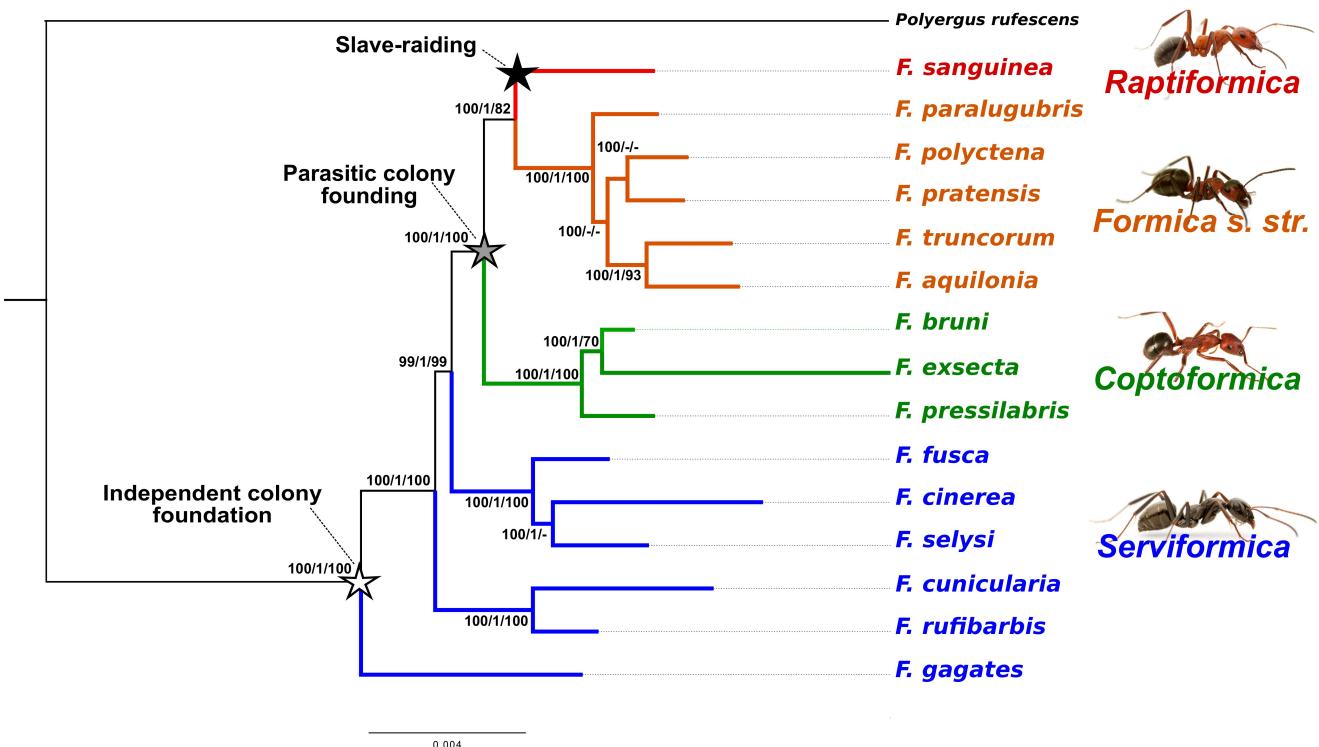


Figure 13 : Phylogénie moléculaire des *Formica*. Les longueurs des branches et la topologie sont basées sur une analyse par maximum de vraisemblance. Le valeurs de supports des trois méthodes utilisées (maximum de vraisemblance, inférence bayésienne et méthode basée sur la coalescence) est indiqué pour chaque nœud. Lorsqu'une méthode ne retrouve pas un nœud, la valeur de soutien est remplacée par un "-". (Romiguier et al. 2018)

1.2.4. Phylogénomique et évolution moléculaire des Formicidae

L'évolution de l'eusocialité a permis aux fourmis de devenir l'un des groupes d'animaux les plus visibles et écologiquement dominants au monde. Une grande majorité des ~14 000 espèces de fourmis actuelles appartiennent aux formicoïdes, un clade de neuf sous-familles qui présentent une division du travail reproducteur très élevée, de grandes tailles de colonie, un fort polymorphisme entre ouvrières et une longévité extrême des reines.

Les huit autres sous-familles non formicoïdes sont quand à elles moins bien étudiées : peu de génomes ont été séquencés jusqu'à présent et leurs relations phylogénétiques ne sont pas claires. Comme pour les mammifères placentaires, l'enracinement de l'arbre des fourmis est particulièrement controversé depuis la découverte de *Martialis heureka*, une espèce extrêmement rare et morphologiquement très divergente qui fut initialement inférée comme étant l'espèce soeur de toutes les autres fourmis vivantes actuellement (Rabeling et al. 2008). Par la suite, deux autres hypothèses contradictoires ont été tour à tour supportées par différentes études, avec la sous-famille des Leptanillinae à la racine des fourmis Kück et al.

2011, Moreau et Bell 2013) puis plus récemment les Leptanillinae et Martialinae pouvant tous les deux être à la racine de l'arbre selon la composition en base des gènes sélectionnés (Borowiec *et al.* 2019).

En séquençant 65 génomes avec au moins une espèce représentative pour chacune des 17 sous-famille de fourmis existant de nos jours, nous avons comblé le manque important de diversité phylogénétique en terme de génomes séquencés (**Romiguier *et al.* 2022**). Cela a permis dans un premier temps à fournir une phylogénie robuste des sous-familles de fourmis, confirmant l'existance du groupe des poneroïdes et de supporter un enracinement de l'arbre au niveau du clade controversé des leptanillomorphes (Leptanillinae + Martialinae) (voir Fig. 14). Ce clade composé exclusivement de petites fourmis souterraines serait ainsi le groupe frère de toutes les autres fourmis existantes.

De plus, en comparant le dN/dS de chaque branche au reste de l'arbre pour 4415 gènes orthologues, nous avons révélés que l'émergence des formicoïdes a été accompagnée d'un nombre relatif d'événements de sélection positive extrêmement élevé (voir Fig. 15). Avec 110 gènes sous sélection positive contre 3,1 en moyenne, cette branche est à l'origine du groupe où ont émergés les formes d'eusocialités les plus avancées (dimorphisme reine/ouvrière le plus important, grande taille de colonie, fécondité et longévité des reines extrême). De manière intéressante, les trois fonctions génétiques les plus sélectionnées le long de ce point chaud de sélection positive sont potentiellement liées à des caractéristiques clés de l'eusocialité avancée : l'acétylation des histones étant impliquée dans la différenciation des castes (notamment démontré via un traitement expérimental, voir Simola *et al.* 2016), l'extinction des gènes par ARN dans la stérilité des ouvrières (Hartfelder *et al.* 2018) et l'autophagie dans la longévité (Madeo *et al.* 2015). Ces résultats montrent que les voies clés associées à l'eusocialité ont fait l'objet d'une forte sélection au cours du Crétacé. Cela suggère que les fondements moléculaires de l'eusocialité complexe ont probablement été rapidement acquis en moins de 20 millions d'années.

FORMICOIDS

PONEROIDS

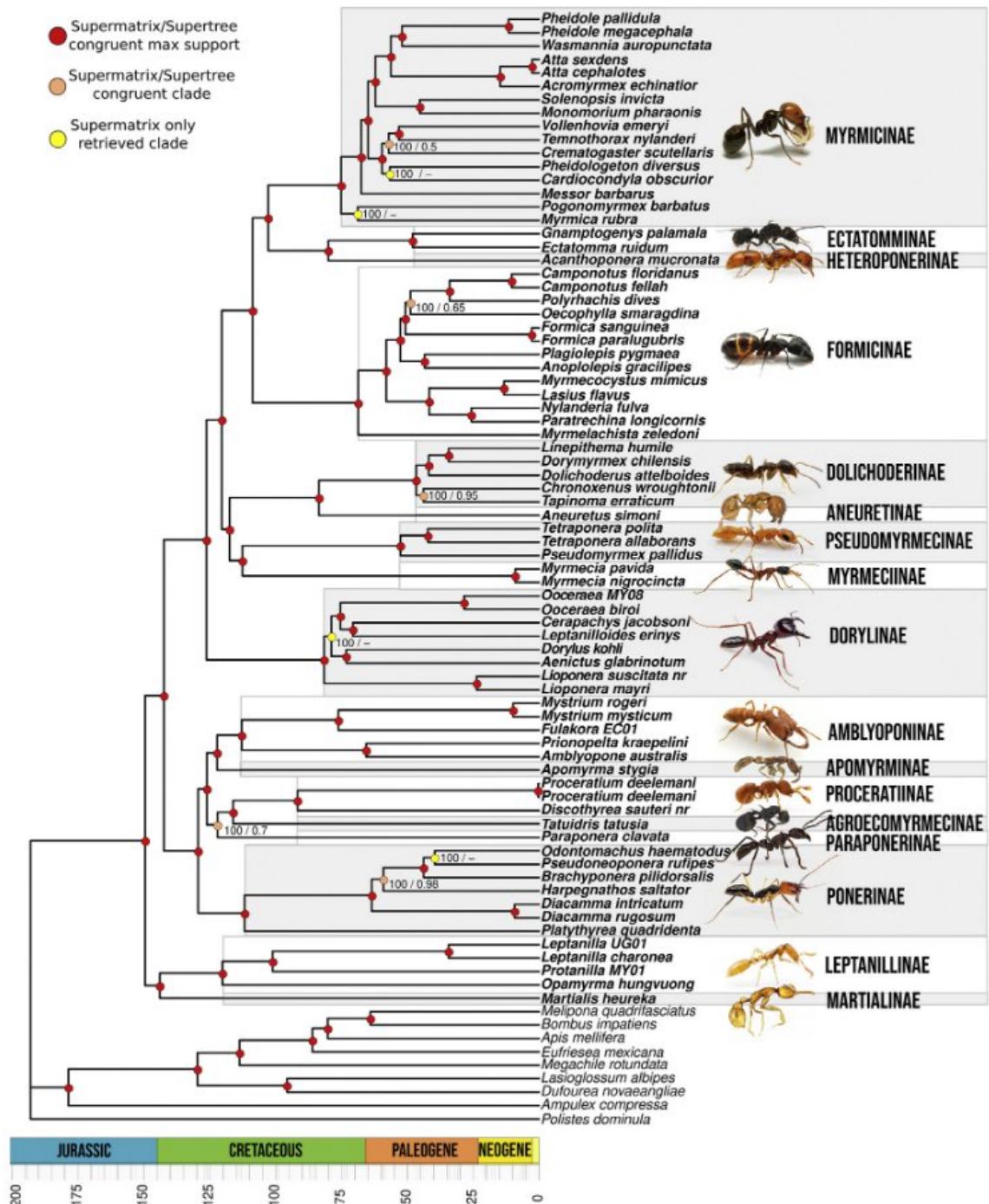


Figure 14 : Phylogénie des fourmis basée sur des données génomiques. La topologie est déduite selon l'analyse principale de la supermatrice (1 692 052 sites d'acides aminés après nettoyage ; maximum de vraisemblance ; modèle de mélange de profils PMSF C20 de IQ-TREE) ; le support bootstrap est affiché en premier lorsque le support des nœuds n'est pas maximal. Le soutien des nœuds des principales analyses de super-arbres (à partir des arbres génétiques des 1 552 alignements de plus de 500 sites d'acides aminés ; arbres génétiques déduits avec la recherche de modèle dans IQ-TREE, et arbre des

espèces avec ASTRAL) est affiché en second lorsque le soutien n'est pas maximal (tiret lorsque le nœud présente une incongruence avec l'analyse de supermatrice). La divergence temporelle a été estimée à l'aide de chronos avec 12 nœuds de calibration. Images de fourmis d'Alex Wild ; utilisées avec permission (**Romiguier et al. 2022**)

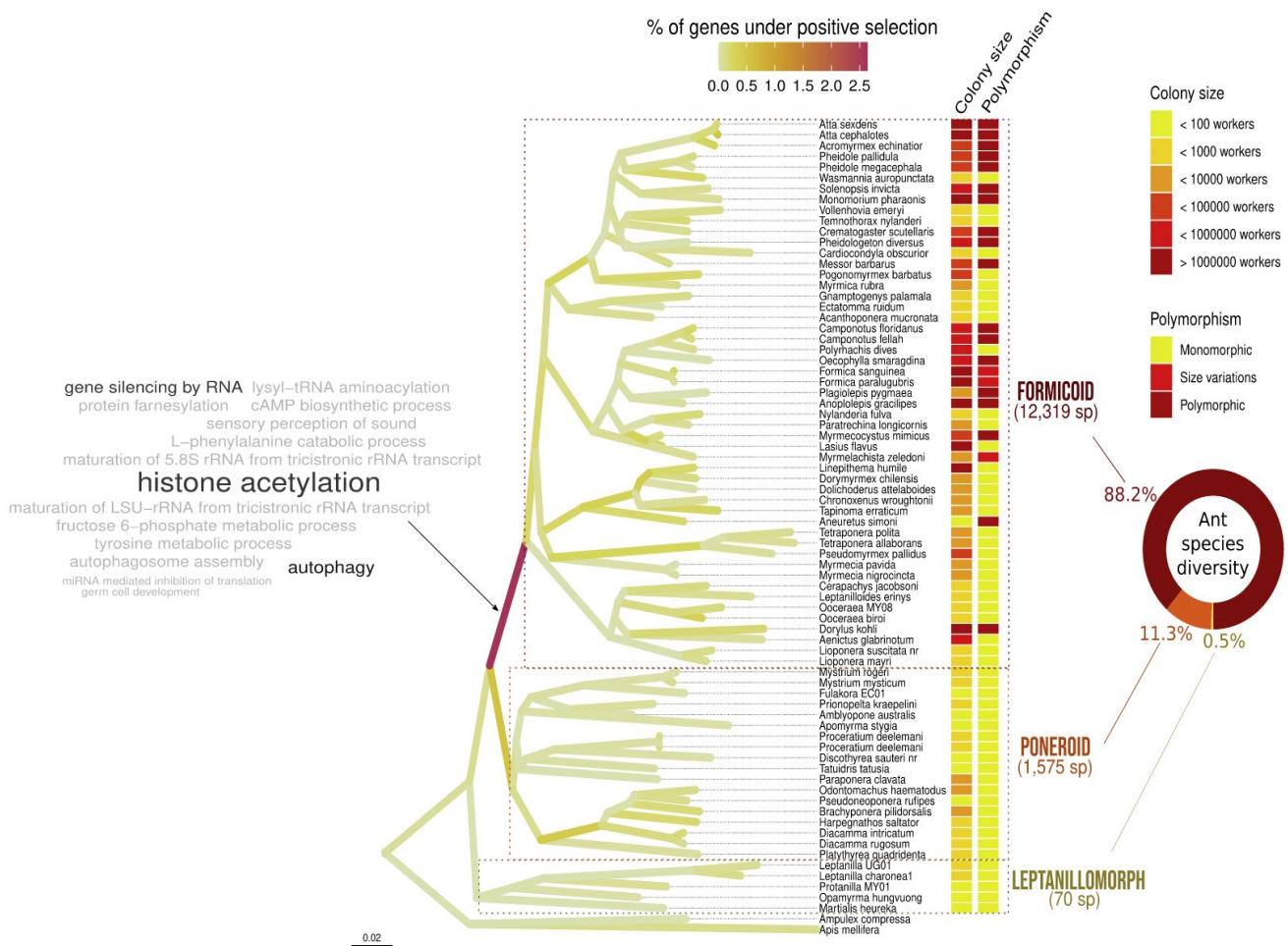


Figure 15 : La branche menant aux formicoïdes montre des taux élevés de sélection positive. Les couleurs indiquent le pourcentage de gènes significativement sous sélection positive (analyse aBSREL) sur chaque branche. Les données relatives à la taille des colonies et au polymorphisme représentent la valeur maximale observée dans le genre. Les catégories fonctionnelles (processus biologiques) significativement enrichies sous sélection positive sont représentées sous forme de nuage de mots pour la branche des formicoïdes. La taille de la police est proportionnelle à la p-value (test exact de Fisher ; une police plus grande indiquant les plus significatives, de 0,011 à 0,047). Les couleurs plus foncées indiquent les trois fonctions, avec le plus grand nombre de gènes annotés significatifs sous sélection positive (**Romiguier et al. 2022**).

1.3. Evolution vers une stratégie reproductive atypique : l'hybridogénèse sociale

1.3.1. Déterminisme de caste : contexte

La part d'inné et d'acquis dans le devenir d'un individu est une vaste question qui taraude régulièrement philosophes et scientifiques. En biologie, cet épineux problème peut se résumer par la part relative qu'ont les gènes et l'environnement sur le phénotype ou le comportement d'un organisme. La variation morphologique intra-spécifique est particulièrement marquée chez les insectes eusociaux. Divisés en castes à la morphologie, physiologie et comportements bien distincts, les insectes eusociaux arborent une impressionnante diversité de formes entre individus d'une même espèce. De manière remarquable, cette diversité phénotypique ne se limite pas au dimorphisme sexuel. Si un œuf non-fécondé d'hyménoptère eusocial se développe quasi-invariablement en mâle haploïde (on parle d'haplo-diploïdie), un œuf fécondé a deux destinées femelles possibles : participer à la reproduction de la colonie en tant que reine, ou s'acquitter des soins et de la recherche de nourriture en tant qu'ouvrière.

Les déterminants qui influencent la différentiation en une caste stérile ou fertile ont longtemps été jugés exclusivement environnementaux. Les premiers arguments sont tout d'abord théoriques : il est en effet difficile d'imaginer la persistance évolutive d'un variant génétique induisant la stérilité (Hölldobler & Wilson 1990). Un allèle qui aurait pour effet systématique de réduire drastiquement ou complètement la fécondité ne pourrait que difficilement se maintenir dans une population, les femelles le portant ne pouvant pas avoir de descendance. Pour qu'un système de caste stérile puisse se maintenir, il est plus simple d'imaginer un gène ouvrière qui ne serait exprimé que conditionnellement à l'environnement. Le fameux exemple de l'ingestion de gelée royale induisant la différentiation d'une larve d'abeille domestique (*Apis mellifera*) en reine a finit d'imposer l'idée de différentiation environnementale de la caste (Smith *et al.* 2008).

L'apport de la génétique a cependant remis en cause le dogme de détermination environnementale de la caste. L'influence directe ou indirecte d'effets génétiques ont en effet été démontrés chez plusieurs populations ou espèces, de telle sorte qu'on parle désormais de continuum de détermination génétique à environnementale de la caste (Schwander *et al.* 2010). Dans la plupart des cas où le déterminisme génétique est quasi-total, une évolution vers la production asexuée des reines est impliquée : un œuf fécondé ne produit dès lors plus que des ouvrières (Schwander & Keller 2012). L'exemple qui a suscité le plus d'intérêt dans la littérature est cependant celui d'un système qui conserve la reproduction sexuée pour la production des deux castes femelles : un œuf fécondé peut toujours se développer en reine ou en ouvrière, mais c'est l'interaction du génome paternel et maternel qui détermine

ce choix. Un tel système est ainsi observable au sein de certaines populations d'origine hybride chez les fourmis moissonneuses nord-américaines du genre *Pogonomyrmex* (Helms Cahan & Keller 2003). Dans ces populations, deux lignées dépendantes l'une de l'autre co-existent et se reproduisent sans jamais se mélanger. Lorsque l'œuf d'une reine est fécondé avec le spermatozoïde d'un mâle de sa lignée, l'œuf se différencie en reine sexuée. A contrario, lorsque l'œuf d'une reine est fécondé par le spermatozoïde d'un mâle de lignée différente, l'œuf se différencie en ouvrière. Ce système de reproduction peu orthodoxe a été nommé hybridogénèse sociale : les croisements intra-lignée produisent des reines, les croisements inter-lignée produisent des ouvrières (voir Fig. 16).

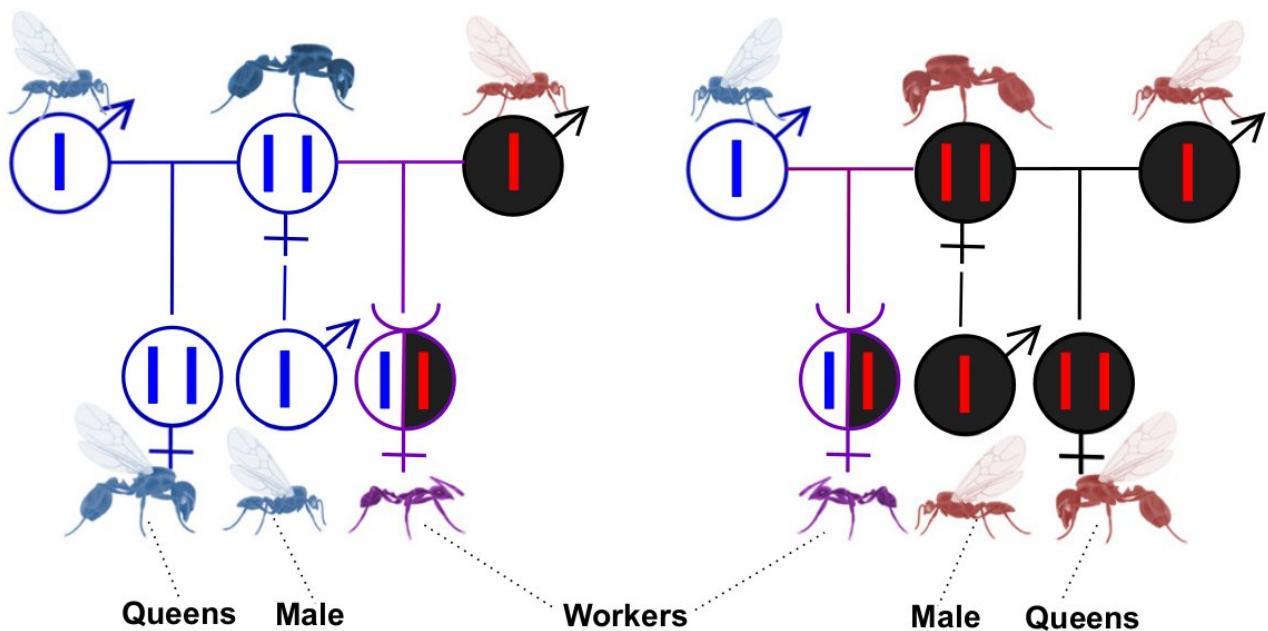


Figure 16 : Hybridogénèse sociale chez les fourmis moissonneuses (Romiguier et al. 2016).

1.3.2. Découverte de l'hybridogénèse sociale chez les fourmis du genre *Messor*

L'hybridogénèse sociale est un système de reproduction où deux lignées distinctes de reines doivent s'hybrider pour produire des ouvrières (voir Fig 16). Nous avons découvert plusieurs nouveaux cas d'hybridogénèse sociale au sein du genre *Messor* (Romiguier et al. 2016). Cette découverte a été faite de manière fortuite, suite à d'importants problèmes rencontrés avec une des 76 espèces de mon travail sur la diversité génétique des animaux (section 1.1.5). Lors de l'échantillonnage de ce projet, j'ai initialement sélectionné la fourmi moissonneuse *Messor barbarus* en tant qu'espèce cible pour son abondance et sa facilité d'échantillonnage. Après avoir échantillonné, séquencé et assemblé 10 transcriptomes

d'ouvrières provenant de différentes populations, j'ai réalisé que les données présentaient une forte anomalie : un fort excès d'hétérozygotie (F_{IS} très négatif, Fig. 16). Afin de vérifier qu'il ne s'agisse pas d'un problème technique (comme exemple, un problème de contamination), j'ai re-séquencé 10 nouveaux individus qui ont confirmés que cette hétérozygotie anormale était bel et bien biologique, du moins chez les ouvrières. En effet, une reine qui avait elle aussi été séquencée parmi les 10 individus présentait elle une niveau d'hétérozygotie totalement normal.

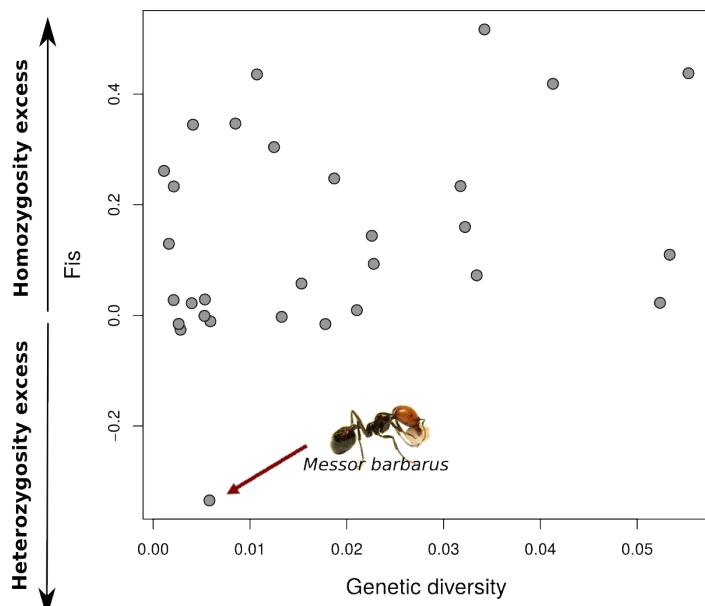


Figure 16 : Excès d'hétérozygotie chez *Messor barbarus* comparé à 29 autres espèces animales

Après avoir séquencé plus d'individus, il s'est avéré que toutes les populations investiguées (25 différentes, réparties en Espagne, France et Maroc) présentaient des ouvrières hybrides auxquelles correspondent deux lignées pures de reines. Un système de reproduction identique à celui trouvé au sein de certaines populations de *Pogonomyrmex* (Fig. 16) a été confirmé via le séquençage complet d'une famille (reine mère, reine fille, ouvrière fille, mâle fils) et du génotypage de plus de 120 individus de toutes castes répartis sur 12 nids (**Romiguier et al. 2016**). Chez *Messor barbarus*, la détermination génétique de la caste ne semble cependant pas uniquement restreinte à certaines populations et paraît bien plus stricte que chez *Pogonomyrmex* : jusqu'à 18% de reines sont hybrides dans certaines populations de *Pogonomyrmex* pratiquant l'hybridogénèse sociale (Schwander et al. 2007), soit un déterminisme génétique de la caste qui n'est pas total.

De surcroît, le séquençage d'une reine mère et d'une ouvrière fille au sein de 8 autres espèces de ce genre a révélé deux autres cas d'hybridogénèse sociale. De manière remarquable, les trois espèces pratiquant ce mode de reproduction (*Messor barbarus*, *Messor structor* et *Messor ebeninus*) sont séparées dans trois clades différents au sein de la

phylogénie du genre, et l'apparition des deux lignées divergentes s'est vraisemblablement produite trois fois indépendamment (99% des positions diagnostiques entre lignées ne sont pas partagées par les trois espèces, voir exemple en Fig. 17).

0.002

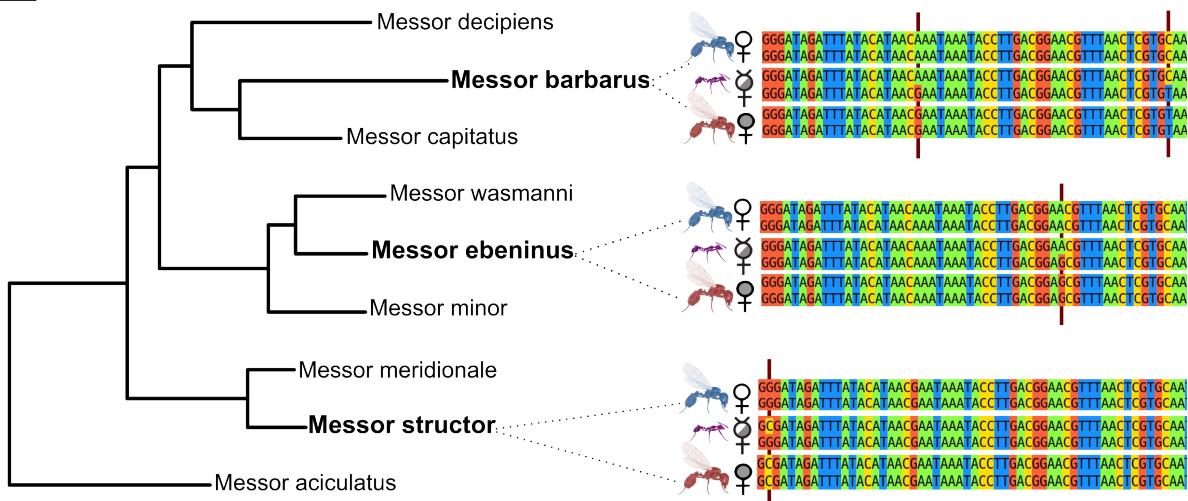


Figure 17 : Hybridogénèse sociale répétée au sein du genre *Messor*

1.3.3. Modèle théorique pour l'origine de l'hybridogénèse sociale

Bien que de plus en plus de cas d'hybridogénèse sociale soient découverts (Helms Cahan *et al.* 2002, Helms Cahan et Vinson 2003, Anderson *et al.* 2006, Romiguier *et al.* 2016, Lacy *et al.* 2019, Kuhn *et al.* 2020), le scénario évolutif pouvant mener à un tel système de reproduction n'est toujours pas bien identifié. C'est dans ce contexte que nous avons développé un modèle simple pouvant expliquer comment un tel système peut évoluer relativement facilement via l'évolution conjointe de l'hybridation et d'allèles biaiseurs de castes (**Weyna *et al.* 2021**).

Le système de caste de tous les insectes eusociaux est en effet théoriquement vulnérable à l'apparition d'éléments génétiques qui favoriseraient le développement de leurs porteurs en reines, favorisant ainsi de manière égoïste la propagation de leurs génotypes aux générations suivantes. L'idée de ce modèle est d'étudier l'évolution de tels allèles biaiseurs de caste et le parasitisme spermatique, phénomène au cours duquel une reine s'hybride avec un mâle d'une autre espèce pour produire des ouvrières stériles (Umphrey 2006). En utilisant des outils de modélisation mathématique (voir Fig 18 et **Weyna *et al.* 2021** pour détails), nous montrons que la coévolution de l'hybridation avec la détermination de la caste déclenche facilement une course aux armements évolutive entre les larves non-hybrides qui se développent de plus en plus en reines et les reines qui s'hybrident de plus en plus pour produire des ouvrières. De manière remarquable, même lorsque l'hybridation réduit

l'efficacité de la fonction ouvrière et la valeur sélective de la colonie, cette course peut conduire à une détermination de la caste purement génétique -voir Fig. 19). Globalement, nos résultats peuvent aider à comprendre l'évolution répétée vers des systèmes avec un déterminisme génétique de la caste, que cela soit dans le cas de l'hybridogénèse sociale ou d'autres systèmes comme l'utilisation conditionnelle du sexe et de la parthénogénèse pour produire ouvrière et reines

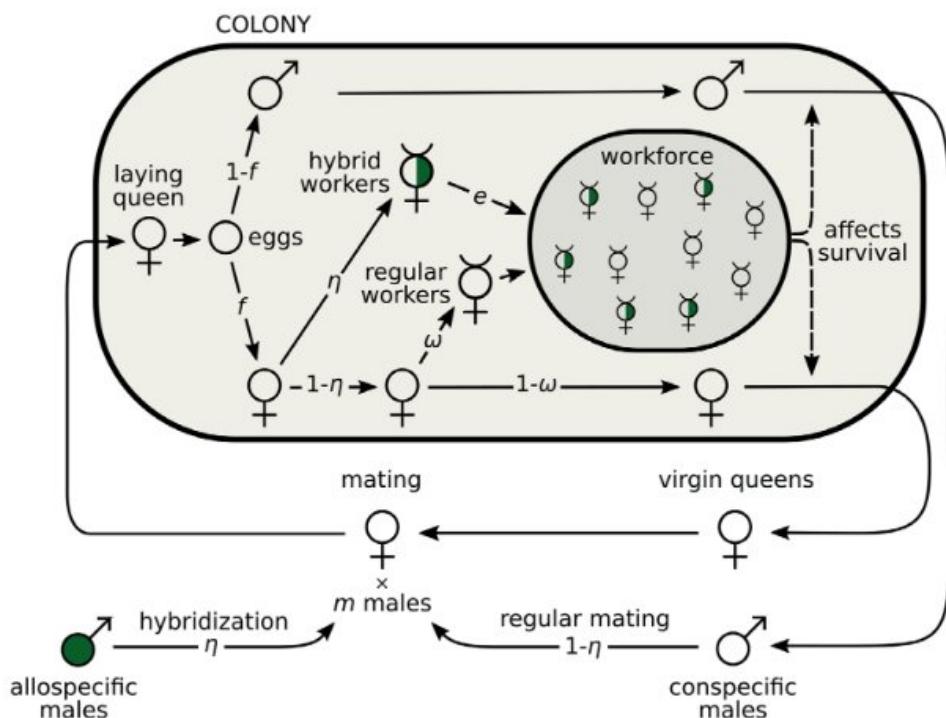


Figure 18 : Schéma général du modèle mathématique principal de cette étude. Le cycle de vie d'une espèce eusociale annuel avec hybridation et parasitisme spermatique (utilisation du sperme d'une autre espèce pour produire des ouvrières stériles) y est représenté. A chaque génération, le cycle de vie commence avec des reines vierges s'accouplant avec m mâles, chacun d'entre eux ayant une probabilité η d'être allospécifique et $1 - \eta$ d'être conspécifique. Après l'accouplement, une reine fonde une colonie et commence à produire des œufs. Les œufs de femelles hybrides (avec une origine paternelle allospécifique) se développent tous en ouvrières. Les femelles non-hybrides (avec une origine paternelle conspécifique) se développent en ouvrières avec une probabilité ω et en reines le cas échéant. La variable η capture donc la tendance des reines à s'hybrider et pratiquer le parasitisme spermatique, tandis que ω contrôle la détermination de la caste. (**Weyna et al. 2021**).

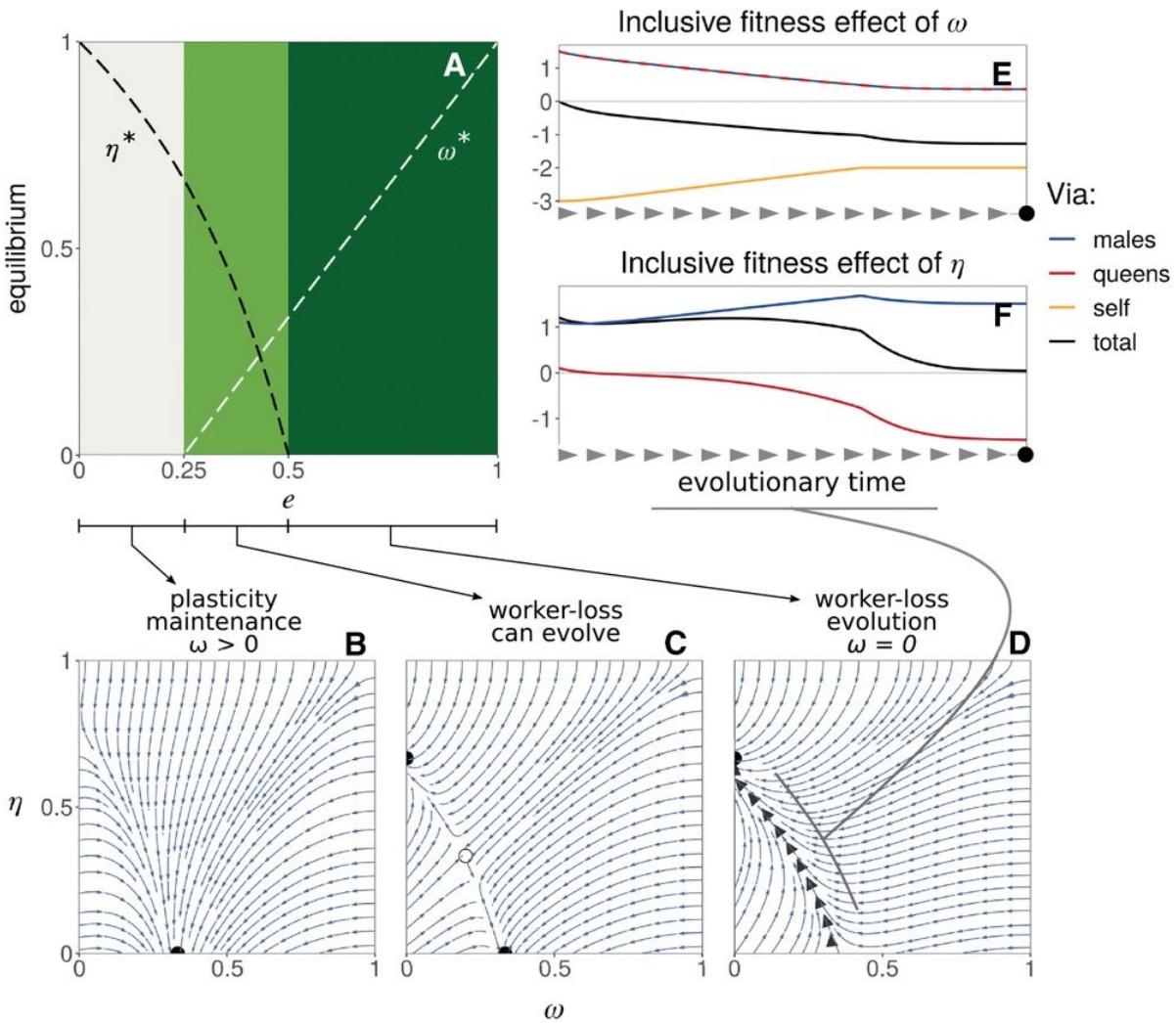


Figure 19 : La coévolution de la détermination de la caste et du parasitisme des spermatozoïdes. (A) Équilibres évolutifs (pour probabilité d'hybridation η en noir et probabilité de se développer en ouvrière ω en blanc) en fonction de l'efficacité des travailleurs hybrides e . Ces équilibres, cependant, sont des répulsifs évolutifs. Par conséquent, trois types de dynamiques coévolutives sont possibles en fonction de e , comme illustré dans les panneaux (B)-(D). Ces panneaux montrent des exemples de trajectoires phénotypiques vers un déterminisme génétique de la caste : Le panneau (B) n'évolue jamais vers un déterminisme génétique de la caste et perte d'ouvrières intra-spécifique ($e = 0.1$) ; le panneau (C) peut évoluer en fonction des conditions initiales ($e = 0.4$) ; le panneau (D) évolue toujours vers un déterminisme génétique de la caste ($e = 0.7$). Les cercles noirs pleins indiquent les deux points finaux de l'évolution : évitement de l'hybridation avec plasticité développementale ($\omega = 1/3$ et $\eta = 0$ dans B et C) ou la perte de travailleurs avec l'hybridation ($\omega = 0$ et $\eta = 2/3$ dans C et D). Le cercle vide dans (C) montre l'équilibre interne instable. Les flèches grises épaisses dans (D) représentent la trajectoire d'une population partant de $\omega = 1/3$ et $\eta = 0$ évoluant vers la perte d'ouvrières.

(E) Effets sur la valeur adaptative de la détermination de caste ω dans une larve mutante via elle-même (en orange), les reines apparentées (rouge) et les mâles apparentés (bleu) le long de la trajectoire menant à la perte d'ouvrières montrée dans le panneau (D) (sélection totale en noir, Annexe B.1.3 pour la dérivation). Nous constatons que les effets négatifs de la fitness via soi-même (ligne orange) conduisent à un effet de sélection total qui est négatif (ligne noire). Ceci indique que les larves mutantes avec des valeurs de plus en plus petites de ω sont sélectionnées parce que ces valeurs augmentent la fitness directe des larves (en augmentant la probabilité qu'elles se développent en reines). (F)

Effets de l'hybridation η dans une reine mutante, via ses fils (bleu) et ses reines filles (rouge) le long de la trajectoire menant à la perte d'ouvrières montrée dans le panneau (D) (sélection totale en noir). La sélection totale positive (en noir) est principalement due à une augmentation de l'aptitude via les mâles (en bleu). Cela signifie que les reines mutantes avec des valeurs de plus en plus grandes de η sont sélectionnées parce que cela augmente leur reproduction, surtout via les mâles. (**Weyna et al. 2021**)

1.3.4. Méthode de détection d'hybride F1 pour scans d'hybridogénèse sociale à large échelle

Comme vu précédemment, de plus en plus de cas d'hybridogénèse sociale sont découverts. Bien que de plus en plus de cas d'hybridogénèse sociale soient découverts par hasard (Helms Cahan *et al.* 2002, Helms Cahan et Vinson 2003, Anderson *et al.* 2006, Romiguier *et al.* 2016, Lacy *et al.* 2019, Kuhn *et al.* 2020). Cela pose de plus en plus la question de la réelle prévalence de tels systèmes de reproduction, et appelle au développement de méthodologies efficaces pour tester la présence d'hybrides F1 dans de larges jeux de données phylogénétiques existants, ce qui pourraient nous indiquer des candidats potentiels à l'hybridogénèse sociale.

De manière plus générale, l'hybridation occupe un rôle central dans de nombreux processus évolutifs fondamentaux, tels que la spéciation ou l'adaptation. Pourtant, malgré son importance capitale dans l'évolution, on sait peu de choses sur la prévalence et la distribution réelle de l'hybridation actuelle dans l'arbre du vivant. Dans **Weyna et al. 2022**, nous avons développés et mis en œuvre une nouvelle méthode statistique permettant la détection des hybrides F1 à partir de données de séquençage du génome d'un seul individu. En utilisant des simulations et des données de séquençage de systèmes hybrides connus, nous démontrons d'abord la spécificité de la méthode et identifions ses limites statistiques. Ensuite, nous présentons la méthode en l'appliquant aux données de séquençage disponibles de plus de 1 500 espèces d'arthropodes, comprenant des hyménoptères, des hémiptères, des coléoptères, des diptères et des arachnides. Parmi ces taxons, nous constatons que les Hyménoptères, et en particulier les fourmis, présentent le plus grand nombre d'hybrides F1 candidats, ce qui suggère des taux plus élevés d'hybridation récente entre des pools génétiques précédemment isolés (voir Fig. 20). La prévalence des hybrides F1 sont distribués de manière hétérogène parmi les fourmis (voir Fig. 21), les taxons comprenant de nombreux candidats ayant tendance à abriter des traits écologiques et

d'histoire de vie spécifiques tels que des régimes alimentaires spécialisés. Ces taxons pourraient en conséquence être des candidats privilégiés pour découvrir de nouveaux cas d'hybridogénèse sociale ou autre systèmes de reproduction atypiques impliquant des ouvrières hybrides de première génération. Ce travail montre également comment des études génomiques comparatives à grande échelle peuvent être mises en œuvre pour mieux comprendre la distribution de l'hybridation dans l'arbre du vivant.

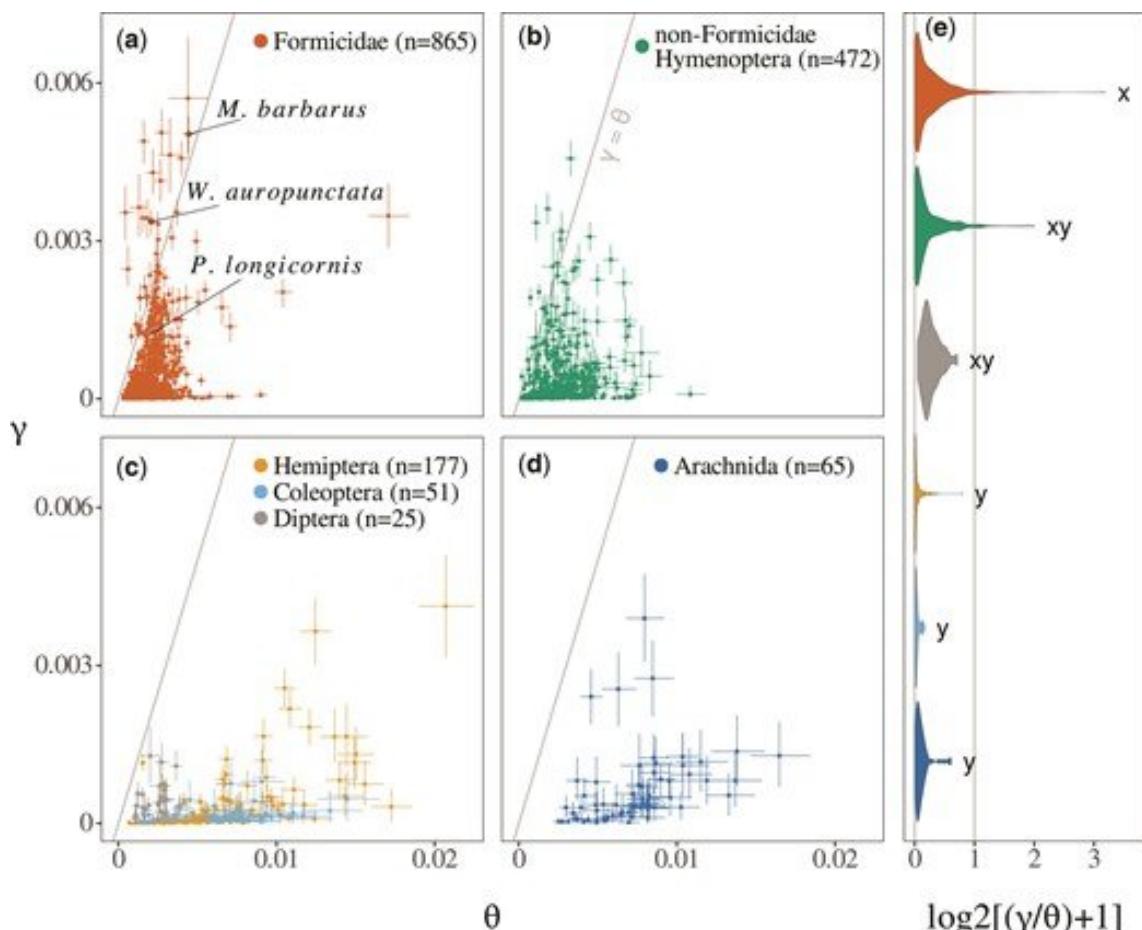


Figure 20 : Scans génomiques pour l'hybridation dans six groupes d'arthropodes. Les estimations du paramètre de divergence γ et du taux de mutation de la population ancestrale θ sont représentées pour les Formicidae (a), les Hyménoptères non-Formicidae (b), les autres insectes (c) et les Arachnida (d). Les points et les lignes colorés représentent respectivement les estimations ponctuelles bayésiennes et les intervalles de crédibilité. (e) Distribution du rapport γ/θ dans chaque groupe. Une échelle \log_2 décalée, sous laquelle la valeur critique de $\gamma/\theta=1$ est inchangée, a été utilisée pour des raisons de commodité visuelle. Les lettres résument le résultat d'un test de signification honnête post-hoc de Tukey, effectué à l'aide de la fonction HSD.test du package R agricolae. Les groupes n'ayant pas de lettres en commun ont des moyennes significativement différentes (avec $\alpha=0,05$). Tous les résultats ont été obtenus en utilisant uniquement des échantillons datés et récents. (**Weyna et al. 2022**)

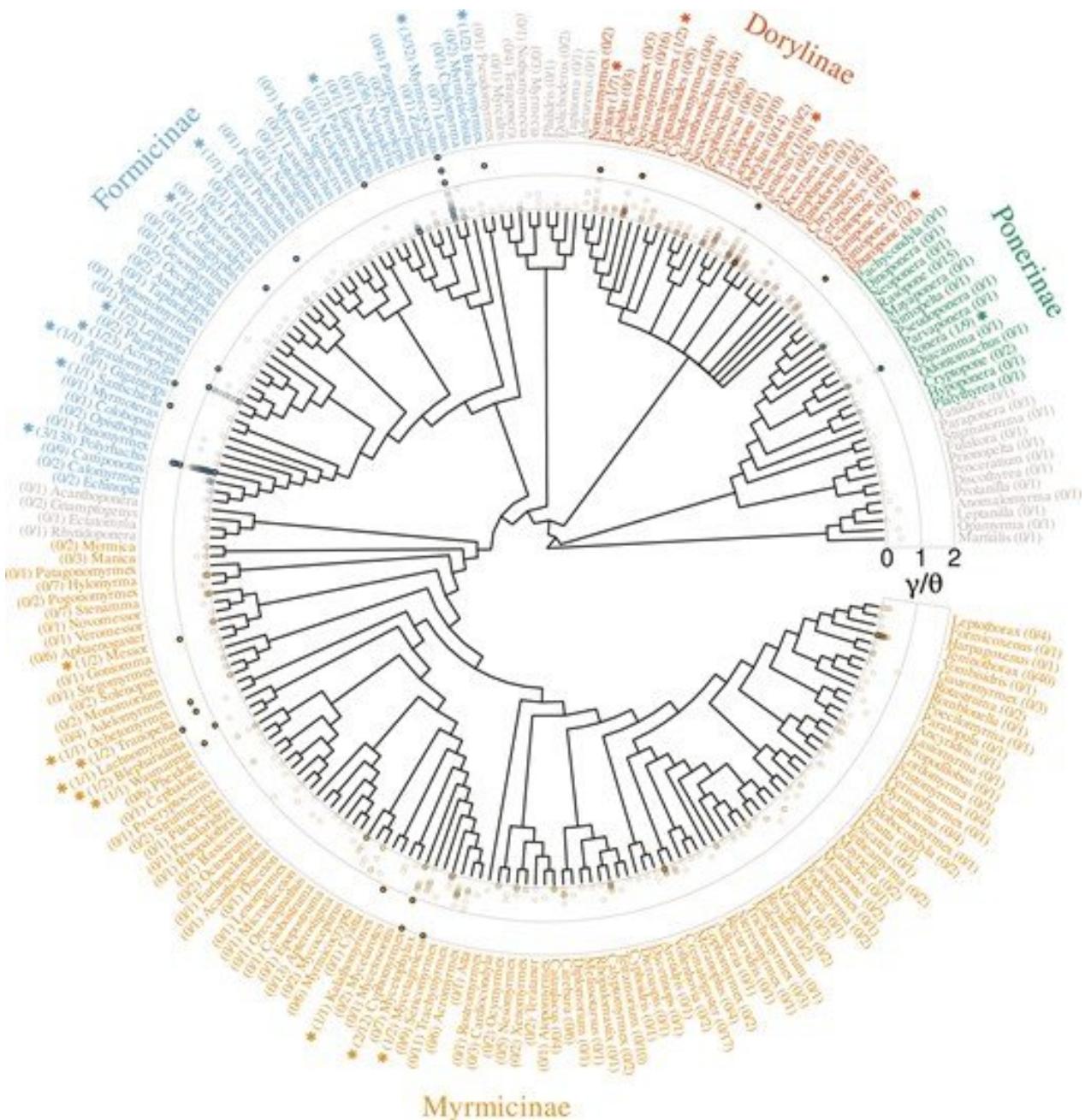


Figure 21 : Occurrence des hybrides F1 à travers les genres de Formicidae. Les estimations du rapport γ/θ obtenues chez les Formicidae sont représentées par rapport à la topologie des genres de ce groupe (extraites de Antwiki). Les genres comptant au moins une espèce avec $\gamma/\theta > 1$ (c'est à dire hybride F1 probable) sont mis en évidence par une étoile. Le nombre de ces espèces par genre, ainsi que le nombre total d'espèces par genre sont donnés pour chaque genre. Les rapports γ/θ supérieurs à deux ont été tronqués à deux pour des raisons de lisibilité. (Weyna et al. 2022).

Chapitre 2

Projet de recherche

2.1. Avant-propos

Mon projet de recherche s'inscrit dans les thématiques de mes axes de recherches précédemment évoqués, mais plus particulièrement encore sur mon 3ème axe de recherche : Déterminisme de caste et hybridogénèse sociale. Pour ce faire, j'utiliserai le modèle de la fourmis moissonneuse barbare pour lequel je viens d'obtenir un financement Starting Grant de l'European Research Council (ERC) pour les 4 prochaines années. Ce projet s'inscrit donc naturellement comme la colonne vertébrale de mon programme de recherche dans les années à venir. Bien que je détaillerai uniquement ce projet, je participe également à trois projets ANR dans lesquels je suis impliqué en tant que collaborateur.

Un projet qui perpétue mon intérêt pour les questions relatives aux liens entre évolution moléculaire et traits d'histoire de vie :

ANR NeGa (Porteur Tristan Lefébure) – Influence de la taille efficace des populations sur l'architecture des génomes animaux ([Résumé NeGa](#))

Deux projets qui prolongent mon intérêt pour les systèmes de reproduction atypiques chez les animaux :

ANR Pseudogamy (Porteuse Marie Delattre) : Origine et conséquences de l'asexualité chez les nématodes du genre *Mesorhabditis* ([Résumé PSEUDOGAMY](#))

ANR MINIGAN (Porteur Patrice David) : Conflits nucléocytoplasmiques et gynodioécie chez les animaux. ([Résumé MINIGAN](#))

Je détaille brièvement dans cet avant-propos le projet **MINIGAN**, dans lequel je co-encadre pour une thèse Fanny Laugier avec Patrice David :

La gynodioécie est un polymorphisme sexuel dans lequel des individus hermaphrodites coexistent avec des individus ayant complètement ou partiellement perdu la fonction mâle. L'origine évolutive de ce système sexuel résulte la plupart du temps d'un conflit entre génome mitochondrial et nucléaire, où la mitochondrie est porteuse d'un élément génétique induisant une stérilité mâle (on parle de stérilité mâle cytoplasmique). La mitochondrie se transmettant exclusivement de mère à fille, elle favorise ainsi sa transmission à plus de descendants, et cela au détriment des gènes nucléaires dont la moitié sont transmis par les mâles. Si ces systèmes sont bien connus chez les plantes à fleurs, nous travaillons sur le premier exemple d'animal connu, l'escargot d'eau douce *Physa acuta* qui possède des individus mâles-stériles avec un mitogénome extrêmement divergent (David et al. 2022).

Dans le cadre de la thèse de Fanny Laugier, j'encadre plus particulièrement la partie bioinformatique, avec le séquençage et l'assemblage d'un génome de référence pour *Physa acuta*, et le séquençage et l'analyse de génomes basse couverture de 300 individus répartis

dans 18 populations naturelles. L'objectif sera de mener des analyses de génomique des population afin d'identifier des gènes ou portions de génomes présentant des traces de sélection selon la proportion de mâles stériles dans les population étudiées. Du fait du conflit entre génome mitochondrial et nucléaire, on s'attend par exemple à une forte pression de sélection pour la fixation de gènes nucléaire restaurateurs de la fonction mâle dans les populations avec une proportion importante de mâles stériles.

Ci-dessous, je consacre l'intégralité du reste de ce chapitre à mon projet principal :

ERC RoyalMess (Porteur : Jonathan Romiguier) - Génomique évolutive de la royauté chez les fourmis moissonneuses du genre *Messor*

2.2. Projet RoyalMess : Résumé

Chez les insectes sociaux, la royauté est généralement acquise par l'environnement plutôt que déterminée génétiquement. Toutefois, ce n'est pas le cas dans un système de reproduction exceptionnel, connu sous le nom d'hybridogénèse sociale. Dans ce cas, deux lignées hybridogénétiques distinctes de fourmis ne produisent que des reines par elles-mêmes alors qu'elles doivent s'hybrider pour produire des ouvrières. Les œufs ayant un génome royal pur sont donc génétiquement destinés à devenir des reines, tandis que les génomes hybrides sont destinés à se développer en ouvrières. L'évolution vers un tel système de reproduction semble répétée chez plusieurs espèces de fourmis moissonneuses (4 fois dans les genres *Pogonomyrmex* et *Messor*), mais son origine est encore mystérieuse. Dans ce projet, je prévois d'utiliser ces espèces modèles uniques pour résoudre les déterminants génomique de la royauté :

Objectif 1. Prévalence de l'hybridogénèse sociale : Nous utiliserons des données génomique pour tester d'autres occurrences d'hybridogénèse sociale chez ~500 nouvelles espèces. Cela me permettra d'identifier si des déterminants écologiques potentiels (par exemple le climat ou le régime alimentaire) favorisent l'évolution de l'hybridogénèse sociale.

Objectif 2. Histoire des lignées hybridogénétiques : Nous allons utiliser la génomique des populations dans 3 paires de lignées hybridogénétiques de *Messor* pour dater et cartographier leur origine. Des méthodes ABC seront utilisées pour cartographier les flux génétiques entre lignées hybridogénétiques et espèces sœurs proches afin de comprendre le maintien de l'hybridogénèse sociale, la spéciation et l'introgression potentielle de gènes de caste.

Objectif 3. Identification des gènes déterminant la caste : Nous utiliserons la transcriptomique comparative dans les œufs de *Messor barbarus* pour identifier les gènes différentiellement exprimés entre les castes avant la divergence développementale. L'accouplement contrôlé entre des mâles recombinants produits en laboratoire et des reines sauvages + technologies CRISPR-cas9 seront utilisés pour valider expérimentalement les gènes candidats.

Ce projet pourrait révéler comment l'un des exemples les plus iconiques de plasticité phénotypique (phénotypes des castes d'ouvrières et de reines) peut devenir génétiquement déterminé par l'évolution répétée d'un système de reproduction fascinant.

2.3. Contexte

La part d'inné et d'acquis dans le devenir d'un individu est une vaste question qui taraude régulièrement philosophes et scientifiques. En biologie, cet épineux problème peut se résumer par la part relative qu'ont les gènes et l'environnement sur le phénotype ou le comportement d'un organisme. La variation morphologique intra-spécifique est particulièrement marquée chez les insectes eusociaux. Divisés en castes à la morphologie, physiologie et comportements bien distincts, les insectes eusociaux arborent une impressionnante diversité de formes entre individus d'une même espèce. De manière remarquable, cette diversité phénotypique ne se limite pas au dimorphisme sexuel. Si un oeuf non-fécondé d'hyménoptère eusocial se développe quasi-invariablement en mâle haploïde (on parle d'haplo-diploïdie), un oeuf fécondé a deux destinées femelles possibles : participer à la reproduction de la colonie en tant que reine, ou s'acquitter des soins et de la recherche de nourriture en tant qu'ouvrière.

Les déterminants qui influencent la différentiation en une caste stérile ou fertile ont longtemps été jugés exclusivement environnementaux. Les premiers arguments sont tout d'abord théoriques : il est en effet difficile d'imaginer la persistance évolutive d'un variant génétique induisant la stérilité (Hölldobler & Wilson 1990). Un allèle qui aurait pour effet systématique de réduire drastiquement ou complètement la fécondité ne pourrait que difficilement se maintenir dans une population, les femelles le portant ne pouvant pas avoir de descendance. Pour qu'un système de caste stérile puisse se maintenir, il est plus simple d'imaginer un gène ouvrière qui ne serait exprimé que conditionnellement à l'environnement, soit une forme de plasticité phénotypique (Evans et Wheeler 2001, West-Eberhard 1989). Le fameux exemple de l'ingestion de gelée royale induisant la différentiation d'une larve d'abeille domestique (*Apis mellifera*) en reine a finit d'imposer l'idée de différentiation environnementale de la caste (Corona et al. 2007, Smith et al. 2008, Simpson et al. 2012).

L'apport de la génétique a cependant remis en cause le dogme de détermination environnementale de la caste. L'influence directe ou indirecte d'effets génétiques ont en effet été démontrés chez plusieurs populations ou espèces, de telle sorte qu'on parle désormais de continuum de détermination génétique à environnementale de la caste (Schwander et al. 2010). Dans la plupart des espèces connues où le déterminisme génétique est quasi-total, une évolution vers la production asexuée des reines est impliquée : un oeuf fécondé ne produit dès lors plus que des ouvrières (Pearcy et al. 2004, Fournier et al. 2005, Ohkawara et al. 2006, Pearcy et al. 2011, Leniaud et al. 2012). Un autre exemple qui a suscité beaucoup d'intérêt dans la littérature est celui d'un système qui conserve la reproduction sexuée pour la production des deux castes femelles : un oeuf fécondé peut toujours se développer en reine ou en ouvrière, mais c'est l'interaction du génome paternel et maternel qui détermine ce choix. Un tel système est ainsi observable au sein de certaines populations d'une zone de contact hybride de fourmis moissonneuses nord-américaines du genre *Pogonomyrmex*

(Helms Cahan & Keller 2003). Dans ces populations, deux lignées dépendantes l'une de l'autre co-existent et se reproduisent sans jamais se mélanger. Lorsque l'oeuf d'une reine est fécondé avec le spermatozoïde d'un mâle de sa lignée, l'oeuf se différencie en reine sexuée. *A contrario*, lorsque l'oeuf d'une reine est fécondé par le spermatozoïde d'un mâle de lignée différente, l'oeuf se différencie en ouvrière (voir Fig. 22). Ce système de reproduction peu orthodoxe a été nommé hybridogénèse sociale : les croisements intra-lignée produisent des reines, les croisements inter-lignée produisent des ouvrières.

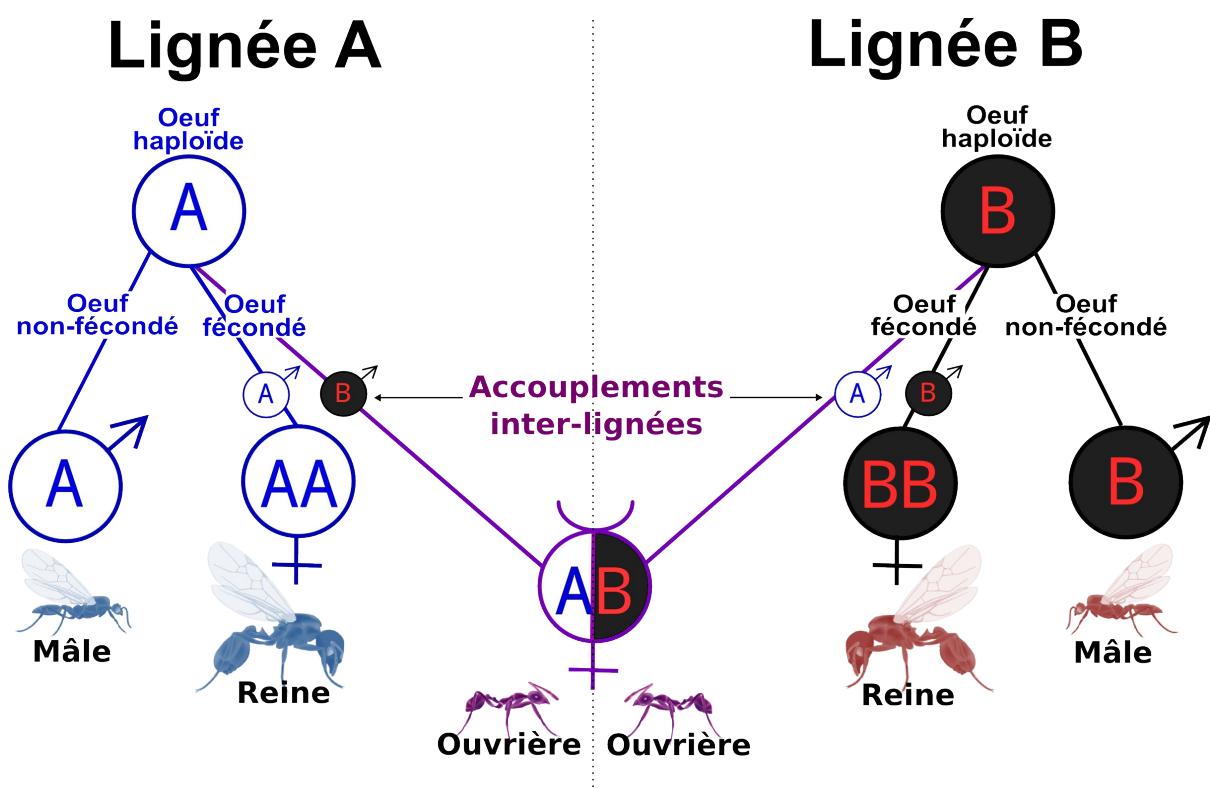


Figure 22 : Déterminisme génétique des castes chez les fourmis moissonneuses

L'hybridogenèse sociale symétrique, combinant des relations symétriques entre deux lignées hybridogénétiques et la production sexuelle de reines, a longtemps été considérée comme anecdotique et limitée aux populations hybrides de *Pogonomyrmex rugosus* et *Pogonomyrmex barbatus* (Helms Cahan et Keller 2003, Schwander et al. 2007, Schwander et al. 2008, Linksvayer et al. 2006, Anderson et al. 2006, Helms Cahan et al. 2022). Cependant, j'ai récemment montré que l'hybridogenèse sociale a évolué de manière convergente à trois reprises chez *Messor*, un autre genre de fourmis moissonneuses (Romiguier et al. 2016). De façon remarquable, ces trois cas ont été découverts à partir des données génétiques de seulement neuf espèces, et plus de 100 espèces de *Messor* doivent encore être étudiées. Cela soulève la question de la prévalence réelle et de l'origine évolutive d'un système reproductif aussi inhabituel.

À court terme, les avantages évolutifs de l'hybridogenèse sociale pourraient être doubles : i) la reine optimise égoïstement la transmission des gènes de sa propre lignée en les transmettant uniquement à la caste fertile, ii) la reine exploite le sperme d'une espèce apparentée (kleptogamie ou parasitisme du sperme, Umphrey 2006) et profite des hybrides stériles (ouvrières) pour former une colonie présentant une diversité génétique accrue et une dépression de consanguinité réduite, ce qui peut augmenter les performances globales de la colonie en conférant des avantages tels qu'une meilleure résistance aux pathogènes (Mattila et Seeley 2007, Olroyd et Fewell 2007). Malgré près de deux décennies de recherche sur ce sujet, le scénario exact menant à un tel système de reproduction reste inconnu. L'objectif principal de ce projet est de comprendre comment une transition évolutive aussi remarquable peut émerger et se maintenir.

2.4. Objectifs

La résolution du scénario évolutif menant à ce système de reproduction alambiqué ne se limite pas à l'étude de l'eusocialité mais englobe plusieurs questions évolutives majeures d'intérêt général :

- Comment la plasticité phénotypique peut-elle devenir génétiquement déterminée (le concept dit "d'assimilation génétique", Waddington 1953) ? Le dimorphisme reine/ouvrière est l'un des exemples les plus impressionnantes de plasticité phénotypique dans la nature. Le genre de fourmis moissonneuses *Messor* offre un cadre unique avec au moins trois (et probablement beaucoup plus) répliques naturelles d'espèces ayant un système de castes génétiques, qui peuvent être directement comparées à des espèces étroitement apparentées ayant un système de castes environnementales (Romiguier *et al.* 2016).
- Comment la reproduction sexuelle classique peut-elle évoluer vers des systèmes de reproduction nécessitant plus de deux partenaires sexuels ? L'hybridogenèse sociale a été interprétée comme une transition évolutive majeure vers plus de deux sexes (Parker 2004). Une reine ne peut pas élever de reines filles sans ouvrières et doit donc s'accoupler avec deux types de mâles différents pour produire des colonies filles. Quatre types parentaux sont nécessaires à la persistance de ce système, qui est unique chez les animaux et reste à décrire du point de vue évolutif.
- Comment empêcher l'accumulation d'incompatibilités entre deux génotypes divergents ? Comme les deux lignées hybridogénétiques ne peuvent pas produire d'hybrides fertiles (Fig. 22), elles peuvent être considérées comme deux pools génétiques isolés (c'est-à-dire deux espèces) du point de vue de la reine/mâle, tout en fonctionnant comme une seule entité génétique au niveau des ouvrières. On s'attend à ce que les lignées divergentes accumulent des mutations qui devraient nuire à la viabilité des hybrides à long terme, ce qui soulève la question du maintien à long terme de l'hybridogenèse sociale (Anderson *et al.* 2008). La production parthénogénétique occasionnelle de mâles recombinants par des ouvrières (démontrée dans des colonies de laboratoire sans reine Romiguier *et al.* 2016) pourrait

induire un peu de flux génétique entre les lignées hybridogénétiques et résoudre cette impasse évolutive. Une autre question connexe est de savoir comment un tel complexe d'espèces hybridogénétique évolue en cas d'isolement géographique et de spéciation ultérieure.

Ces questions seront explorées à travers trois "objectifs" principaux à trois échelles taxonomiques centrées sur la fourmi moissonneuse *Messor barbarus* (voir Fig 23), une espèce méditerranéenne répandue et écologiquement dominante, qui est considérée comme un ingénieur majeur de l'écosystème pour structurer ou restaurer les communautés végétales (Bulot *et al.* 2014, De Almeida *et al.* 2020).

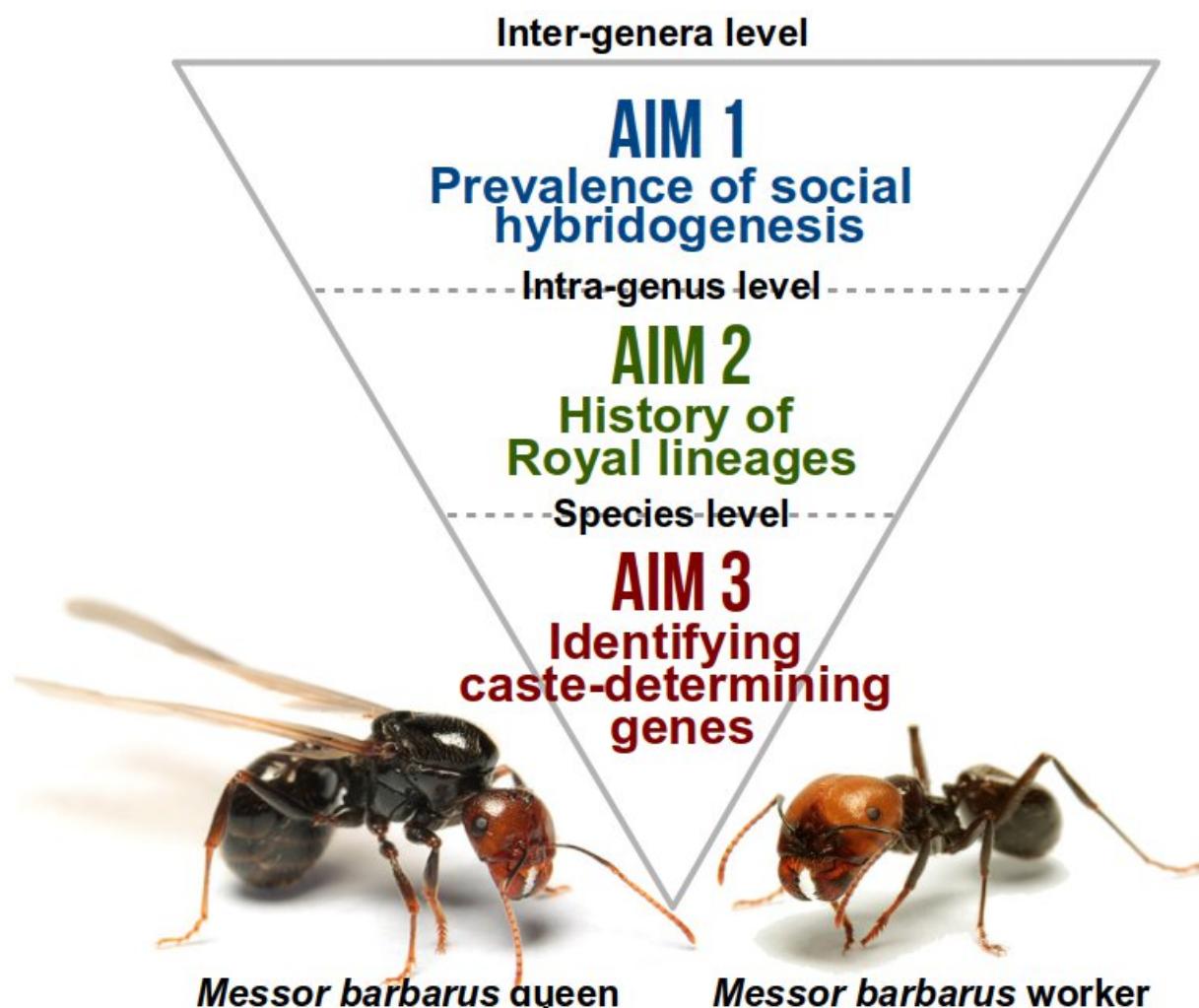


Figure 23 : Résumé graphique des objectifs du projet

- L'objectif 1 se situera à un **niveau taxonomique de la tribu** et étudiera la prévalence de l'hybridogenèse sociale dans l'ensemble du genre *Messor* et dans sept autres genres apparentés dans la tribu des *Stenammini*.

- L'**objectif 2** se situera à un **niveau taxonomique intra-genre** et retracera précisément l'origine géographique et génomique de trois paires de lignées hybridogénétiques ainsi que leur réseau de flux géniques avec trois espèces proches.
- L'**objectif 3** sera centré au **niveau de l'espèce** et tentera d'identifier plus directement les gènes de la royauté par des approches expérimentales telles que des analyses d'expression différentielle aux premiers stades larvaires, l'accouplement contrôlé de mâles recombinants produits en laboratoire et l'édition génétique via CRISPR-Cas9 pour manipuler le développement des castes.

En donnant une vision holistique de l'hybridogenèse sociale via trois niveaux taxonomiques complémentaires, le projet RoyalMess vise à comprendre comment la plasticité phénotypique peut devenir génétiquement déterminée, d'une large échelle macro-évolution à une fine échelle fonctionnelle. En cas de succès, ce projet pourrait permettre la première manipulation génétique de la royauté des fourmis, un exemple emblématique de l'évolution phénotypique vers une longévité et une fécondité extrêmes.

2.5. Méthodes

2.5.1. Objectif 1 - Prévalence de l'hybridogenèse sociale

L'hybridogenèse sociale non-clonale a été détectée chez trois espèces de *Messor* (Romiguier *et al.* 2016) et chez quatre paires de lignées dans des populations hybrides de *Pogonomyrmex* (Helms Cahan 2022). Il est intéressant de noter que *Messor* et *Pogonomyrmex* sont deux genres de fourmis moissonneuses non apparentés qui présentent d'importantes convergences écologiques : tous deux présentent des spécialisations liées à la granivorie et vivent dans des habitats arides. Il a été proposé que l'écologie de ces espèces puisse faciliter les transitions vers l'hybridogenèse sociale (Romiguier *et al.* 2016).

- H1) Régime alimentaire. L'alimentation différentielle des larves par les ouvrières contrôle classiquement le déterminisme environnemental de la caste (Haydak 1943, Kamakura 2011), donc un régime alimentaire granivore spécialisé pourrait avoir limité ce contrôle environnemental et facilité l'évolution du déterminisme génétique de la caste. Chez une espèce de *Pogonomyrmex* ayant un régime mixte graines/insectes, les larves consommant des proportions plus élevées de protéines (c'est-à-dire d'insectes) par rapport aux graines ont tendance à se développer en reines, ce qui semble analogue à la gelée royale riche en protéines dans la détermination des castes des abeilles domestiques. Un régime alimentaire pauvre en protéines en cas de granivorie stricte pourrait diminuer la capacité des ouvrières à contrôler le déterminisme des castes par une alimentation différentielle des larves, ce qui augmenterait l'effet d'un facteur génétique affectant la détermination des castes. Cela pourrait ouvrir la voie à l'évolution d'un allèle de caste conduisant à l'hybridogenèse sociale (Weyna *et al.* 2021).

- H2) Milieu aride. Une contrainte majeure de l'hybridogenèse sociale est que les reines doivent s'accoupler au moins une fois avec un mâle de chaque lignée hybridogénétique pour

produire à la fois des ouvrières et de nouvelles reines, ce qui devrait être plus facile en cas de vols massifs synchrones entre les colonies. Les pluies d'été (lorsque la température est élevée et que le nid est facile à creuser) sont le déclencheur environnemental le plus typique pour les essaimages, donc les habitats arides avec des pluies rares devraient produire des essaimages plus rares, plus massifs et plus synchrones entre colonies d'une même population. A l'appui de cette hypothèse, les fourmis *Pogonomyrmex* et *Messor* se rencontrent principalement dans les climats méditerranéens et sont connues pour avoir des vols nuptiaux massifs et hautement synchronisés (Markl *et al.* 1977, Davidson 1982, Gomez et Abril 2012).

Soutenant encore ces hypothèses, la fourmi de feu *Solenopsis xyloni* parfois impliquée dans des systèmes d'hybridogenèse sociale asymétrique (une étape intermédiaire probable vers l'hybridogenèse sociale, Helms Cahan et Vinson 2003) vit typiquement dans des habitats arides et consomme de grandes quantités de graines (Valone et Kaspari 2005, Tschinkel 2006). Compte tenu des quelques exemples connus à ce jour, il est donc difficile de savoir dans quelle mesure ce système de reproduction remarquable est répandu et s'il est lié à l'écologie d'une espèce spécifique. Trois espèces connues de *Messor* pratiquent l'hybridogenèse sociale, mais seules neuf espèces sur les 113 que compte le genre ont été étudiées jusqu'à présent (Romiguier *et al.* 2016). Nous utiliserons des données génomiques pour documenter les nouvelles occurrences d'hybridogenèse sociale dans l'ensemble des genres *Messor* et *Pogonomyrmex* ainsi que dans sept genres apparentés, pour un total de 499 espèces cibles comprenant quatre convergences indépendantes vers la granivorie et divers habitats.

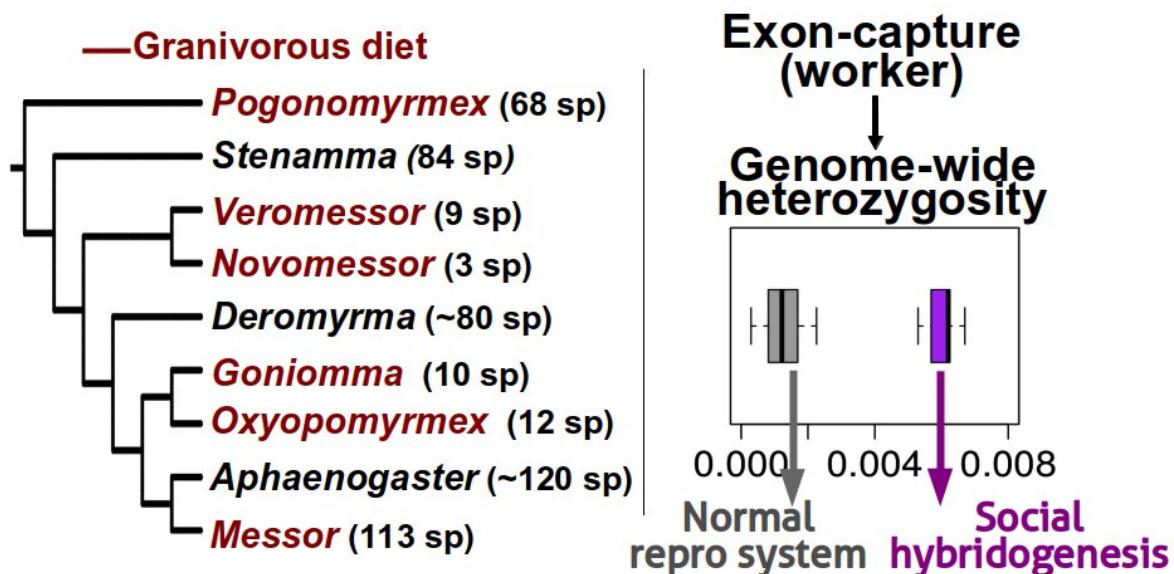


Figure 24 : Résumé graphique de l'Objectif 1

Étape 1 : Une analyse phylogénétique préliminaire pour l'hybridogenèse sociale sera d'abord effectuée sur jusqu'à 499 espèces (Fig 24).

Stratégie d'échantillonnage

La plupart des espèces cibles existent sous forme d'échantillons de musée, et il est possible d'effectuer des séquençages non destructifs de spécimens anciens, qu'ils soient conservés dans de l'alcool ou à sec (Castalanelli *et al.* 2010, Tin *et al.* 2014, Santos *et al.* 2018). La base de données AntWeb (www.antweb.org/) compile les collections de spécimens de fourmis de 11 musées et contient la localisation de 408 des 499 espèces cibles. Je collabore actuellement avec son administrateur (Brian L. Fisher) avec qui j'ai collaboré récemment (Romiguier *et al.* 2022) et ai entamé des premiers tests avec succès sur une soixantaine de spécimen du Museum d'Histoire Naturelle de Londres.

Stratégie de séquençage

L'hétérozygotie à l'échelle du génome sera évaluée par séquençage de capture d'exon au lieu du séquençage du transcriptome utilisé dans mon ancien travail sur les *Messor* (Romiguier *et al.* 2016). La capture d'exon est ici idéale car elle fournit des quantités comparables de séquences codantes pour une estimation fiable de l'hétérozygotie, tout en nécessitant de l'ADN au lieu de l'ARN, ce qui permettra l'échantillonnage de spécimens de musée. Des membres de mon équipe de laboratoire (Rousselle *et al.* 2020) ont réussi avec succès à concevoir des sondes personnalisées pour le genre *Formica* vieux de 44,1 millions d'années (Romiguier *et al.* 2018) à partir d'un seul transcriptome. Sur la base de ces chiffres et de la divergence temporelle entre les espèces cibles, les données déjà séquencées de *Messor barbarus* (Romiguier *et al.* 2016), *Pogonomyrmex barbatus* (Smith *et al.* 2011) et *Aphaenogaster subterranea* (disponibles sur la base de données ncbi SRA id SRX2960342) devraient être suffisants pour concevoir les sondes de toutes mes espèces cibles. Si nécessaire, des échantillons de *Stenamma*, *Novomessor* et *Goniomma*, congelés et stockés dans un réfrigérateur à -80, sont disponibles pour un séquençage de transcriptome et pourraient être utilisés pour concevoir des sondes plus spécifiques.

Analyses

La détection des espèces pratiquant l'hybridogenèse sociale se fera en calculant l'hétérozygotie d'une ouvrière à l'échelle du génome : les résultats préliminaires sur les transcriptomes issus de projets passés et en cours (Romiguier *et al.* 2014, 2016) indiquent que chez *Messor*, les ouvrières des espèces hybridogénétiques ont une hétérozygotie plus de deux fois supérieure (moyenne de 0,006 chez 17 individus, 3 espèces de *Messor*) par rapport aux ouvrières des espèces non-hybridogénétiques (moyenne de 0,00125, 52 individus, 22 espèces dont 7 espèces de *Messor* et 16 espèces des genres *Pheidole*, *Camponotus*, *Formica*, *Polyergus*, *Crematogaster*, *Temnothorax* et *Myrmica*) (Fig. 24). Une telle distribution bimodale et sans chevauchement permet de discriminer clairement les ouvrières hybridogénétiques, d'autant plus que les fourmis non-hybridogénétiques testées ici semblent ne jamais avoir d'hétérozygoties supérieures à 0,0025. La méthode développée par Weyna *et al.* (2022) sera utilisée pour les cas plus intermédiaires (lignées hybridogénétiques moins divergentes ou espèces non-hybridogénétiques plus hétérozygotes).

Étape 2 : Le séquençage des reines sera effectué chez les espèces identifiées comme potentiellement hybridogénétiques à l'étape 1.

Si l'hétérozygotie des reines est confirmée comme étant faible par rapport à celle des ouvrières, un échantillonnage supplémentaire sera effectué. Au moins une reine par lignée hybridogénétique devra être identifiée par espèce candidate pour une confirmation formelle de l'hybridogénèse sociale. Les espèces de *Messor* et les ouvrières les plus hybrides seront sélectionnés en priorité en cas d'un trop grand nombre d'espèces candidates. Toutes les données seront ensuite utilisées pour construire un arbre phylogénétique presque exhaustif décrivant les relations entre espèces et lignées hybridogénétiques. La phylogénie de la Figure 24 et l'estimation du nombre d'espèces sont basées sur AntWeb et des études concluant que le grand genre *Aphaenogaster* de 200 espèces est polyphylétique et divisé en au moins deux clades (nommés ici *Aphaenogaster* et *Deromyrma* avec des nombres d'espèces spéculatifs, Branstetter *et al.* 2022), ce qui sera confirmé ou pas dans notre nouvelle phylogénie. La prévalence de l'hybridogenèse sociale sera également comparée à des facteurs écologiques tels que le régime alimentaire et l'habitat. Six genres ont un régime granivore, tandis que les trois genres restants sont plus généralistes. L'habitat, l'aire de répartition et les conditions climatiques peuvent être facilement trouvés pour chaque espèce grâce à l'impressionnant travail de compilation (8811 publications, 25 bases de données existantes) de la base de données Global Ant Biodiversity Informatics (Janicki *et al.* 2016).

Impact

Il s'agira de la première analyse à grande échelle de l'hybridogenèse sociale dans plusieurs genres. Elle pourrait conduire à la découverte de plusieurs dizaines de nouveaux cas d'espèces hybridogénétiques et fournira la première image globale de leur prévalence évolutive. Les clarifications taxonomiques et phylogénétiques nécessaires nous aideront à comprendre comment les lignées hybridogénétiques sont liées les unes aux autres et aux autres espèces, donnant un aperçu de leur origine, de leur stabilité et de leurs modes de spéciation. Ce travail offrira également l'opportunité de vérifier si les transitions vers la détermination génétique des castes sont liées à des facteurs écologiques, tels que le régime alimentaire ou l'habitat. En outre, les données génétiques sur les nouveaux cas d'hybridogenèse sociale aideraient considérablement à identifier les gènes candidats responsables de la détermination de la caste génétique (par exemple en identifiant des sites d'hétérozygotes constant parmi les ouvrières de plusieurs espèces hybridogénétiques ou via des méthodes de détection de convergence moléculaire, Rey *et al.* 2019).

2.5.2. Histoire des lignées hybridogénétiques

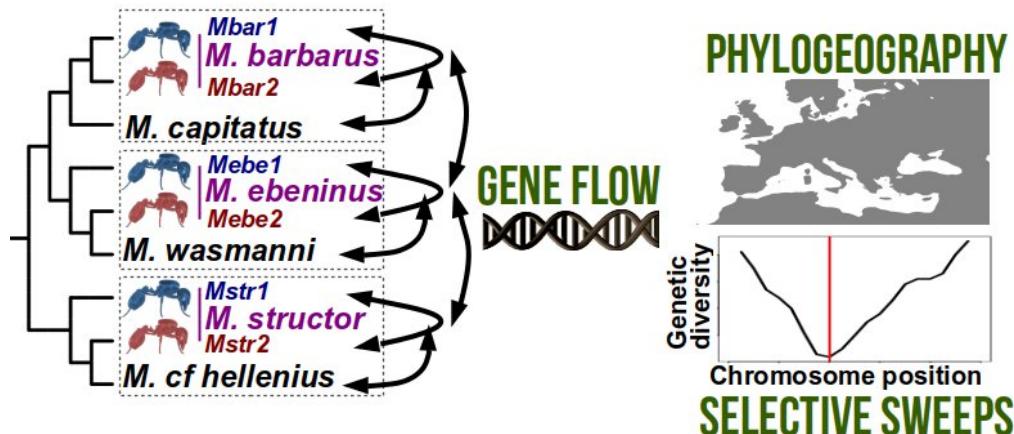


Figure 25 : Résumé graphique de l'Objectif 2

Savoir comment, où et quand une paire de lignée hybridogénétique a commencé sa divergence est une condition préalable pour valider ou réfuter un scénario pour l'origine de l'hybridogenèse sociale (voir Fig. 25). Ont-elles d'abord divergé en raison d'un isolement géographique (refuge glaciaire différent ?) ou en tant que populations sympatriques ? Sont-elles complètement isolées ou échangent-elles du matériel génétique par le biais de flux géniques occasionnels ? Les réponses à ces questions permettront d'étayer ou de rejeter les scénarios évolutifs candidats conduisant à l'hybridogenèse sociale. Je présente ci-dessous deux hypothèses générales suggérées dans la littérature.

Hypothèse 1 :

Un premier scénario suggère que les lignées hybridogénétiques de *Pogonomyrmex* sont apparues de manière sympatrique après l'hybridation de deux espèces parentales normales (Helms Cahan *et al.* 2003). Des niveaux élevés de dérive génétique causés par la petite taille de la population effective de ces hybrides auraient entraînés la fixation de deux ensembles alternatifs d'allèles épistatiques incompatibles (A1A1 B2B2 et A2A2 B1B1, voir Fig 26) qui ne peuvent pas déclencher le développement des ouvrières indépendamment. Pour restaurer la production d'ouvrières, les deux lignées devraient alors se croiser, comme c'est le cas dans l'hybridogenèse sociale. Soutenant cette hypothèse chez *Messor*, les hybrides interspécifiques et les colonies mixtes semblent communs dans le genre (Steiner *et al.* 2011). En outre, une histoire évolutive complexe et des flux génétiques entre les lignées hybridogénétiques et d'autres espèces sont probables, comme l'illustrent deux espèces hybridogénétiques (*M. structor* et *M. ebeninus*) où les lignées hybridogénétiques peuvent être des espèces sœurs d'espèces non hybridogénétiques (voir Fig 4). Les modèles génomiques soutenant l'hypothèse 1 (voir Fig 26) :

- **Événements d'hybridation entre l'ancêtre des lignées hybridogénétiques et une espèce étroitement apparentée**
- **Goulot d'étranglement démographique (diminution homogène de la diversité génétique)**
- **Divergence sympatrique des lignées hybridogénétiques.**

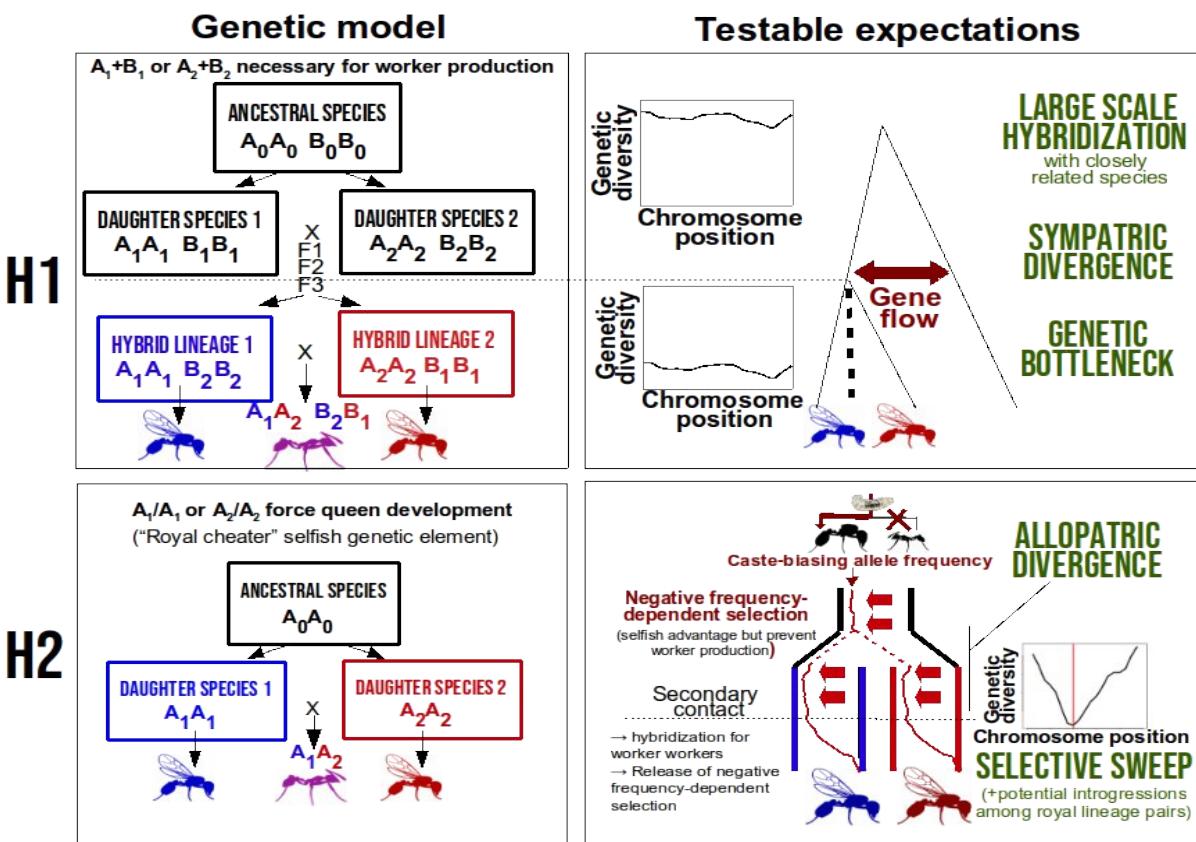


Figure 26 : Modèles génétiques et attendus les hypothèses 1 et 2 pour l'origine évolutive de l'hybridogénèse sociale

Hypothèse 2 :

Un scénario concurrent suggère que les deux lignées hybridogénétiques ont d'abord divergées en allopatrie et que la détermination génétique de la caste est apparue à cause d'un allèle biaissant le développement larvaire vers les reines (Anderson *et al.* 2008, Weyna *et al.* 2021) (A1A1 ou A2A2, voir Fig 26). Étant surreprésenté dans la caste fertile, un tel allèle favorise égoïstement sa propre propagation (Hughes et Boomsma 2008). On s'attend à ce qu'un tel élément génétique égoïste soit présent à une faible fréquence dans une population donnée, car sa fixation empêcherait complètement la production d'ouvrières. Cependant, la libération d'une telle sélection négative fréquence-dépendante est théoriquement possible si une lignée peut contourner l'impossibilité de produire des ouvriers en utilisant le sperme d'une lignée/espèce divergente, phénomène qualifié de parasitisme du spermatique (Umphrey 2006). Après un contact secondaire avec une lignée/espèce divergente et une hybridation fréquente produisant des ouvrières, on peut s'attendre à la fixation rapide d'un allèle biaiseur de caste, scénario que nous avons d'ailleurs récemment validé théoriquement (Weyna *et al.* 2021) Un modèle génétique alternatif de ce scénario suggère qu'un complexe cyto-nucléaire co-évolué peut biaiser le développement larvaire vers la reine (A1A1 mito1 et A2A2 mito2) et que les incompatibilités épistatiques dans les hybrides (A1A2 mito1 ou A1A2

mito2) forcent le développement en ouvrières (Linksvayer *et al.* 2006). L'incapacité à produire des ouvrières pour une seule des deux lignées en combinaison avec le parasitisme spermatique d'une espèce étroitement apparentée pourrait avoir été une étape intermédiaire vers l'hybridogénèse sociale, et a été effectivement observée dans les zones hybrides de deux fourmis de feu (Helms Cahan et Vinson 2003). Dans une variante de cette hypothèse, il a été suggéré que l'allèle biaiseur de caste aurait pu être introgressé de la lignée parasite à la lignée parasitée ou de paires de lignées hybridogénétiques à d'autres paires de lignées hybridogénétiques du même genre (Anderson *et al.* 2008). Des variantes de ce scénario incluent également des mutations indépendantes de-novo de biais de caste après la divergence des lignées qui peuvent être sur le même ou différents locus. Quel que soit le scénario exact, la fixation d'un tel allèle de caste forçant le développement larvaire vers des reines devrait être analogue à la fixation d'un élément génétique égoïste. Il a été démontré que la fixation d'un élément génétique égoïste est rapide et analogue à la fixation d'un allèle sélectionné positivement (Derome *et al.* 2004, Presgraves *et al.* 2009, Didion *et al.* 2016), qui est classiquement caractérisée par un balayage sélectif, c'est-à-dire une diminution locale de la diversité génétique (voir Fig 25 et 26). Par conséquent, le transfert horizontal d'un biaiseur de caste égoïste serait caractérisé par une introgression suivie d'un balayage sélectif. Ce scénario où la détermination génétique des castes peut être transmise entre espèces/lignées pourrait expliquer pourquoi l'hybridogénèse sociale est apparue à plusieurs reprises chez *Messor* et *Pogonomyrmex*.

Les modèles génomiques soutenant l'hypothèse 2 (voir Fig 26) :

- Divergence allopatrique des lignées hybridogénétiques
- Fixation rapide d'allèles égoïstes biaisant la caste conduisant à des balayages sélectifs (+ introgression potentielle entre les lignées et les paires de lignées hybridogénétiques).

Au-delà de l'origine de l'hybridogénèse sociale, les mécanismes de son maintien à long terme pourraient être démontrés via l'inférence d'un flux de gènes entre lignées hybridogénétiques qui limite leur divergence et assure leur compatibilité à long terme pour la production d'ouvrières hybrides. Ici, nous proposons de produire des données de génomique des populations pour trois trios (une paire de lignées hybridogénétiques + l'espèce la plus proche) dans le genre *Messor* et de développer de nouvelles méthodologies pour déduire simultanément la phylogéographie, les flux de gènes et les balayages sélectifs (c'est-à-dire la réduction de la diversité génétique près d'un allèle sélectionné atteignant la fixation) suite à l'introgression potentielle de gènes de caste entre lignées.

Étape 1 : 25 individus par lignée hybridogénétique/espèce seront séquencés en trois trios (deux lignées hybridogénétiques + parent le plus proche).

Stratégie d'échantillonnage

Nous allons échantillonner trois trios d'espèces/lignées : *M. barbarus* (Mbar1/Mbar2) + *M. capitatus*, *M. structor* (Mstr1/Mstr2) + *M. cf hellenius* et *M. ebeninus* (Mebe1/Mebe2) + *M. wasmanni* (voir Fig 25). Ces 225 individus seront collectés de manière à représenter au mieux leur aire de distribution native respective. Dans la mesure du possible, j'essaierai obtenir des individus fraîchement collectés. Ces espèces de *Messor* sont communes, faciles

à repérer et sont principalement distribuées en Europe méditerranéenne (à l'exception de *M. ebeninus*). Sur la base des occurrences des espèces, les pays potentiels pour des collaborations ou du travail de terrain seront la France, l'Espagne, l'Italie, la Croatie, la Grèce, la Turquie, le Liban, Israël, le Maroc, l'Algérie, la Tunisie, l'Egypte. J'ai déjà rassemblé 53 des individus requis (principalement d'Europe occidentale et du Maroc). D'après mon expérience, les *Messor* sont parmi les fourmis les plus faciles à échantillonner, les colonies étant faciles à repérer avec de grands groupes d'ouvrières en quête de graines dans des zones ouvertes près des routes et des villes.

Stratégie de séquençage

Pour cartographier précisément l'hétérogénéité des flux génétiques et les balayages sélectifs potentiels, des génomes entiers seront séquencés via Illumina (couverture de 60X, 150 paires de paires) et alignés sur un assemblage de référence de haute qualité que je suis en train de séquencer (séquençage PacBio long-reads + séquençage Illumina short-reads) pour chaque espèce cible dans le cadre de l'appel à projet du consortium GAGA (Boomsma *et al.* 2017). Les génomes entiers sont ici absolument nécessaires pour détecter les introgressions potentielles d'un élément égoïste de caste et les balayages sélectifs ultérieurs dans les régions génomiques voisines. Pour les espèces non-hybridogénétiques, n'importe quelle caste peut être utilisée et traitée avec une préparation classique de bibliothèque Illumina (TrueSeq). Pour les lignées hybridogénétiques, seuls les mâles/reines des colonies peuvent être utilisés de la même manière. Un risque important est qu'il soit impossible d'échantillonner les sexués dans toutes les populations, car ils peuvent être délicats à collecter selon la saison ou l'espèce. Pour surmonter ce défi, deux ouvrières hybridogénétiques d'une population ne disposant pas d'individus sexués seront échantillonnées et traitées via un séquençage long-reads ou type Chromium 10X (technologie de séquençage en phase permettant l'identification des haplotypes paternels et maternels), ce qui permettra d'échantillonner les deux lignées royales en même temps en collectant simplement des ouvrières hybrides de n'importe quelle colonie. Cette stratégie de séquençage va considérablement faciliter l'échantillonnage.

Ajustements potentiels de la stratégie d'échantillonnage

Le choix des espèces dépend de notre compréhension actuelle de la phylogénie de *Messor*, compréhension qui pourrait évoluer avec la phylogénie et la révision de la taxonomie de l'Objectif 1. Les trois espèces hybridogénétiques cibles ne seront à priori pas modifiées, car nous sommes sûrs qu'elles se trouvent dans des clades séparés et qu'elles font partie des espèces de *Messor* les plus répandues et les plus faciles à échantillonner. En fonction des premiers résultats de l'Objectif 1, d'autres espèces non hybridogénétiques pourraient être échantillonnées et séquencées, en particulier les espèces étroitement liées à *M. structor* (Steiner *et al.* 2018). Le budget de séquençage est conçu pour être flexible, et jusqu'à huit espèces non-hybridogénétiques supplémentaires peuvent être ajoutées pour le même prix en réduisant le nombre d'individus échantillonnés à 20 par espèce/lignée.

Étape 2 : Développement d'un nouvel outil informatique pour déduire simultanément flux génétiques et balayages sélectifs.

Cette étape sera réalisée à l'aide d'approches ABC (Approximate Bayesian Computation), une approche qui a démontré sa capacité à quantifier l'étendue des barrières reproductives entre des populations divergentes (Roux *et al.* 2016). En génomique des populations, une approche ABC est un cadre statistique inférentiel qui classe un ensemble de scénarios évolutifs en fonction de leur capacité à reproduire les modèles génomiques observés. L'ABC permet ainsi de calculer les probabilités postérieures relatives de divers scénarios d'évolution, par exemple un scénario simple où du flux génétique existe entre deux lignées hybridogénétiques face un scénario où les lignées sont totalement isolées. En pratique, ce type d'analyse implique de nombreuses simulations qui seront comparées aux données observées à l'aide d'un ensemble de statistiques résumées. Par exemple, la statistique ABBA-BABA, qui est très sensible à l'asymétrie des événements d'introgression, sera calculée et utilisée pour les inférences (Martin *et al.* 2014). Je collabore et continuerai à collaborer activement avec C. Roux qui développe et améliore ces outils ABC pour la génomique des populations. La méthode sera d'abord adaptée à la structure en trio de l'échantillonnage en intégrant plus d'une paire de populations/espèces divergentes. La détection par balayage sélectif sera ensuite intégrée. Une version préliminaire de la détection de balayages sélectifs via ABC est déjà fonctionnelle et donne des résultats prometteurs avec des validations par simulation. La validation empirique et les ajustements seront menés jusqu'à l'objectif final, qui consiste à tester explicitement divers scénarios d'origine de l'hybridogenèse sociale. Par exemple, plusieurs variations du scénario 1 seront comparées à plusieurs variations du scénario 2 (représentation graphique simple dans la Fig 26) pour classer leur capacité à produire les motifs observés le long de plusieurs fenêtres glissantes des génomes de lignées hybridogénétique. Les résultats des trois trios seront comparés avant de tester le flux génétique entre eux pour vérifier si l'hybridogenèse sociale a pu être transmise horizontalement.

Impact

L'application de ces méthodes devrait permettre d'aider à élucider l'origine mystérieuse de l'hybridogenèse sociale en révélant dans trois études de cas répliquées la provenance géographique et génomique des lignées hybridogénétiques. Dans le cadre de l'Objectif 2, je vais également retracer le transfert d'une espèce à l'autre d'éléments génétiques égoïstes potentiels qui biaisent le développement des castes, et plus généralement caractériser l'introgression entre les lignées hybridogénétiques et non-hybridogénétiques.

2.5.3. Objectif 3 - Identifier les gènes déterminant la caste

Les fourmis moissonneuses constituent un modèle idéal pour découvrir des gènes déterminant les castes. Si les objectifs 1 et 2 pourraient aider à identifier les gènes candidats via des approches comparatives (par exemple, des positions hétérozygotes dans toutes les ouvrières hybrides de plusieurs espèces ou des introgressions/balayages sélectifs parmi les lignées hybridogénétiques), l'Objectif 3 est plus précisément dédié à ce but. *M. barbarus* est l'une des espèces de fourmis les plus faciles à maintenir et a une fécondité extrêmement élevée dans des conditions artificielles. En se concentrant principalement sur cette espèce,

l'Objectif 3 utilisera des approches plus directes et expérimentales pour identifier ou valider les locus de caste.

2.5.3.1 Objectif 3a : Expression différentielle embryonnaire

Les castes et les génotypes étant intrinsèquement liés dans l'hybridogenèse sociale, celle-ci offre l'opportunité unique de prédire le développement de la reine ou de l'ouvrière à partir d'un œuf indifférencié. Cela ouvre la porte à des analyses d'expression différentielle entre les futures reines et ouvrières à un stade embryonnaire très précoce (c'est-à-dire avant tout indice morphologique ou important biais d'expression génique). Les données des transcriptomes d'œufs visuellement indifférenciés seront générées à partir de trois colonies matures de *M. barbarus* élevées dans des conditions contrôlées. Le transcriptome d'œufs d'âge connu sera séquencé. Les œufs seront ensuite affectés à un groupe "reine" ou "ouvrière" à partir de leur génotype et une analyse classique d'expression différentielle des gènes (5 futures ouvrières vs 5 futures reines) sera réalisée (voir Walsh *et al.* 2018 pour exemple). En l'absence de gènes différentiellement exprimés entre les deux castes, les œufs plus âgés seront échantillonnés et analysés. Au contraire, la présence de nombreux gènes exprimés de manière différentielle entre les futures reines et les futures ouvrières signifiera que le développement est déjà orienté vers la caste et que des œufs plus jeunes doivent être séquencés. En affinant progressivement l'âge des individus séquencés, le premier objectif sera d'identifier à quel jour commence le développement spécifique à la caste (Fig. 27). Enfin, l'âge sera affiné au niveau de l'heure afin d'identifier les tout premiers gènes architectes déclenchant une cascade d'événements conduisant à la caste.

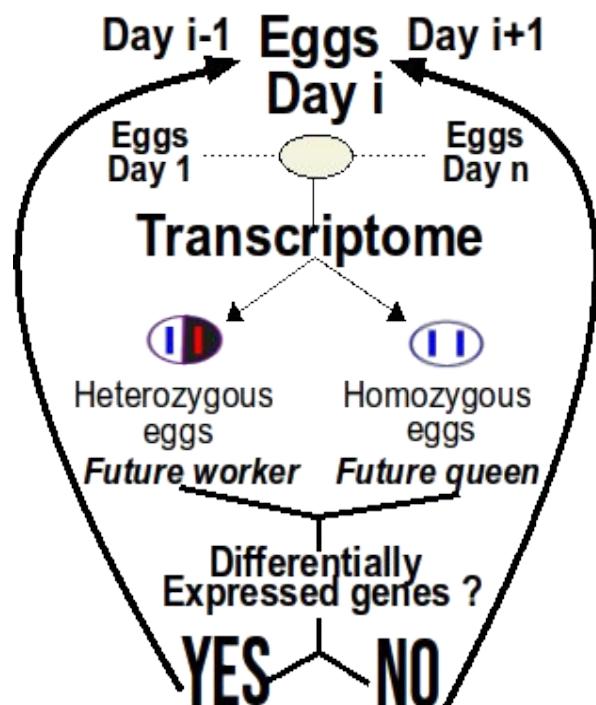


Figure 27 : Résumé graphique de l'Objectif 3a

Étape 1 Suivi de l'âge et échantillonnage des œufs

Une première difficulté majeure de cette expérience est de suivre précisément l'âge d'un œuf donné, ce qui est un défi dans une grande colonie de plusieurs milliers d'individus et une reine qui peut pondre des dizaines d'œufs par jour. Un problème majeur est qu'il est difficile de marquer visuellement les œufs d'une manière non invasive qui ne peut pas potentiellement biaiser leur expression génétique, en particulier compte tenu du fait qu'ils doivent être constamment nettoyés par les ouvrières pour éviter l'avortement après 1-2 jours. Lorsque les œufs et un groupe d'ouvrières sont isolés hors du nid, les œufs sont mangés et les ouvrières pondent leurs propres œufs non fécondés après quelques jours, probablement parce que l'absence de phéromones de la reine libère ce comportement égoïste. Une solution efficace est d'isoler physiquement les œufs et les ouvrières mais de les laisser sentir le reste de la colonie en les mettant dans des boîtes semi-isolées. Pour cette expérience, j'utiliserais un grand nid avec 20 boîtes semi-isolées fermées par un grillage fin et rempli de 20 ouvrières nourrices et de nourriture. La reine sera ensuite déplacée chaque jour d'une boîte à l'autre pendant 20 jours. A la fin du 15ème jour, environ ~100 œufs âgés de 1 à 20 jours seront ainsi disponibles pour l'échantillonnage. Nous avons mené avec succès une version pilote de ce protocole et collectés des œufs âgés de 1 à 10 jours. Puisqu'une reine pond en moyenne 5 œufs par heure, il devrait être possible à terme de réaliser des analyses d'expression différentielles par heures à partir de 3 œufs futures ouvrières vs 3 futures reines. Le nombre d'œufs pouvant fluctuer d'heure en heure pourrait obliger à concevoir un échantillonnage plus raisonnable toutes les 6 heures, ce qui reste extrêmement précis pour une espèce dont le développement larvaire peut atteindre 40 jours. Le temps de développement pouvant varier fortement en fonction de la température, la température des nids artificiels sera étroitement contrôlée à 28° degrés.

Étape 2 Séquençage du transcriptome

L'extraction d'ARN d'un œuf est un défi mais a été réalisée avec succès suite à l'étude pilote. Le séquençage du transcriptome de 10 œufs pondus le même jour a révélé que i) la qualité du transcriptome obtenu est excellente et ii) les œufs destinés aux reines et aux ouvrières semblent être pondus dans des proportions approximativement égales (6 génotypes de reines, 4 génotypes d'ouvrières).

Étape 3 Comparaison avec une espèce dont la caste est déterminée par l'environnement

Les gènes de caste différemment exprimés au stade embryonnaire précoce seront comparés à des données similaires chez une espèce témoin proche ayant un système de reproduction classique mais une taille et une biologie similaires (*M. capitatus*). Cela nous permettra de voir dans quelle mesure la cascade génétique qui suit le développement différentiel des castes est conservée et si *M. barbarus* présente des profils d'expression génique spécifiques ou non spécifiques pendant le premier jour/heure qui suit le début du développement spécifique de la caste. Comme nous connaîtrons déjà les signatures d'expression génique spécifiques du développement de la reine ou de l'ouvrière chez *M. barbarus*, nous pouvons espérer distinguer les futures reines des futures ouvrières à partir des données d'expression génique d'un œuf de *M. capitatus* suffisamment vieux. Ces données seront utilisées pour tester si un patrilignage est surreprésenté chez les futures

reines, ce qui pourrait être une indication potentielle d'un allèle biaisant le développement de la caste et ségrégant à basse fréquence^{48,67} (Boomsma et Hughes 2008, Withrow et Tarpy 2018). Un risque majeur de cette partie est que, comme les œufs sont censés être totipotents chez *M. capitatus*, ils pourraient ne que très rarement commencer à se développer en reines sans indices environnementaux appropriés. Les colonies seront exposées au froid (12-14°C) pendant les 2 mois précédant les expériences, car il semble déclencher le développement des reines chez de nombreuses espèces (Petersen-Braun 1977 , Suzzoni *et al.* 1980, Vargo et Passera 1992, Schwander *et al.* 2008), et les nids de *M. barbarus/M. capitatus* contiennent généralement des nymphes de reines au printemps, après l'hibernation. A des fins de contrôle, la même procédure sera appliquée pour l'échantillonnage des œufs de *M. barbarus* décrit ci-dessus.

2.5.3.2. Objectif 3b : accouplement des mâles recombinés inter-lignées

En condition de laboratoire, les colonies orphelines de *M. barbarus* produisent des mâles recombinés entre les deux lignées hybridogénétiques (Fig 28 et Romiguier *et al.* 2016). Ces mâles sont obtenus en séparant certaines ouvrières de leur reine mère, les ouvrières orphelines pondant généralement des mâles après quelques jours d'isolement. Dans les populations naturelles, cela pourrait ouvrir la porte à un flux génétique occasionnel entre les lignées hybridogénétiques, ce qui pourrait à son tour être un moyen de prévenir une divergence excessive et l'inviabilité des ouvrières hybrides. Les reines de *M. barbarus* ont probablement une durée de vie de plus de 20 ans (source : des colonies élevées dans mon laboratoire postdoc à l'Université de Lausanne sont en vie depuis plus de 20 ans) et les ouvrières vivent ~2 ans, ce qui signifie que théoriquement, une colonie devient orpheline après 20 ans et peut libérer des mâles recombinants interlignées pendant 1-2 ans. Les niveaux de flux génétique seront estimés dans l'Objectif 2 et comparés à des estimations expérimentales.

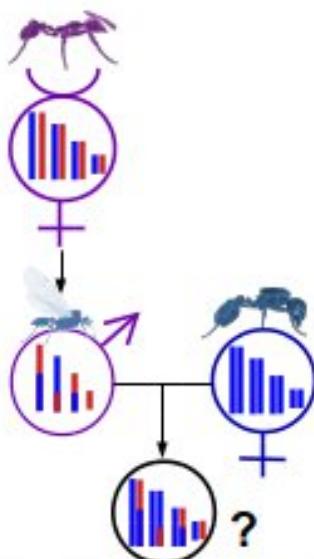


Figure 28 : Résumé graphique de l'Objectif 3b

Pour cela, nous essaierons d'accoupler des reines sauvages capturées lors d'un essaimage et de les accoupler avec un mâle recombinant produit en laboratoire (voir Fig 28). Le principal défi est que les essaimages ont lieu une fois par an pendant un laps de temps très limité de quelques jours, typiquement en octobre après une pluie. Une grande population naturelle de *M. barbarus* vit sur le campus de l'Université de Montpellier, permettant une détection relativement facile des émergences et l'échantillonnage de plusieurs centaines de nouvelles reines. Cependant, les accouplements peuvent être difficiles à observer car se déroulant à priori en plein vol. Pour surmonter ces limitations, nous développerons des stratégies d'insémination artificielle chez *M. barbarus*. L'insémination du sperme des mâles aux reines est extrêmement bien développé et optimisé chez les abeilles (Ruttner 1969), principalement en raison de son utilisation commerciale. Des inséminations artificielles basées sur le kit commercial utilisé pour les abeilles domestiques ont été utilisées avec succès sur les bourdons, mais aussi sur deux espèces de fourmis, *Solenopsis invicta* (Ball *et al.* 1983) et *Atta colombica* (den Boer *et al.* 2012).

Un objectif de 100 reines, fécondées par des mâles recombinants inter-lignées, sera élevé dans des nids artificiels. La ponte d'œufs dans certains d'entre eux indiquera que les mâles recombinants peuvent se reproduire avec succès avec des reines de lignée pure. Pour vérifier leur succès reproductif relatif, 100 autres reines seront inséminées avec le sperme d'un mâle normal de la lignée pure. Comme il est peu probable que ces œufs se développent en ouvrières et en reines dans une jeune colonie, les œufs seront adoptés par une colonie mature, en suivant le protocole de l'Objectif 3a avec de petites boîtes semi-isolées et quelques ouvrières pour s'en occuper. L'adoption des œufs est facile chez *M. barbarus* : lorsqu'ils sont recouverts d'eau sucrée, les œufs étrangers sont léchés et adoptés par toute colonie réceptrice. Les ouvrières et les reines qui vont éclore seront séquencées pour identifier quel chromosome ou partie de chromosome porte les gènes contrôlant la royauté. En principe, un chromosome ou une partie chromosomique systématiquement hybride chez toutes les ouvrières et/ou systématiquement de lignée pure chez toutes les reines sera considéré comme une région génomique candidate portant un déterminant génétique de la caste. Ces régions candidates potentielles seront comparées aux locus candidats potentiellement identifiés dans l'Objectif 1, l'Objectif 2 et l'Objectif 3a.

Un risque majeur de l'expérience est que très peu d'œufs atteignent le stade de reine, en raison d'une sous-optimalité naturelle des mâles recombinants inter-lignées, de soins sous-optimaux des ouvrières de la colonie réceptrice ou d'une sous-optimalité expérimentale de l'insémination artificielle. Si tel est le cas, davantage de reines peuvent être fécondées ; le nombre de reines vierges n'est pas limitatif car les populations naturelles de *M. barbarus* sont extrêmement répandues autour de mon laboratoire. En outre, le transcriptome de la moitié des œufs au stade de développement précoce sera séquencé. Comme nous connaîtrons les profils d'expression génétique spécifiques à chaque caste au cours du développement larvaire de *M. barbarus* (résultats de l'Objectif 3a), il sera possible de savoir si un œuf a commencé à se développer en reine ou en ouvrière sans attendre l'éclosion ou sans espérer que les ouvrières s'occupent suffisamment bien d'eux pour le développement complet de la reine en adulte.

En effectuant des accouplements contrôlés entre des mâles inter-lignées produits en laboratoire et des reines sauvages de lignée pure, cette partie 1) testera expérimentalement si un flux de gènes entre lignées royales peut se produire et 2) identifiera potentiellement une région chromosomique identifiée comme portant un déterminant génétique de la royauté.

2.5.3.3. Objectif 3c : Edition de gènes par CRISPR-Cas9

La technologie CRISPR-Cas9 a révolutionné le domaine de la génétique fonctionnelle et a ouvert la porte à des stratégies d'édition du génome relativement bon marché et faciles sur des espèces non-modèles (Doudna et Charpentier 2014, Taning et al. 2017, Sanders 2014, Mazo-Vargas et al. 2017, Martin et al. 2017). Ce domaine de recherche est particulièrement prometteur chez les insectes eusociaux pour mieux comprendre leur extraordinaire plasticité phénotypique. Récemment, deux études indépendantes ont montrées chez deux espèces différentes que des micro-injections de CRISPR-Cas9 dans l'œuf peuvent modifier avec succès le génome d'un embryon de fourmi (Trible et al. 2017, Yan et al. 2017). En raison de la détermination particulière de leur caste génétique et de leur haute fécondité en captivité, *M. barbarus* est l'espèce modèle parfaite pour valider expérimentalement un gène architecte potentiel de la royauté.

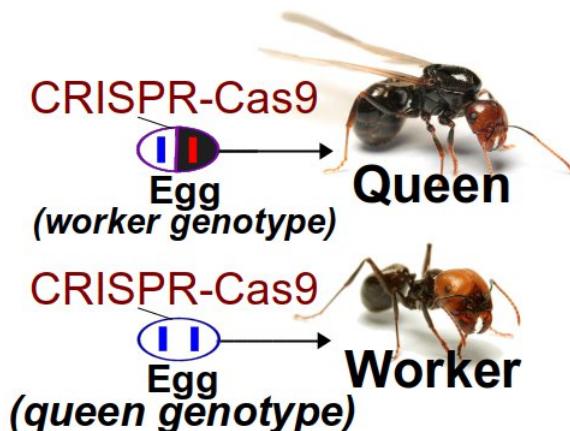


Figure 29 : Résumé graphique de l'Objectif 3c

Étape 1

Tout d'abord, le protocole de micro-injection dans l'œuf sera établi en essayant d'éliminer le même gène qui a été utilisé dans des études précédentes sur les fourmis (*orco*), en suivant exactement le même protocole avant d'ajuster certains paramètres (volume de l'injection, pression, concentrations...) pour optimiser le processus pour *M. barbarus*. Le risque est que, contrairement aux deux autres espèces de fourmis, un knock-out de ce gène n'entraîne aucun changement phénotypique visible ou soit létal chez *M. barbarus*. Si tel est le cas, un gène de pigmentation classique (*yellow*, Miyazaki et al. 2014) sera utilisé. Un autre risque est que les ouvrières détruisent les œufs micro-injectés lorsqu'ils sont replacés dans le nid. Pour surmonter ce problème potentiel, nous placerons les œufs dans un incubateur ou sur des plaques de gélose jusqu'à leur éclosion, comme cela a été fait avec succès dans des

études précédentes sur des fourmis similaires (Trible *et al.* 2017, Yan *et al.* 2017). Après l'éclosion des œufs micro-injectés avec succès, les larves seront semi-isolées dans des boîtes avec des ouvrières, en suivant le protocole de l'Objectif 3a.

Étape 2

Dans un deuxième temps, je me concentrerai sur un gène candidat pour la royauté (*catalase*) que j'ai déjà identifié (Romiguier *et al.* 2016). Ce gène est le seul portant une position hétérozygote non synonyme partagée par tous les ouvriers hybridogénétiques de toutes les espèces hybridogénétiques connues de *Messor* (*M. barbarus*, *M. structor* et *M. ebeninus*). La catalase est une enzyme clé qui protège les cellules des dommages oxydatifs causés par les espèces réactives de l'oxygène produites par l'activité mitochondriale (Michiels *et al.* 1994, Lee *et al.* 2010, Schriner *et al.* 2014). Il est intéressant de noter que les premiers stades des larves de reine d'abeille surexpriment la catalase (Cameron *et al.* 2013) ainsi que les composants protéiques de complexes cyto-nucléaires, probablement parce qu'elles nécessitent des taux respiratoires élevés (Eder *et al.* 1983, Begna *et al.* 2011). Il a même été suggéré que l'hybridogenèse sociale résultait d'une épistasie cyto-nucléaire, avec des gènes mitochondriaux et nucléaires co-évoluant en interaction pour permettre le développement de la larve en reine (Linksvayer *et al.* 2006). La catalase est donc un candidat sérieux pour contrôler le développement des castes chez *M. barbarus* et pourrait être un élément clé du modèle génétique qui sous-tend l'hybridogenèse sociale. L'édition de génome via Crispr-Cas9 est une occasion unique de vérifier expérimentalement si son expression ou ses mutations affectent le devenir d'un œuf destiné à devenir reine. En pratique, cela sera réalisé par des méthodes capables de manipuler l'expression d'un gène (Chavez *et al.* 2016, Huynh *et al.* 2018) ou de modifier son statut hétérozygote/homozygote (Paquet *et al.* 2016) au lieu de simples knockouts pour voir si les œufs destinés à devenir reine se développent en ouvrières ou si les œufs destinés à devenir ouvrières se développent en reines (voir Fig. 29). Les œufs micro-injectés avec succès seront éclos dans un incubateur ou sur des plaques d'agar puis placés dans des boîtes semi-isolées avec des nourrices jusqu'à leur développement complet. Un risque majeur est qu'ils ne se développent jamais en adultes à cause de conséquences inattendues de la micro-injection ou de la réaction des ouvrières nourrices. Si tel est le cas, le transcriptome des œufs sera séquencé afin de vérifier si leurs patrons d'expression génique spécifiques à la caste sont modifiés par rapport aux œufs non micro-injectés issus des résultats de l'Objectif 3a. Cela permettra de détecter un effet de nos modifications même si nous perturbons le développement complet vers l'adulte ou si les ouvrières ont tendance à négliger ces larves modifiées.

Étape 3

Enfin, ces développements méthodologiques Crispr-Cas9 permettront de tester et de valider davantage de gènes candidats potentiellement identifiés dans l'Objectif 1, l'Objectif 2 ou l'Objectif 3a/b comme étant responsables de la détermination génétique des castes dans l'hybridogenèse sociale et au-delà.

C'est la partie la plus novatrice et risquée du projet, mais les gains attendus sont potentiellement plus important et les autres objectif du projet RoyalMess ne dépendent pas du succès de cette expérience. La fécondité élevée de *Messor barbarus* (120 œufs par jour par reine) fourniront la quantité d'œufs nécessaires.

Chapitre 3

Curriculum Vitae

Jonathan Romiguier

Chargé de Recherche CNRS

Université de Montpellier

Né le 20/10/1986

Nationalité Française

1 enfant

CR CNRS (depuis février 2018)

Email : jonathan.romiguier@umontpellier.fr

ORCID : 0000-0002-2527-4740

Section CNU 67 : Biologie des populations et écologie

Institut des Sciences de l'Evolution de Montpellier (UMR 5554)

Expérience professionnelle

Post-doctorat Génomique évolutive des fourmis **Université de Lausanne (2014 - 2018)**

Supervision: Laurent Keller

Laboratoire : Department of Ecology and Evolution

Comprends 3 ans de salaire financé par une FEBS Long-term fellowship accordée à J. Romiguier

Post-doctorat Diversité génomique des animaux **Université de Montpellier (2012 - 2013)**

Supervision: Nicolas Galtier

Laboratoire : Institut des Sciences de l'Evolution (ISEM)

Financement ERC Advanced Grant (Projet PopPhyl) accordée à Nicolas Galtier

Cursus Universitaire

Thèse de Doctorat

Université de Montpellier (2012)

Supervision: Nicolas Galtier, Vincent Ranwez

Laboratoire : Institut des Sciences de l'Evolution (ISEM)

Titre : Phylogénomique et stratégies d'histoire de vie des mammifères placentaires : apports de la théorie de la conversion génique biaisée.

Master 2 - Ecologie, Biodiversité, Evolution

Université de Montpellier (2009)

Supervision : Nicolas Galtier, Vincent Ranwez, Emmanuel Douzery

Mention Très Bien (1^{er})

Laboratoire : Institut des Sciences de l'Evolution (ISEM)

Stage : Comment la conversion génique biaisée façonne-t-elle les génomes mammaliens ?

Master 1 - Interactions Symbiotiques et Parasitaires

Université de Montpellier (2008)

Supervision : Johan Michaux

Mention Très Bien (1^{er})

Laboratoire : Centre de Biologie pour la Gestion des Populations (CBGP)

Stage : Phylogéographie du mulot sylvestre et son parasite *Heligmosomoides polygyrus*.

Licence Biologie des Organismes

Université de Montpellier (2007)

Contrats de recherche en tant que porteur ou co-porteur

- **European Research Council Starting grant (ERC)**, 1 500 000€ (Porteur), 2021 - présent.

Génomique évolutive de la royaute chez les fourmis hybridogénétiques du genre *Messor* (Acronym: RoyalMess)

- **Bourse ANR t-ERC**, 120 000 € (Porteur), 2020 - 2021.

Génomique évolutive de la royaute chez les fourmis hybridogénétiques du genre *Messor*

- **Bourse ANR**, 605 640,24€ (co-porteur avec Patrice David et Emilian Lucquet, 92 340€ pour l'ISEM), 2018 - présent.

Conflits nucléocytoplasmiques et Gynodioécie chez les animaux (acronyme MINIGAN, porteur Patrice David)

- **Bourse CNRS (PICS)**, 10 000 € (Porteur), 2019.

- **Bourse Long-term fellowship de la Federation of European Biochemical Societies (FEBS)**, 180 000 € (Porteur).

Projets en collaboration

Bourse ANR PSEUDOGAMY (2018 – présent), Origine et conséquences de l'asexualité chez les nématodes *Mesorhabditis* (Porteuse : Marie Delattre)

Bourse ANR NeGa (2020 – présent), Influence la taille efficace de population sur l'architecture des génomes (Porteur : Tristan Lefébure)

Supervision

- POST-DOC

- Célia Lutrat

2022 - présent

Financement ERC RoyalMess : Déterminants génomiques de la caste chez les fourmis

- Arthur Weyna

2021 - présent

Financement ERC RoyalMess : Génomique des populations des fourmis moissonneuses du genre *Messor*

- DOC

- Yannick Juvé

2022 - présent

Financement ERC RoyalMess, co-dirigée avec Benoît Nabholz : Prévalence de l'hybridogénèse sociale chez les fourmis

- Fanny Laugier

2020 - présent

Financement ANR MINIGAN, co-dirigée avec Patrice David : Gynodioécie chez *Physa acuta*

- Arthur Weyna

2018 - 2021

Financement Ecole doctorale GAIA, co-dirigée avec Nicolas Galtier : Evolution et prévalence de l'hybridogénèse sociale chez les fourmis (Formicidae).

Actuellement en post-doctorat sous ma direction

– MASTER

- Yannick Juvé 2021**
Master à l’Université de Montpellier : Hybridogénèse sociale et caractérisation du microbiote chez les fourmis du genre *Messor*
Actuellement en doctorat sous ma direction
- François Monnet 2020**
Master à l’Université de Montpellier, en co-direction avec Camille Roux : Histoire démographique et flux de gènes au sein du genre *Messor*
Actuellement en doctorat à l’Université de Lille
- Lucille Bourouina 2019**
Master à l’Université de Montpellier : Patrons génomiques associés aux systèmes de reproduction atypiques chez les fourmis.
- Jean-Loup Imbert-Claret 2019**
Master à l’Université de Montpellier : L’activité des éléments transposables des génomes de Formicidae est-elle liée à leur niveau d’eusocialité ?
Actuellement en doctorat à l’Université de Montpellier
- Arthur Weyna 2018**
Master à l’Université de Montpellier : L’eusocialité module-t-elle l’efficacité de la sélection naturelle chez les Hyménoptères ?
Actuellement en post-doctorat sous ma direction
- Tomas Key 2017**
Master à l’Université de Lausanne en co-direction avec Laurent Keller : Expression génétique et déterminisme de caste des oeufs et larves de la fourmi moissonneuse *Messor barbarus*.
Actuellement en doctorat à l’Université de Lausanne
- Axel Fournier 2015**
Master à l’Université de Lille, en co-direction avec Laurent Keller : Hybridogénèse sociale chez les fourmis moissonneuses du genre *Messor*.
- Sébastien Ravel 2013**
Master à l’Université de Clermont-Ferrand, en co-direction avec Nicolas Galtier : Plutôt mammifères que mouches ? L’organisation sociale affecte l’efficacité de la sélection naturelle sur les protéomes de fourmis.
Actuellement Ingénieur au CIRAD (Montpellier)
- Emeric Figuet 2012**
Master à l’Université de Montpellier, en co-direction avec Nicolas Galtier : Origine des isochores chez les endothermes : ce qu’en disent les génomes des reptiles.

– LICENCE

- Alice Ha 2022**
Licence à l’Université de Lille : Hybridogénèse sociale chez les fourmis moissonneuses
Actuellement en Master
- Jean-Baptiste Sarraute 2020**

Licence à l'Université de Montpellier : Essaimages et accouplements chez *Messor barbarus*

Actuellement en doctorat à l'Université de Lille

Enseignement

| | |
|---|----------------------------------|
| 2019 - 2020 | Université de Montpellier |
| Initiation à la bioinformatique avec le langage de programmation Python (CM/TP/TD Licence 36h) | |
| 2014 – 2018 | Université de Lausanne |
| Génétique des populations (TD License 45h), Evolutionary Biology (CM Master 3h) | |
| 2009 – 2013 | Université de Montpellier |
| Génétique des populations (TD Licence 81h), Evolution moléculaire phylogénie avancée (TD Master 12h), Biologie animale et végétale (TP Licence 48h), Initiation à la ligne de commande UNIX sur environnement Linux (TP/TD Licence 60h) | |

Organisation d'évènement scientifique

ESEB meeting 2015, Lausanne, Suisse, 1400 participants (Organisation d'une salle pendant un symposium)
Biology 16 (2016) meeting, Lausanne, Suisse (Organisation d'une salle pendant un symposium)

Responsabilité institutionnelle

Délégué des post-doctorats au conseil du département écologie évolution (DEE), Université de Lausanne (élu, 2015 - 2018).

Expertise

Reviewer pour Nature, Biology Letters, Molecular Biology and Evolution, Genome Biology and Evolution Proceedings of the Royal Society B, Plos One, Peer J, BMC Evolutionary Biology, Molecular Ecology, Science Advances, PCI Evolutionary Biology

Membre du Conseil Scientifique du Museum d'Histoire Naturelle d'Orléans

Vulgarisation

Présentation d'une fourmilière au public lors de l'inauguration du Museum d'Histoire Naturelle d'Orléans (2021)

Organisation et participation des journées portes ouvertes de l'Université de Lausanne (2014 - 2017)

Animation dans le "Musée de la Main" (Lausanne, 2015)

Interview presse de vulgarisation scientifique : [NewsScientist](#), [Science News](#)

Communication orales et posters

August 2018 Ant phylogenomics based on the genome sequencing of 65 species. Joint ESEB-Evolution congress, Montpellier (poster)

August 2018 Comparative genomics of 65 ants species covering all subfamilies. IUSSI 2018 Garujá (Invited speaker)

August 2015 Parental investment predicts genetic diversity of animal species. ESEB (European Society for Evolutionary Biology) international congress, Lausanne (Speaker)

October 2014 What determines genetic diversity of animal species?. ARCAD days, Montpellier, (Invited Speaker)

February 2014 Can we predict genetic diversity from species biology ? An answer from 414 animal transcriptomes. ALPHY/PhyloSIB conference, Geneva (Speaker)

February 2013 Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. ALPHY (Alignments and Phylogeny) conference, Lyon (Speaker)

April 2012 Genomic evidence for large, long-lived mammalian ancestors. Jacques Monod Conference, Roscoff (Speaker)

February 2011 Probabilistic substitution mapping to explore the heterogeneity of sequence evolutionary process across a phylogenetic tree. ALPHY (Alignments and Phylogeny) conference, Lyon (Speaker)

July 2010 Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. SMBE (Society for Molecular Biology and Evolution) international congress, Lyon (Poster)

February 2010 Contrasting GC-content dynamics across 33 mammalian genomes: influence of body mass and chromosome size. ALPHY (Alignments and Phylogeny) conference, Marseille (Speaker)

Publications

32 articles (11 premiers auteurs, 4 derniers auteurs). Mon nom est souligné, le nom des personnes que j'ai supervisé en gras, des astérisques signalent mes publications en tant que dernier ou co-dernier auteur. Détails des citations par articles à jour disponible [ici](#).

Romiguier J, Borowiec ML, **Weyna A**, Helleu Q, Loire E, La Mendola C, Rabeling C, Fisher BL, Ward PS, Keller L. 2022. Ant phylogenomics reveals a natural selection hotspot preceding the origin of complex eusociality. *Current Biology*.

Weyna A, **Bourouina L**, Galtier N*, **Romiguier J***. 2022. Detection of F1 hybrids from single-genome data reveals frequent hybridization in Hymenoptera and particularly ants. *Molecular Biology and Evolution* 39:msac071.

Weyna A, **Romiguier J***. 2021. Relaxation of purifying selection suggests low effective population size in eusocial Hymenoptera and solitary pollinating bees. *Peer Community Journal* 1.

Weyna A, **Romiguier J***, Mullon C*. 2021. Hybridization enables the fixation of selfish queen genotypes in eusocial colonies. *Evolution letters* 5:582–594.

Fraïsse C, Popovic I, Mazoyer C, Spataro B, Delmotte S, **Romiguier J**, Loire E, Simon A, Galtier N, Duret L, et al. 2021. DILS: Demographic inferences with linked selection by using ABC. *Molecular Ecology Resources* 21:2629–2644.

Rolland J, Schluter D, **Romiguier J***. 2020. Vulnerability to fishing and life history traits correlate with the load of deleterious mutations in teleosts. *Molecular biology and evolution* 37:2192–2196.

- Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. 2020. MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Molecular Ecology Resources* 20:892–905.
- Ben Chehida Y, Thumloup J, Schumacher C, Harkins T, Aguilar A, Borrell A, Ferreira M, Rojas-Bracho L, Robertson KM, Taylor BL, Vikingsson A, **Weyna A**, Romiguier J, Morin AM, Fontaine MC. 2020. Mitochondrial genomics reveals the evolutionary history of the porpoises (Phocoenidae) across the speciation continuum. *Scientific Reports* 10:1–18.
- Fraïsse C, Roux C, Gagnaire P-A, Romiguier J, Faivre N, Welch JJ, Bierne N. 2018. The divergence history of European blue mussel species reconstructed from Approximate Bayesian Computation: the effects of sequencing techniques and sampling strategies. *PeerJ* 6:e5198.
- Galtier N, Roux C, Rousselle M, Romiguier J, Figuet E, Glémin S, Bierne N, Duret L. 2018. Codon usage bias in animals: disentangling the effects of natural selection, effective population size, and GC-biased gene conversion. *Molecular biology and evolution* 35:1092–1103.
- Romiguier J, Rolland J, Morandin C, Keller L. 2018. Phylogenomics of paleartic Formica species suggests a single origin of temporary parasitism and gives insights to the evolutionary pathway toward slave-making behaviour. *BMC evolutionary biology* 18:1–8.
- Romiguier J, **Fournier A**, Yek SH, Keller L. 2017. Convergent evolution of social hybridogenesis in Messor harvester ants. *Molecular Ecology* 26:1108–1117.
- Romiguier J, Roux C. 2017. Analytical biases associated with GC-content in molecular evolution. *Frontiers in Genetics* 8:16.
- Lucas ER, Romiguier J, Keller L. 2017. Gene expression is more strongly influenced by age than caste in the ant *Lasius niger*. *Molecular ecology* 26:5058–5073.
- Leroy T, Roux C, Villate L, Bodénès C, Romiguier J, Paiva JA, Dossat C, Aury J-M, Plomion C, Kremer A. 2017. Extensive recent secondary contacts between four European white oak species. *New Phytologist* 214:865–878.
- Romiguier J, Cameron SA, Woodard SH, Fischman BJ, Keller L, Praz CJ. 2016. Phylogenomics controlling for base compositional bias reveals a single origin of eusociality in corbiculate bees. *Molecular biology and evolution* 33:670–678.
- Roux C, Fraisse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS biology* 14:e2000234.
- Rolland J, Loiseau O, Romiguier J, Salamin N. 2016. Molecular evolutionary rates are not correlated with temperature and latitude in Squamata: an exception to the metabolic theory of ecology? *BMC evolutionary biology* 16:1–6.
- Figuet E**, Ballenghien M, Romiguier J, Galtier N. 2015. Biased gene conversion and GC-content evolution in the coding sequences of reptiles and vertebrates. *Genome Biology and Evolution* 7:240–250.

Galtier N, Romiguier J. 2015. Parental investment predicts genetic diversity in animal species. *Medecine Sciences: M/S* 31:17–19.

Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515:261–263.

Romiguier J, Lourenco J, Gayral P, Faivre N, Weinert LA, **Ravel S**, Ballenghien M, Cahais V, Bernard A, Loire E, et al. 2014. Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *Journal of Evolutionary Biology* 27:593–603.

Weber CC, Boussau B, Romiguier J, Jarvis ED, Ellegren H. 2014. Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. *Genome biology* 15:1–16.

Weber CC, Nabholz B, Romiguier J, Ellegren H. 2014. K r/K c but not d N/d S correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome biology* 15:1–13.

Douzery EJ, Scornavacca C, Romiguier J, Belkhir K, Galtier N, Delsuc F, Ranwez V. 2014. OrthoMaM v8: a database of orthologous exons and coding sequences for comparative genomics in mammals. *Molecular biology and evolution* 31:1923–1928.

Figuet E, Romiguier J, Dutheil JY, Galtier N. 2014. Mitochondrial DNA as a tool for reconstructing past life-history traits in mammals. *Journal of evolutionary biology* 27:899–910.

Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJ. 2013. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. *Molecular biology and evolution* 30:2134–2144.

Romiguier J, Ranwez V, Douzery E, Galtier N. 2013. Genomic evidence for large, long-lived ancestors to placental mammals. *Molecular biology and evolution* 30:5–13.

Loire E, Chiari Y, Bernard A, Cahais V, Romiguier J, Nabholz B, Lourenço JM, Galtier N. 2013. Population genomics of the endangered giant Galápagos tortoise. *Genome biology* 14:1–11.

Romiguier J, **Figuet E**, Galtier N, Douzery EJ, Boussau B, Dutheil JY, Ranwez V. 2012. Fast and robust characterization of time-heterogeneous sequence evolutionary processes using substitution mapping. *PLoS One* 7:e33852.

Dutheil JY, Galtier N, Romiguier J, Douzery EJ, Ranwez V, Boussau B. 2012. Efficient selection of branch-specific models of sequence evolution. *Molecular biology and evolution* 29:1861–1874.

Romiguier J, Ranwez V, Douzery EJ, Galtier N. 2010. Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. *Genome research* 20:1001–1009.

Chapitre 4

Bibliographie

1. Almeida EAB, Porto DS. (2014). Investigating eusociality in bees while trusting the uncertainty. *Sociobiology*, 61.
2. Anderson KE, Gadau J, Mott BM *et al.* (2006). Distribution and evolution of genetic caste determination in *Pogonomyrmex*-seed-harvester ants. *Ecology*, 87, 2171–2184.
3. Anderson KE, Gadau J, Mott BM, Johnson RA, Altamirano A, Strehl Cet al. (2006). Distribution and evolution of genetic caste determination in *Pogonomyrmex*-seed-harvester ants. *Ecology* 87:2171–2184.
4. Anderson KE, Linksvayer TE, Smith CR. (2008). The causes and consequences of genetic caste determination in ants (Hymenoptera: Formicidae). *Myrmecol. News* 11, 119–132.
5. Ball DE, Mirenda JT, Sorensen AA, Vinson SB. (1983). Instrumental insemination of the fire ant, *Solenopsis invicta*. *Entomol. Exp. Appl.* 33, 195–202.
6. Begna D, Fang Y, Feng M, Li J. (2011). Mitochondrial proteins differential expression during honeybee (*Apis mellifera*) queen and worker Larvae caste determination. *J. Proteome Res.* 10, 4263–4280.
7. Boomsma JJ et al. (2017). The Global Ant Genomics Alliance (GAGA). *Myrmecol. News* 25, 61–66.
8. Boussau B, Gouy M. (2006). Efficient likelihood computations with nonreversible models of evolution. *Systematic biology*, 55, 756–768.
9. Borowiec ML, Cover SP, Rabeling C. (2021). The evolution of social parasitism in *Formica* ants revealed by a global phylogeny. *Proceedings of the National Academy of Sciences*, 118(38), e2026029118.
10. Borowiec ML, Rabeling C, Brady SG, Fisher BL, Schultz TR, Ward PS. (2019). Compositional heterogeneity and outgroup choice influence the internal phylogeny of the ants. *Molecular phylogenetics and evolution*, 134, 111-121.
11. Branstetter MG, Longino JT, Reyes-López JL, Brady SG, Schultz TR. (2022). Out of the temperate zone: A phylogenomic test of the biogeographical conservatism hypothesis in a contrarian clade of ants. *Journal of Biogeography*, 49(9), 1640-1653.
12. Bulot A, Dutoit T, Renucci M, Provost E. (2014). A new transplantation protocol for harvester ant queens *Messor barbarus* (Hymenoptera: Formicidae) to improve the restoration of species-rich plant communities. *Myrmecol. News* 20, 43–53.
13. Buschinger A, Alfred B. (1986). Evolution of social parasitism in ants. *Trends in ecology & evolution*, 1, 155–160.
14. Cameron RC, Duncan EJ, Dearden PK. (2013). Biased gene expression in early honeybee larval development. *BMC Genomics* 14, 903.
15. Castalanelli MA et al. (2010). A rapid non-destructive DNA extraction method for insects and other arthropods. *J. Asia. Pac. Entomol.* 13, 243–248.

16. Chavez A et al. (2016). Comparison of Cas9 activators in multiple species. *Nat. Methods* 13, 563–567.
17. Corona M et al. .(2007). Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proceedings of the National Academy of Sciences* 104, 7128–7133.
18. Darwin C.(1871). *On the origin of species*. (John Murray).
19. Danforth BN, Cardinal S, Praz C, Almeida EAB, Michez D. (2013). The impact of molecular data on our understanding of bee phylogeny and evolution. *Annual review of entomology*, 58, 57–78.
20. Davidson DW. (1982). Sexual selection in harvester ants (Hymenoptera: Formicidae: Pogonomyrmex). *Behav. Ecol. Sociobiol.* 10, 245–250.
21. Doudna JA, Charpentier E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346, 1258096.
22. Douzery EJP, Scornavacca C, Romiguier J *et al.*(2014). OrthoMaM v8: a database of orthologous exons and coding sequences for comparative genomics in mammals. *Molecular biology and evolution*, 31, 1923–1928.
23. De Almeida T, Blight O, Mesléard F, Bulot A, Provost E, Dutoit T. (2020). Harvester ants as ecological engineers for Mediterranean grassland restoration: Impacts on soil and vegetation. *Biological Conservation*,245, 108547.
24. den Boer SPA, Boomsma JJ, Baer B. (2012). A technique to artificially inseminate leafcutter ants. *Insectes Soc.* 60, 111–118.
25. Derome N, Métayer K, Montchamp-Moreau C, Veuille M. (2004). Signature of selective sweep associated with the evolution of sex-ratio drive in *Drosophila simulans*. *Genetics* 166, 1357–1366.
26. Didion JP et al. (2016). R2d2 Drives selfishsweeps in the housemouse. *Mol. Biol. Evol.* 33, 1381–1395.
27. Duret L, Semon M, Piganeau G, Mouchiroud D, Galtier N. (2002).Vanishing GC-rich isochores in mammalian genomes. *Genetics* 162: 1837–1847
28. Duret L, Galtier N. (2009). Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual review of genomics and human genetics*,10(1), 285-311.
29. EvansJD, Wheeler DE. (2001). Gene expression and the evolution of insect polyphenisms. *Bioessays* **23**,62–68.
30. Eder J, Kremer J.-P, Rembold H. (1983). Correlation of cytochrome titer and respiration in *Apis mellifera*: Adaptive response to caste determination defines workers, intercastes and queens. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* 76, 703–716.

31. Figuet E, Ballenghien M, Romiguier J, Galtier N. (2015). Biased gene conversion and GC-content evolution in the coding sequences of reptiles and vertebrates. *Genome biology and evolution*, 7, 240–250.
32. Figuet E, Romiguier J, Dutheil JY, Galtier N. (2014). Mitochondrial DNA as a tool for reconstructing past life-history traits in mammals. *Journal of evolutionary biology*, 27, 899–910.
33. Gomez C, Abril S. (2012). Nuptial flights of the seed-harvester ant *Messor barbarus*(Linnaeus, 1767) (Hymenoptera: Formicidae) in the Iberian Peninsula: synchrony, spatial scale and weather conditions. *Myrmecol. News* 16, 25–29.
34. Goropashnaya AV, Fedorov VB, Seifert B, Pamilo P. (2012). Phylogenetic relationships of Palaearctic *Formica* species (Hymenoptera, Formicidae) based on mitochondrial cytochrome B sequences. *PloS one*, 7, e41697.
35. Hamilton WD. (1964). The genetical evolution of social behaviour. II. *J. Theor. Biol.* 7, 17–52.
36. Hartfelder K, Tiberio GJ, Lago DC, Dallacqua RP, Bitondi MMG. (2018). The ovary and its genes—developmental processes underlying the establishment and function of a highly divergent reproductive system in the female castes of the honey bee, *Apis mellifera*. *Apidologie*, 49(1), 49-70.
37. Haydak MH. (1943). Larval food and development of castes in the honeybee. *J. Econ. Entomol.* 36, 778–792.
38. Helms Cahan S, JD Parker, SW Rissing, RA Johnson, TS Polony, MD Weiser *et al.* (2002). Extreme genetic differences between queens and workers in hybridizing *Pogonomyrmex* harvester ants. *Proc. R. Soc.* 269:1871–1877.
39. Helms Cahan S, Keller L. (2003). Complex hybrid origin of genetic caste determination in harvester ants. *Nature*, 424, 306–309.
40. Helms Cahan S and Vinson SB. (2003). Reproductive division of labor between hybrid and nonhybrid offspring in a fire ant hybrid zone. *Evolution* 57:1562–1570.
41. Helms Cahan S, Nguyen AD, Zhou Y. (2022). Population genomics supports multiple hybrid zone origins of socially hybridogenetic lineages of *Pogonomyrmex* harvester ants. *Evolution*, 76(5), 1016-1032.
42. Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T. (2011). Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome research*, 21, 349–356.
43. Hölldobler B, Wilson EO (1990) *The Ants*. Harvard University Press.
44. Hughes WOH, Boomsma JJ. (2008). Genetic royal cheats in leaf-cutting ant societies. *Proc. Natl. Acad. Sci. U. S. A.* 105, 5150–5153.

45. Janicki J, Narula N, Ziegler M, Guénard B, Economo EP. (2016). Visualizing and interacting with large-volume biodiversity data using client–server web-mapping applications: The design and implementation of antmaps.org. *Ecol. Inform.* 32, 185–193.
46. Kamakura M. (2011). Royalactin induces queen differentiation in honeybees. *Nature* 473, 478–483.
47. Kapheim KM, Pan H, Li C *et al.* (2015). Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. *Science*, 348, 1139–1143.
48. Kent CF, Minaei S, Harpur BA, Zayed A. (2012). Recombination is associated with the evolution of genome structure and worker behavior in honey bees. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 18012–18017.
49. Huynh N, Zeng J, Liu W, King-Jones K. (2018). A *Drosophila* CRISPR/Cas9 toolkit for conditionally manipulating gene expression in the prothoracic gland as a test case for polytene tissues. *G3: Genes, Genomes, Genetics*, 8(11), 3593–3605.
50. Kück P, Hita Garcia F, Misof B, Meusemann K. (2011). Improved phylogenetic analyses corroborate a plausible position of *Martialis heureka* in the ant tree of life. *PloS one*, 6(6), e21031.
51. Kuhn, A., H. Darras, O. Paknia, and S. Aron. (2020). Repeated evolution of queen parthenogenesis and social hybridogenesis in *Cataglyphis* desert ants. *Mol. Ecol.* 29:549–564.
52. Lacy KD, Shoemaker D, Ross KG. (2019). Joint evolution of asexuality and queen number in an ant. *Current Biology*, 29:1394–1400.
53. Lee HY *et al.* (2010). Targeted expression of catalase to mitochondria prevents age-associated reductions in mitochondrial function and insulin resistance. *Cell Metab.* 12, 668–674.
54. Leffler EM, Kevin B, Matute DR *et al.* (2012). Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS biology*, 10, e1001388.
55. Linksvayer TA, Wade MJ, Gordon DM (2006) Genetic caste determination in harvester ants: possible origin and maintenance by cyto-nuclear epistasis. *Ecology* 87, 2185–2193.
56. Loire E, Chiari Y, Bernard A, Cahais V, Romiguier J, Nabholz B, Lourenço JM, Galtier N. (2013). Population genomics of the endangered giant Galápagos tortoise. *Genome biology*, 14, R136.
57. Madeo F, Zimmermann A, Maiuri MC, Kroemer G. (2015). Essential role for autophagy in life span extension. *The Journal of clinical investigation*, 125(1), 85–93.
58. Martin A *et al.* (2016). CRISPR/Cas9 mutagenesis reveals versatile roles of Hox genes in crustacean limb specification and evolution. *Curr. Biol.* 26, 14–26.

59. Mazo-Vargas A *et al.* (2017). Macroevolutionary shifts of WntA function potentiate butterfly wing-pattern diversity. *Proceedings of the National Academy of Sciences* 114, 10701–10706.
60. Markl H, Hölldobler B, Hölldobler T. (1977). Mating behavior and sound production in harvester ants (*Pogonomyrmex*, Formicidae). *Insectes Soc.* 24, 191–212.
61. Martin SH, Davey JW, Jiggins CD. (2014). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Mol. Biol. Evol.* 32, 244–257.
62. Mattila HR, Seeley TD. (2007). Genetic diversity in honey bee colonies enhances productivity and fitness. *Science* 317, 362–364.
63. Michiels C, Raes M, Toussaint O, Remacle J. (1994). Importance of Se-glutathione peroxidase, catalase, and Cu/Zn-SOD for cell survival against oxidative stress. *Free Radic. Biol. Med.* 17, 235–248.
64. Miyazaki S *et al.* (2014). Sexually dimorphic body color is regulated by sex-specific expression of yellow gene in ponerine ant, *Diacammasp*. *PLoS One* 9, e92875.
65. Moreau CS, Bell CD. (2013). Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. *Evolution*, 67(8), 2240-2257.
66. Nikolaev SI, Montoya-Burgos JI, Popadin K *et al.* (2007). Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 20443–20448.
67. Ohta T. (1992). The nearly neutral theory of molecular evolution. *Annual review of ecology and systematics*, 263-286.
68. Oldroyd BP, Fewell JH. (2007). Genetic diversity promotes homeostasis in insect colonies. *Trends Ecol. Evol.* 22, 408–413.
69. Parker JD. (2004). A major evolutionary transition to more than two sexes? *Trends Ecol. Evol.* 19, 83–86 .
70. Peters RS, Krognann L, Mayer C, Donath A, Gunkel S, Meusemann K *et al.* (2017). Evolutionary history of the Hymenoptera. *Current Biology*, 27(7), 1013-1018.
71. Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K. (2007). Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 13390–13395.
72. Presgraves DC, Gérard PR, Cherukuri A, Lyttle TW. (2009). Large-scale selective sweep among Segregation Distorter chromosomes in African populations of *Drosophila melanogaster*. *PLoS Genet.* 5, e1000463.

73. Rabeling C, Brown JM, Verhaagh M. (2008). Newly discovered sister lineage sheds light on early ant evolution. *Proceedings of the National Academy of Sciences*, 105(39), 14913-14917.
74. Rey C, Lanore V, Veber P, Guéguel L, Lartillot N, Sémon M, Boussau B. (2019). Detecting adaptive convergent amino acid evolution. *Philosophical Transactions of the Royal Society B*, 374(1777), 20180234.
75. Rolland J, Schluter D, Romiguier, J. (2020). Vulnerability to fishing and life history traits correlate with the load of deleterious mutations in teleosts. *Molecular biology and evolution*, 37(8), 2192-2196.
76. Romiguier J, Cameron SA, Woodard SH *et al.* (2015). Phylogenomics controlling for basecompositionalbiasreveals a singleorigin of eusociality in corbiculatebees. *Molecular biology and evolution*.
77. Romiguier J, Figuet E, Galtier N *et al.* (2012). Fast and robust characterization of time-heterogeneous sequence evolutionary processes using substitution mapping. *PloS one*, 7, e33852.
78. Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, Loire E, Lourenco JM, Nabholz B, Roux C, Tsagkogeorga G, Weber AAT, Weinert LA, Belkhir K, Bierne N, Glémin S, Galtier N. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515(7526), 261-263.
79. Romiguier J, Lourenco J, Gayral P, Faivre N, Weinert LA, Ravel S, Ballenghien M, Cahais V, Bernard A, Loire E, Keller L, Galtier N. (2014b). Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *Journal of evolutionary biology*, 27, 593–603.
80. Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJP. (2013a). Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. *Molecular biology and evolution*, 30, 2134–2144.
81. Romiguier J, Ranwez V, Douzery EJP, Galtier N. (2010). Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. *Genome research*, 20, 1001–1009.
82. Romiguier J, Ranwez V, Douzery EJP, Galtier N. (2013b). Genomic evidence for large, long-lived ancestors to placental mammals. *Molecular biology and evolution*, 30, 5–13.
83. Romiguier J, Fournier A, Yek SH, Keller L. (2016). Convergent evolution of social hybridogenesis in *Messor*harvester ants. *Molecular ecology*, doi:10.1111/mec.13899.
84. Romiguier J, Roux C. (2017). Analytical biases associated with GC-content in molecular evolution. *Frontiers in Genetics*, 8, 16.
85. Romiguier J, Rolland J, Morandin C, Keller L. (2018). Phylogenomics of paleartic *Formicaspecies* suggests a single origin of temporary parasitism and gives insights to the evolutionary pathway toward slave-making behaviour. *BMC evolutionary biology*, 18(1), 1-8.

86. Rousselle M, Simion P, Tilak MK, Figuet E, Nabholz B, Galtier N. (2020). Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. *PLoS Genetics*, 16(4), e1008668.
87. Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *Plos Biology* 14, 12, doi:10.1101/059790
88. Ruttner F. (1969). Apimondia. Die Instrumentelle Besamung der Bienenkönigin.
89. Sanders L. (2014). Body & brain: Bizarre vision may save brainpower: Mantis shrimp perceive color in a way unlike any other animal. *Sci. News* 185, 7–7.
90. Santos D, Ribeiro GC, Cabral AD, Sperança MA. (2018). A non-destructive enzymatic method to extract DNA from arthropod specimens: Implications for morphological and molecular studies. *PLoS One* 13, e0192200.
91. Schriner SE *et al.* (2005). Extension of murine life span by overexpression of catalase targeted to mitochondria. *Science* 308, 1909–1911.
92. Schwander T, Keller L, Cahan SH. (2007). Two alternate mechanisms contribute to the persistence of interdependent lineages in *Pogonomyrmex* harvester ants. *Molecular ecology*, 16, 3533–3543.
93. Schwander T *et al.* (2008). Maternal effect on femalecastedetermination in a socialinsect. *Curr. Biol.* 18, 265–269.
94. Schwander T, Lo N, Beekman M, Oldroyd BP, Keller L. (2010). Nature versus nurture in social insect caste differentiation. *Trends in ecology & evolution*, 25, 275–282.
95. Schwander T, Keller L. (2012). Evolution: sociality as a driver of unorthodox reproduction. *Current biology*, 22, R525–7.
96. Simola DF, Graham RJ, Brady CM *et al.* (2016). Epigenetic (re) programming of caste-specific behavior in the ant *Camponotus floridanus*. *Science*, 351(6268), aac6633.
97. Simpson SJ, Sword GA, Lo N. (2012). Polyphenism in Insects. *Curr. Biol.* 22, 352.
98. Smith CR, Toth AL, Suarez AV, Robinson GE. (2008). Genetic and genomic analyses of the division of labour in insect societies. *Nature reviews. Genetics*, 9, 735–748.
99. Smith CR, Suarez AV. (2010). The trophic ecology of castes in harvester ant colonies. *Funct. Ecol.* 24, 122–130.
100. Smith CR *et al.* (2011) Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 5667–5672.
101. Steiner FM *et al.* (2011). Mixed colonies and hybridisation of *Messor* harvester ant species (Hymenoptera: Formicidae). *Org. Divers. Evol.* 11, 107–134.

102. Steiner FM *et al.* (2018). Turning one into five: Integrative taxonomy uncovers complex evolution of cryptic species in the harvester ant *Messor 'structor'*. *Mol. Phylogenet. Evol.* 127, 387–404.
103. Taning CNT, Van Eynde B, Yu N, Ma S, Smagghe G. (2017). CRISPR/Cas9 in insects: Applications best practices and biosafety concerns. *J. Insect Physiol.* 98, 245–257.
104. Tin MMY, Economo EP, Mikheyev AS. (2014). Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *PLoS One* 9, e96793.
105. Trible W *et al.* (2017). orco mutagenesis causes loss of antennal lobe glomeruli and impaired social behavior in ants. *Cell* 170, 727–735.e10.
106. Tschinkel WR. (2006). The Fire Ants. (Harvard University Press).
107. Umphrey JG. (2006). Sperm parasitism in ants: selection for interspecific mating and hybridization. *Ecology* 87:2148–2159.
108. Valone TJ, Kaspari M. (2005). Interactions between granivorous and omnivorous ants in a desert grassland: results from a long-term experiment. *Ecol. Entomol.* 30, 116–121.
109. Waddington CH. (1953). Genetic assimilation of an acquired character. *Evolution* 7, 118–126.
110. Wallberg A, Glémin S, Webster MT. (2015). Extreme recombination frequencies shape genome variation and evolution in the honeybee, *Apis mellifera*. *PLoS genetics*, 11, e1005189.
111. Walsh JT, Warner MR, Kase A, Cushing BJ, Linksvayer TA. (2018). Ant nurse workers exhibit behavioural and transcriptomic signatures of specialization on larval stage. *Anim. Behav.* 141, 161–169.
112. West-Eberhard M. (1989). Phenotypic plasticity and the origins of diversity. *Annu. Rev. Ecol. Syst.* 20, 249–278.
113. Weber CC, Benoit N, Jonathan R, Hans E. (2014a). K r /K c but not d N /d S correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome biology*, 15.
114. Weber CC, Boussau B, Romiguier J, Jarvis ED, Ellegren H. (2014b). Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. *Genome biology*, 15, 549.
115. Withrow JM, Tarpy DR. (2018). Cryptic 'royal' subfamilies in honey bee (*Apis mellifera*) colonies. *PLoS One* 13, e0199124.
116. Wilfert L, Gadau J, Schmid-Hempel P. (2007). Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity*, 98, 189–197.

117. Woodard SH, Fischman BJ, Venkat A et al. (2011). Genes involved in convergent evolution of eusociality in bees. *Proceedings of the National Academy of Sciences*, 108, 7472–7477.
118. Weyna A, Romiguier J. (2021). Relaxation of purifying selection suggests low effective population size in eusocial Hymenoptera and solitary pollinating bees. *Peer Community Journal*, 1.
119. Weyna A, Romiguier J, Mullon C. (2021). Hybridization enables the fixation of selfish queen genotypes in eusocial colonies. *Evolution letters* 5:582–594.
120. Weyna A, Bourouina L, Galtier N, Romiguier J. (2022). Detection of F1 hybrids from single-genome data reveals frequent hybridization in Hymenoptera and particularly ants. *Molecular Biology and Evolution*, 39(4), msac071.
121. Yan H et al. (2017). An engineered orco mutation produces aberrant social behavior and defective neural development in ants. *Cell* 170, 736–747.e9.
122. Yang Z, Ziheng Y, Rasmus N. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of molecular evolution*, 46, 409–418.

Annexe

Sélection d'articles

Comparative population genomics in animals uncovers the determinants of genetic diversity

J. Romiguier^{1,2}, P. Gayral^{1,3}, M. Ballenghien¹, A. Bernard¹, V. Cahais¹, A. Chenuil⁴, Y. Chiari⁵, R. Dernat¹, L. Duret⁶, N. Faivre¹, E. Loire¹, J. M. Lourenco¹, B. Nabholz¹, C. Roux^{1,2}, G. Tsagkogeorga^{1,7}, A. A.-T. Weber⁴, L. A. Weinert^{1,8}, K. Belkhir¹, N. Bierne¹, S. Glémén¹ & N. Galtier¹

Genetic diversity is the amount of variation observed between DNA sequences from distinct individuals of a given species. This pivotal concept of population genetics has implications for species health, domestication, management and conservation. Levels of genetic diversity seem to vary greatly in natural populations and species, but the determinants of this variation, and particularly the relative influences of species biology and ecology versus population history, are still largely mysterious^{1,2}. Here we show that the diversity of a species is predictable, and is determined in the first place by its ecological strategy. We investigated the genome-wide diversity of 76 non-model animal species by sequencing the transcriptome of two to ten individuals in each species. The distribution of genetic diversity between species revealed no detectable influence of geographic range or invasive status but was accurately predicted by key species traits related to parental investment: long-lived or low-fecundity species with brooding ability were genetically less diverse than short-lived or highly fecund ones. Our analysis demonstrates the influence of long-term life-history strategies on species response to short-term environmental perturbations, a result with immediate implications for conservation policies.

Since the early studies of evolutionary genetics, there has been no understanding of how and why genetic diversity levels vary between species. This old puzzle, considered four decades ago as ‘the central problem in population genetics’¹, is still essentially unsolved in the genomic era². Meanwhile, there is increasing evidence that genetic diversity is central to many conservation challenges, such as species response to environmental changes, ecosystem recovery, and the viability of recently endangered populations^{3–7}. In this context, our ability to understand and predict this key aspect of biodiversity seems critical. But is it possible to quantify the contributory ecological and genetic factors? How predictable is the level of genetic diversity of a given species?

Population genetic theory states that neutral genetic polymorphism (that is, diversity) increases with effective population size, N_e , which in a panmictic population is equal to the number of individuals contributing to reproduction. One would therefore expect the genetic diversity of a species to be linked to biological traits associated with abundance, such as body size or fecundity. However, this intuitive prediction has not yet been clearly confirmed by empirical data^{2,8–10}. This is typically explained by invoking the many confounding factors potentially affecting genetic polymorphism, such as mutation rate, population structure, population bottlenecks, selective sweeps, and, more generally, ecological disturbances^{11,12}. Whether demographic or adaptive, historical contingency is often considered to be the main driver of genetic diversity¹¹. According to this viewpoint, polymorphism levels would be expected to fluctuate in time more or less randomly, irrespective of life-history traits.

In the absence of compelling empirical evidence, the relative importance of species biology and ecology (on the one hand) and historical, contingent factors (on the other) in shaping the genetic diversity of species is

still highly contentious. Indeed, current knowledge on species genetic diversity is based on just a handful of model organisms, or small sets of molecular data^{2,8,13}. Various animal taxa and lifestyles, particularly across the invertebrates, have yet to be explored. Here we fill this gap and present the first distribution of genome-wide polymorphism levels across the metazoan tree of life.

We focused on 31 families of animals spread across eight major animal phyla. In each family we produced high-coverage transcriptomic data (RNAseq) for about ten individuals of a particular species. In 25 of these families, we sampled one to three additional species of similar biology and ecology (two to seven individuals each), thus producing taxonomic replicates. The total data set consisted of 374 individual transcriptomes from 76 non-model species (Fig. 1, Extended Data Fig. 1 and Supplementary Tables 1 and 2), from which we predicted protein coding sequences¹⁴ and identified diploid genotypes and single nucleotide polymorphisms^{15,16} (Methods). Across species the number of analysed loci varied from 804 to 20,222 (median 5,347) and the number of polymorphic sites from 1,759 to ~230,000 (median 17,924).

Estimates of the synonymous nucleotide diversity (π_s) spanned two orders of magnitude across species, a range far wider than is usually observed in surveys restricted to a single taxonomic group. The extreme values of π_s were observed in two invertebrate species: 0.1% in the subterranean termite *Reticulitermes grassei*; 8.3% in the slipper shell *Bostrycapulus aculeatus*. Figure 1 illustrates the patchy distribution of low-diversity (green) and high-diversity (red) species across the metazoan phylogeny. It also shows that species π_s values tend to be similar within families, but distinct between families (analysis of variance; $P < 10^{-12}$). Such a strong taxonomic effect would be unexpected if stochastic disturbances and contingent effects were the main drivers of genetic diversity, because species from a given family are not particularly expected to share a common demographic history. Testing this hypothesis more thoroughly, we detected no strong relationship between π_s and any variable related to geography, such as the average distance between GPS records (regression test, $P = 0.19$, $r^2 = 0.02$), maximum distance between GPS records ($P = 0.02$, $r^2 = 0.07$), average distance to Equator ($P = 0.87$, $r^2 = 0.0003$), population structure (measured as F_{ST} , $P = 0.22$, $r^2 = 0.02$), invasive status (Student’s *t*-test, $P = 0.14$) and marine versus continental environment (Student’s *t*-test, $P = 0.52$).

To test whether species biology can explain variations in π_s , we collected data for six life-history traits potentially related to N_e : adult size, body mass, maximum longevity, adult dispersion ability, fecundity and propagule size (Supplementary Table 3). In contrast to the geographic variables, all these traits were significantly correlated with the nucleotide diversity (Extended Data Fig. 2) and collectively explained 73% of the variance in π_s in a multiple linear regression test ($P < 10^{-10}$). Propagule size, here defined as the size of the stage that leaves its parents and disperses (egg or juvenile depending on species), is by far the most predictive

¹UMR 5554, Institute of Evolutionary Sciences, University Montpellier 2, Centre national de la recherche scientifique, Place E. Bataillon, 34095 Montpellier, France. ²Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland. ³UMR 7261, Institut de Recherches sur la Biologie de l’Insecte, Centre national de la recherche scientifique, Université François-Rabelais, 37200 Tours, France. ⁴Aix-Marseille Université, Institut Méditerranéen de Biodiversité et d’Écologie marine et continentale (IMBE) – CNRS – IRD – UAPV, 13007 Marseille, France. ⁵Department of Biology, University of South Alabama, Mobile, Alabama 36688-0002, USA. ⁶UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1, CNRS, 69622 Lyon, France. ⁷The School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK. ⁸Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK.

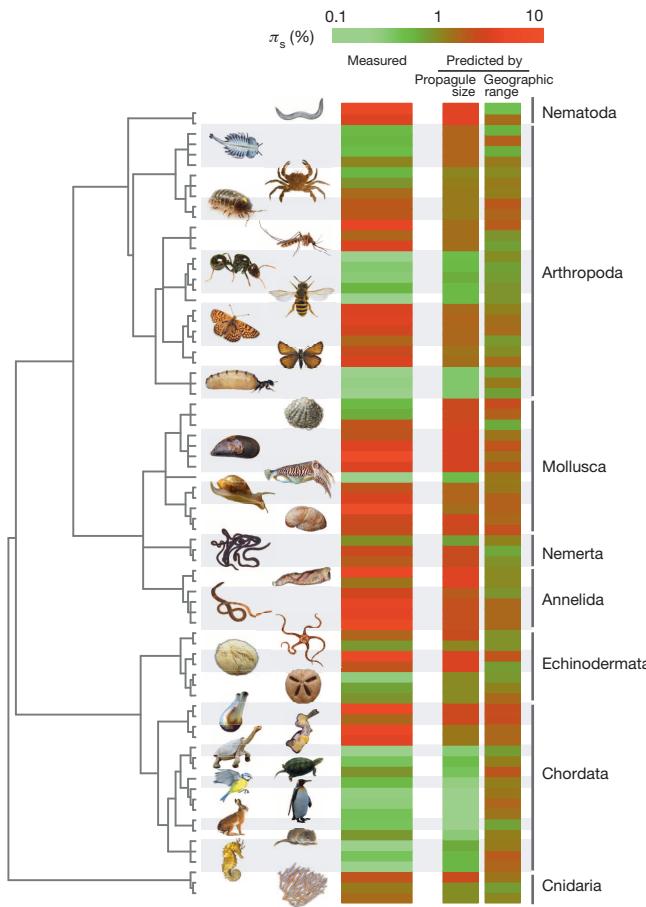


Figure 1 | Genome-wide genetic diversity across the metazoan tree of life. Each branch of the tree represents a species ($n = 76$). The leftmost vertical coloured bar is the estimated genome-wide genetic diversity (π_s), the central bar is the prediction of π_s based on a linear model with propagule size as the explanatory variable ($P < 10^{-14}$, $r^2 = 0.56$), and the rightmost bar is the prediction of π_s based on a linear model with average distance between GPS records, maximal distance between GPS records, average distance to Equator and invasive status as explanatory variables ($P = 0.16$). Each thumbnail corresponds to one metazoan family. Species are in the same order as in Supplementary Table 2 (from top to bottom).

of these variables (linear regression test, $r^2 = 0.56$; Fig. 2a). This is illustrated in Fig. 1 by the good agreement between the observed distribution of π_s (leftmost coloured vertical bar) and the π_s value predicted from propagule size (central bar). The predicted π_s based on four demographic metrics is plotted alongside (rightmost bar) for visual comparison.

We explored in more detail the relative impact on π_s of the various life-history traits of interest here (Extended Data Fig. 2). Figure 2b plots the relationship between π_s and species adult size, a variable typically taken as a proxy for population size in some taxa⁹. Although significant, the correlation is not particularly strong ($P = 0.018$, $r^2 = 0.07$). In particular, species with low genetic diversity cover a large range of body sizes, from less than 1 cm to more than 1 m. Low-polymorphism species include amniotes (turtles, mammals and birds), but also brooding marine species (seahorses, brooding urchins, nemerteans and brittle-stars), eusocial insects (ants, bees and termites) and cuttlefish. These phylogenetically unrelated species have in common a large parental investment in their offspring, as represented in Fig. 2b by the ratio of propagule size to adult size (red). In contrast, species with minimal parental investment (blue) tend to carry high genetic diversity given their size. This is typically the case of highly fecund, broadcast spawning sessile species (such as mussels, non-brooding urchins, nemerteans and brittle-stars, sea squirts and gorgonians). The trade-off between offspring quantity (fecundity) and

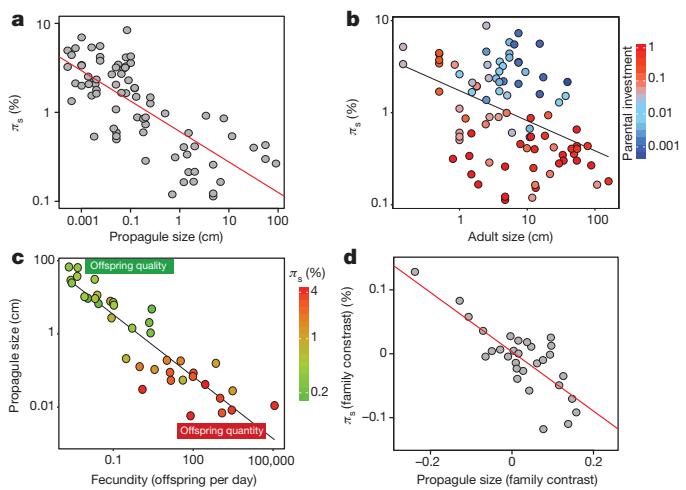


Figure 2 | Life-history traits and genetic diversity relationships. **a**, Relationship between propagule size and π_s ($P < 10^{-14}$, $r^2 = 0.56$, 76 species included; see Fig. 1). **b**, Relationship between adult size and π_s ($P < 0.05$, $r^2 = 0.07$, 76 species included). The colour scale represents the degree of parental investment, here defined as the ratio of propagule size to adult size. **c**, Effect of fecundity per day (x axis) and propagule size (y axis) on genetic diversity (colour scale; $P < 10^{-6}$, $r^2 = 0.69$, 29 family-averaged data points). **d**, Phylogenetic contrasts of family-averaged π_s versus family-averaged propagule size ($P < 10^{-6}$, $r^2 = 0.62$).

quality (propagule size) seems to be the most relevant factor explaining variations in polymorphism between species in the animal kingdom (Fig. 2c). We shall for simplicity hereafter categorize as *K*-strategists the species that tend to invest in the quality of their progeny, and as *r*-strategists those that favour quantity¹⁷.

The correlation we report between life-history traits and π_s is not due to phylogenetic non-independence of the sampled species: taking family averages from Fig. 1 increased the correlation coefficients (from $r^2 = 0.56$ to $r^2 = 0.66$ with propagule size alone, from $r^2 = 0.73$ to $r^2 = 0.79$ with the six life-history traits). When we took into account the between-family phylogenetic tree using independent contrasts, this still resulted in highly significant correlations between π_s and life-history traits ($r^2 = 0.62$ for propagule size; Fig. 2d and Extended Data Fig. 3). These relationships were also unaffected by sampling strategy, sequencing depth, gene expression levels or contaminants (Methods, Supplementary Table 4 and Extended Data Figs 4–6). Finally, our conclusions were unchanged when we included 14 previously published species of mammals¹⁰ or when we restricted the analysis to a subset of common orthologous genes (Supplementary Table 4).

The relationship between π_s and life-history traits, however strong, could in principle be mediated by causative variables that were not included in the analysis. One of these potential confounding factors is the mutation rate: a higher average per-generation mutation rate in *r*-strategists than in *K*-strategists could explain our results irrespective of the population size effect. However, theoretical models and empirical measurements actually support the opposite; that is, an increased per-generation mutation rate in large, long-lived organisms due to a larger number of germline cell divisions per generation and a reduced efficacy of natural selection on the fidelity of polymerases¹⁸. Therefore, as far as we can tell, across-species variations of mutation rate are likely to oppose, not strengthen, the main effect we are reporting here.

We computed the non-synonymous nucleotide diversity, π_n , and this was also found to be correlated with species life-history traits (Extended Data Fig. 2). We found substantial variation in π_n/π_s across metazoan species, and significant correlations with life-history traits, the best predictor in this case being longevity (Extended Data Fig. 7). This positive correlation is predicted by the nearly neutral theory of molecular evolution¹⁹: in small populations (long-lived species), the enhanced genetic drift counteracts purifying selection and promotes the segregation of weakly

deleterious, non-synonymous mutations at high allele frequency. These results also confirm that the relationships we uncovered between life-history traits and diversity patterns are mediated in the first place by an effect of N_e , not of the mutation rate; synonymous and non-synonymous positions being physically interspersed, the π_n/π_s ratio is unaffected by the mutation rate.

Our analysis reveals that polymorphism levels are well predicted by species biology, whereas historical and contingent factors are only minor determinants of the genetic diversity of a species. This unexpected result opens new questions. How can life-history traits be so predictive of π_s in spite of the overwhelming evidence for the impact of ecological perturbations on patterns of genetic variation^{11,12}? Why does the 'r/K' gradient affect genetic polymorphism so strongly?

In an attempt to resolve these paradoxes, we suggest that life-history strategies might influence the response of species to environmental perturbations. Because K-strategy species have been selected for survival and the optimization of offspring quality in complex, stable environments¹⁷, we speculate that they might experience fewer occasional disturbances (or be less sensitive to them), thus ensuring the long-term viability of even small populations. In contrast, only species with a large population-carrying capacity could sustain the 'riskier' r-strategy in the long term, thus buffering the frequent bottlenecks experienced in the context of high environmental sensitivity (see Supplementary Equations for a model formalizing these arguments). According to this hypothesis, environmental perturbations would be a common factor affecting every species, but their demographic impact would depend on the life-history strategy of each species.

This study highlights the importance of species life-history strategy when it comes to turning genetic diversity measures into conservation policy. So far, conservation efforts have mainly been focused on large-sized vertebrates. Here we show that these popular animals represent only a subset of the existing low-diversity, K-strategists. Invertebrate species with strong parental investment are probably equally vulnerable to genetic risks. Our results also indicate that r-strategists will typically show elevated amounts of genetic diversity irrespective of their current demography, which suggests that species of this kind might face significant extinction risks²⁰ without any warning genetic signal.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 March; accepted 17 July 2014.

Published online 20 August; corrected online 12 November 2014 (see full-text HTML version for details).

1. Lewontin, R. *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, 1974).
2. Leffler, E. M. et al. Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol.* **10**, e1001388 (2012).
3. Keller, L. F. & Waller, D. M. Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17**, 19–23 (2002).
4. Reusch, T. B. H., Ehlers, A., Hämerli, A. & Worm, B. Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proc. Natl Acad. Sci. USA* **102**, 2826–2831 (2005).
5. Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N. & Vellend, M. Ecological consequences of genetic diversity. *Ecol. Lett.* **11**, 609–623 (2008).
6. Johnson, W. E. et al. Genetic restoration of the Florida panther. *Science* **329**, 1641–1645 (2010).
7. Nair, P. Conservation genomics. *Proc. Natl Acad. Sci. USA* **111**, 569 (2014).

8. Bazin, E., Glémén, S. & Galtier, N. Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570–572 (2006).
9. Nabholz, B., Mauffrey, J.-F., Bazin, E., Galtier, N. & Glemin, S. Determination of mitochondrial genetic diversity in mammals. *Genetics* **178**, 351–361 (2008).
10. Perry, G. et al. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res.* **22**, 602–610 (2012).
11. Banks, S. C. et al. How does ecological disturbance influence genetic diversity? *Trends Ecol. Evol.* **28**, 670–679 (2013).
12. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000).
13. Ellegeren, H. Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* **29**, 51–63 (2014).
14. Cahais, V. et al. Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol. Ecol. Resources* **12**, 834–845 (2012).
15. Tsagkogeorga, G., Cahais, V. & Galtier, N. The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis*. *Genome Biol. Evol.* **4**, 740–749 (2012).
16. Gayral, P. et al. Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genet.* **9**, e1003457 (2013).
17. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, 1967).
18. Lynch, M. Evolution of the mutation rate. *Trends Genet.* **26**, 345–352 (2010).
19. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).
20. Pinsky, M. L. et al. Unexpected patterns of fisheries collapse in the world's oceans. *Proc. Natl Acad. Sci. USA* **108**, 8317–8322 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the following for providing samples: F. Delsuc, E. Douzery, M. Tilak, G. Dugas, S. Harispé, C. Benoist, D. Bouchon (woodlice), J. Bierne, M. Bierne, B. Houseaux, M. Strand, C. Lemaire, D. Lallias, Service Modèle Biologique Station Marine Roscoff (nemertines), X. Turon, S. Lopez-Legentil (Cystodytes), P. Jarne, P. David, R. Dillon, J. Auld, R. Relyea, C. Lively, J. Jokela, V. Poullain, T. Stewart (snails), S. Lapègue, V. Boulo, F. Batista, D. Lallias, L. Fast Jensen, M. Cantou (oysters), J. Do Nascimento, C. Daguin-Thiébaut, M. Cantou (crabs), L. Bonnaud (cuttlefish), D. Aurelle (gorgonians), F. Viard, Y. Pechenik, A. Cahill, R. Collins (slipper limpets), L. Dupont (earthworms), D. Jollivet (trumpet worms), M. A. Felix, I. Nuez (nematodes), N. Rodes, T. Lenormand, E. Flaven (brine shrimps), Rotterdam Zoo, Zurich Zoo, C. Libert, Montpellier Zoo, S. Martin, la Ferme aux Crocodiles, O. Verneau, C. Ayres, M. Carretero, M. Vanbergher, K. Pobolsaj, M. Zuffi, C. Palacios, L. du Preez, B. Halpern, Budapest Zoo (turtles), P. Peret, C. Doutrelant, B. Halpern, B. Rosival (tit), M. de Dinechin, B. Rey (penguins), Z. Melo-Ferreira, P. Alves (hares), N. Brand, M. Chapuisat (bees), R. Blatrix, A. Lenoir, I. Nodet, A. Lugagne, S. Blanquart, L. Serres-Giardi, V. Roustang, N. François, G. Ballantyne, A. Carbonnel, Y. Samuel, G. James, G. Kalyta, F. Guerrini, S. Stenzel, J. Beekman, X. Cerda, S. Ikoen (ants), I. Hanski, S. Ikonen, J. Kullberg, Z. Kolev (fritillary butterflies), F. Viard, X. Turon, Di Jiang, D. Chourrout, B. Vercaemer, E. Newman-Smith, Ascidian Stock Center, Service Modèle Biologique Station Marine Roscoff (ciona), L. Excoffier, G. Heckel (voles), F. Dedeine (termites), C. Atyame, O. Duron, M. Weill (mosquitoes), M. Cantou, H. Violette, F. Batista, J. Hondeville (seahorses), C. Fraisse, G. Pogson, N. Saarman, J. Normand (mussels), E. Poulin, C. Gonzalez-Weivar, and J. P. Feral (sea urchins). This work was supported by European Research Council advanced grant 232971 (PopPhy).

Author Contributions N.G. conceived the project. P.G., M.B., N.F., Y.C., L.A.W., G.T., A.C., A.W., J.R., N.G. and N.B. performed sampling and laboratory work. A.B., V.C., E.L., J.R., J.M.L., C.R., P.G., G.T., B.N., R.D., K.B., S.G. and N.G. developed the data analysis pipeline. J.R. collected life-history/geographic variables and produced figures. J.R., A.B., V.C., L.D., E.L. and N.G. analysed the data. S.G., N.B., B.N., J.R. and N.G. provided interpretations and models. J.R., N.B., S.G. and N.G. wrote the paper.

Author Information Data sets are freely available from the Sequence Read Archive (SRA) database (<http://www.ncbi.nlm.nih.gov/sra>) under project ID SRP042651 and from the Datasets section of the PopPhy website (<http://kimura.univ-montp2.fr/> PopPhy), in which predicted single nucleotide polymorphisms and genotypes are provided as .vcf files. Scripts and executable files are freely available from the Tools section of the PopPhy website. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.G. (nicolas.galtier@univ-montp2.fr).

METHODS

Sampling and sequencing. The 76 analysed species were selected based on phylogenetic and ecological criteria with the goal of gathering a panel representative of the metazoan diversity. In each species, from two to eleven individuals were collected in various localities of their natural geographic range (Supplementary Table 1) or from zoos (*Chelonoidis nigra*). Depending on the species, the whole body, body parts, organs or tissues were dissected and preserved in RNA later, flash-frozen, or processed immediately (Supplementary Table 2). For each sample, total RNA was extracted using standard and improved protocols²¹, and a non-normalized complementary DNA library was prepared. The libraries were sequenced on a Genome Analyzer II or Hiseq 2000 (Illumina) to produce 100-base-pair (bp) or 50-bp single-end fragments. In 12 species, an additional normalized random-primed cDNA library was prepared and sequenced for half a run using a 454 Genome Sequencer (GS) FLX Titanium Instrument (Roche Diagnostics). Illumina reads from two to four individuals in 14 mammalian species¹⁰ were downloaded from the SRA database. Reads were trimmed of low-quality terminal portions with the SeqClean program (<http://compbio.dfci.harvard.edu/tgi/>). The fastQ program was applied to Illumina reads and revealed only a limited amount of motif enrichment: the number of motifs in significant excess varied between 0 and 17 across species, its median being 1.

Transcriptome assembly, read mapping, coding sequence prediction. *De novo* transcriptome assembly based on the 454 (when available) and Illumina reads was performed by following strategies B and D in ref. 14, using a combination of the programs Abyss and Cap3. Illumina reads were mapped to predicted cDNAs (contigs) with the BWA program. Contigs with a per-individual average coverage below $\times 2.5$ were discarded. Open reading frames (ORFs) were predicted with the Trinity package. Contigs carrying no ORF longer than 200 bp were discarded. In contigs including ORFs longer than 200 bp, 5' and 3' flanking non-coding sequences were deleted, thus producing predicted coding sequences that are hereafter referred to as loci.

Calling single nucleotide polymorphisms (SNPs) and genotypes. At each position of each locus and for each individual, diploid genotypes were called according to the method described in ref. 15 (model M1) and improved in ref. 16, using the reads2snps program. This method first estimates the sequencing error rate in the maximum-likelihood framework, calculates the posterior probability of each possible genotype, and retains genotypes supported at $>95\%$; otherwise missing data are called. A minimum of ten reads per position and per individual were required to call a genotype. Then polymorphic positions were filtered for possible hidden paralogues (duplicated genes), using a likelihood ratio test based on explicit modelling of paralogy¹⁶. The across-species average percentage of SNPs discarded for suspicion of paralogy was 7.65%. Positions at which a genotype could not be called in a sufficient number of individuals were discarded. Calling k the number of sampled individuals for a given species ($2 \leq k \leq 11$), the minimum number of genotyped individuals required to retain a position was set to $k/2$ when $k > 5$, to $k - 1$ when $k = 4$ or $k = 5$, and to k when $k < 4$.

Control for contamination. Each locus of each species was translated to protein and compared with the non-redundant NR database using BlastP in search for possible contaminants. The percentage of loci for which no significant hit (e-value <0.001) was retrieved varied greatly between species, reflecting the taxonomic representation of sequences in NR. The percentage of no-hits was below 10% in mammals, but reached values above 20% in echinoderms and cnidarians. When one or several hits were found, the GenBank taxonomy of the first hit was recorded. Overall, 98.7% of first hits were assigned to Metazoa. The percentage of non-metazoan first hits was below 2% in 63 species out of 76, reaching its maximum (5.5%) in the trumpet worm *Pectinaria koreni*. Contamination by known microbes therefore seems negligible in our data set. Extended Data Figure 6 displays the taxonomic distribution of hits for four representative species—a mammal, an insect, a mollusc and an annelid. The results of our analyses were qualitatively unchanged when we removed loci that hit a non-metazoan and/or no-hit loci. In all cases, linear regression tests between π_s and propagule size yielded the same r^2 of 0.55 ($P < 10^{-13}$) as in our main analysis.

Life-history, ecological and geographical variables. Species life-history traits (adult size, juvenile size, body mass, longevity, fecundity and adult speed) were retrieved from the literature (Supplementary Table 3). Invasive/non-invasive status was obtained from the Global Invasive Status Database (<http://www.issg.org/database/species>List.asp>). The Global Biodiversity Information Facility database (<http://www.gbif.org/>) was queried to retrieve the GPS records corresponding to documented observations of individuals from the species of interest here. These data were merged with the GPS records of our own samples. For each species, the average and maximum distance between two distinct GPS records and the average distance to the Equator were computed (Supplementary Table 2).

Statistical analyses. Population genomic statistics π_s , π_n and F_{it} were calculated by using home-made programs that rely on the Bio++ libraries²². The F_{it} calculation was corrected for small sample sizes in accordance with ref. 23. Confidence intervals were obtained by bootstrapping loci. Regression analyses were conducted in R. Variables were log-transformed before linear regressions were performed. The linear model

including π_s and propagule size alone met the required assumptions of normally distributed residual errors (Shapiro's test, $P = 0.19$) and homoscedasticity (Fligner-Killeen's test, $P = 0.48$). The same remark is valid for the multiple linear model including π_s and the six life-history traits (Shapiro's test, $P = 0.31$; Fligner-Killeen's test, $P = 0.49$). Family-level phylogenetic independent contrast analysis was performed with the APE package based on the tree shown in Extended Data Fig. 3, in which branch lengths are proportional to time. Divergence time estimates were retrieved from the TimeTree database (expert result, or average value if expert result was missing). When divergence time estimates were not available (Polycitoridae–Cionidae, Hesperiidae–Nymphalidae, Calyptraeidae–Physidae, Mytilidae–Ostreidae), they were inferred on the basis of the divergence dates of neighbouring nodes.

SNP calling quality controls. The main analyses of this study were reproduced in three ways: first, with an increased minimum number of reads per position per individual of 30 instead of 10, second, removing five bases from each end of each read, and third, not using 454 data, thus controlling for a potential effect of insufficient sequencing depth, low-quality base calls near read ends and sequencing technology. In all three cases the results were highly similar to the main analysis (Supplementary Table 4; columns 'depth = 30X', 'clip_ends' and 'no_454', respectively), indicating that the analysis was robust to these technical caveats. No difference in π_s was detected between species showing versus not showing a significant excess of certain motifs by fastQC.

GC content. In each species, the correlation coefficient between contig GC content and contig π_s was calculated. It was significantly positive in 37 species, significantly negative in 18 species, and non-significantly different from zero in 21 species. The squared correlation coefficients (r^2) were relatively low (median $r^2 = 0.007$; maximum $r^2 = 0.16$ in *Physa acuta*), suggesting only a weak effect of GC content on π_s . For each species, the average contig GC content was calculated and correlated to the average π_s or propagule size. No significant relationship was detected, which indicates that the variation in GC content across genes and across species has no substantial impact on the results of this study.

Individual and locus sampling. No significant relationship was found between π_s and the number of sampled individuals per species, or between π_s and the number of sampled loci per species (Extended Data Fig. 4). The robustness to sampling of the relationship between propagule size and π_s was further assessed in two ways. First, for each species, loci were randomly subsampled. Extended Data Figure 5 displays the squared coefficient of correlation between propagule size and π_s as a function of the per species number of analysed loci. It shows that as few as 50 loci are enough to capture the relationship with a good probability. Second, for each species, only two individuals were randomly selected and the analyses were conducted again. Results were highly similar to the main analysis: the relationship between propagule size and π_s was unchanged and highly significant ($P < 10^{-15}$, $r^2 = 0.55$), thus indicating that population sample size is not an issue.

Orthologous genes. The coding sequences of 129 genes or gene fragments previously identified as orthologous across metazoans²⁴ (hereafter called 'core genes') were downloaded. In each of our species, contigs predicted to be orthologous to one of the core genes by reciprocal best BLAST hit were selected (expected e-value 0.0001, hits with a number of identical matches less than 80 and a bitscore of less than 1,200 were discarded). The number of such predicted core gene orthologues varied between 40 and 122 among species. We restricted the data set to the 39 species including at least 21 core gene orthologues, and reproduced the analysis. Colon 'orthologues' from Supplementary Table 4 shows that the correlation between π_s and life-history traits was still strong and significant when a subset of common genes was considered.

Expression level. In each species, the expression level of each locus was estimated as the average number of bases read per position. Correlating π_n/π_s to expression level across genes revealed no significant relationship in 33 species, and weak relationships ($r^2 < 0.27$) in 57 species. The relationship between π_n/π_s and expression level, when detected, was negative, as expected under the hypothesis or a stronger selective pressure acting on high-expressed genes. Then, for each species, loci were grouped into three equal-sized bins of genes with high, medium or low expression. Each of these categories, taken separately, provided a strong correlation between species propagule size and π_s ($r^2 = 0.57$, 0.62 and 0.62, respectively).

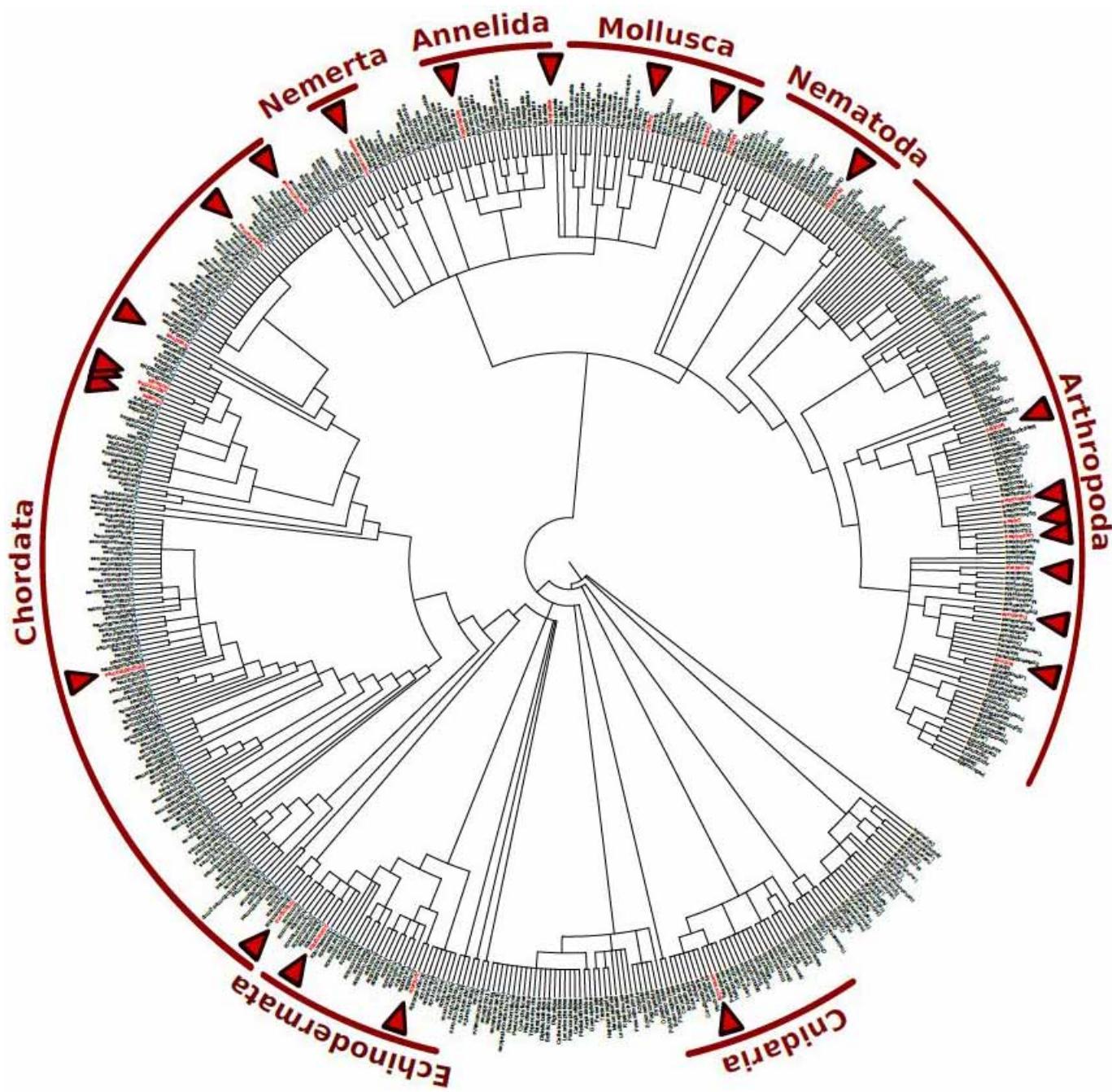
Linkage. The diversity of a neutral locus might be affected by selection at linked sites. This is particularly true of synonymous sites, which reside within coding sequences; that is, targets for natural selection. One would therefore predict a lower genetic diversity in species experiencing a low genomic average recombination rate. We lacked a recombination map in most of the analysed species; we therefore relied on generic taxonomic patterns to approach this issue. Eusocial hymenopterans are known to experience a recombination rate one order of magnitude higher than most animals²⁵. In contrast, dipteran Culicidae (mosquitoes) experience relatively small amounts of recombination (median recombination rate in eusocial hymenoptera, 9.7 centimorgans per megabase ($cM Mb^{-1}$); median recombination rate in Culicidae, 0.3 $cM Mb^{-1}$)²⁵. The linkage effect would therefore predict a decreased π_s in Culicidae and an elevated π_s in eusocial hymenopterans. We observed the opposite: π_s varied from 0.0016 to 0.0058 in our five eusocial hymenopteran species, which is below the average metazoan

π_s (0.015), and one order of magnitude below the π_s of our three Culicidae species (0.016–0.041). This result, which is consistent with the propagule-size hypothesis, does not suggest that the between-species variation in genomic average recombination rate strongly influences our results.

Population structure. The genetic distance between individuals was defined as $(H_b - H_w)/H_w$, where H_b is the probability of drawing two distinct alleles when sampling one copy from each of the two considered individuals, and H_w is the average heterozygosity of the two considered individuals. In species containing more than four individuals, the genetic distance was calculated for each pair of individuals and correlated to the geographic distance; the squared coefficient correlation, r^2 , which measures genetic isolation by distance, ranged from 0.0008 to 0.73 (Supplementary Table 5). Consistent with the phylogeographic literature, it was high ($r^2 > 0.35$) and significant in, for example, *Ciona intestinalis* A, *Melitaea cinxia* and *Sepia officinalis*, and low ($r^2 < 0.02$, n.s.) in, for example, *Culex pipiens*, *Lepus granatensis* and *Crepidula fornicate*. The F_{it} statistic was significantly higher, on average, in terrestrial (median $F_{it} = 0.25$) than in marine (median $F_{it} = 0.02$) species (*t*-test, $P = 0.029$) when only species including at least five individuals were considered. No significant relationship was detected between π_s and absolute values of F_{it} ($P = 0.22$, $r^2 = 0.05$, only species with more than four individuals included), which does not suggest any confounding effect of population structure in our analysis.

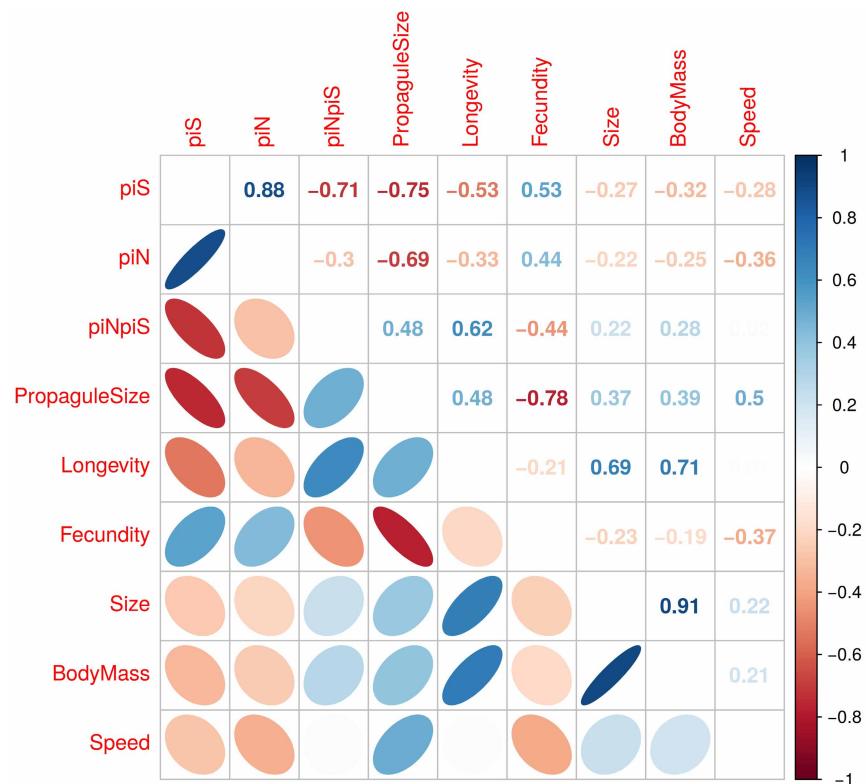
Ethical statement. Living animals were manipulated according to the ‘Charte Nationale Portant sur l’Ethique de l’Expérimentation Animale’. Sampling of protected species was performed under permits 53/2009 (Galicia, Spain, *Emys orbicularis*), 2009/11/12 (Aude, France, *Emys orbicularis*), 503/05/07/2006 (Pisa, Italy, *Emys orbicularis*), 35601-60/2005-4 (Slovenia, *Emys orbicularis*), and 009-01-1230/a34-455 (France, *Parus caeruleus*). *Aptenodytes patagonicus*, *Eudyptes moseleyi* and *Eudyptes filholi* individuals were sampled by Institut Polaire Français Paul Emile Victor, program IPEV 131. *Chelonoidis nigra* individuals were handled and sampled by the veterinarians and staff of the Zurich zoo (Switzerland), Rotterdam zoo (the Netherlands), and A Cupulatta zoo (France) in accordance with the Code of Practice and Code of Ethics established by the European Association of Zoos and Aquaria.

21. Gayral, P. et al. Next-generation sequencing of transcriptomes: a guide to RNA isolation in non-model animals. *Mol. Ecol. Resources* **11**, 650–661 (2011).
22. Guéguen, L. et al. Bio++: efficient, extensible libraries and tools for molecular evolution. *Mol. Biol. Evol.* **30**, 1745–1750 (2013).
23. Weir, B. S. & Cockerham, C. C. Estimating F -statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
24. Philippe, H. et al. Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19**, 706–712 (2009).
25. Wilfert, L., Gadau, J. & Schmid-Hempel, P. Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity* **98**, 189–197 (2007).

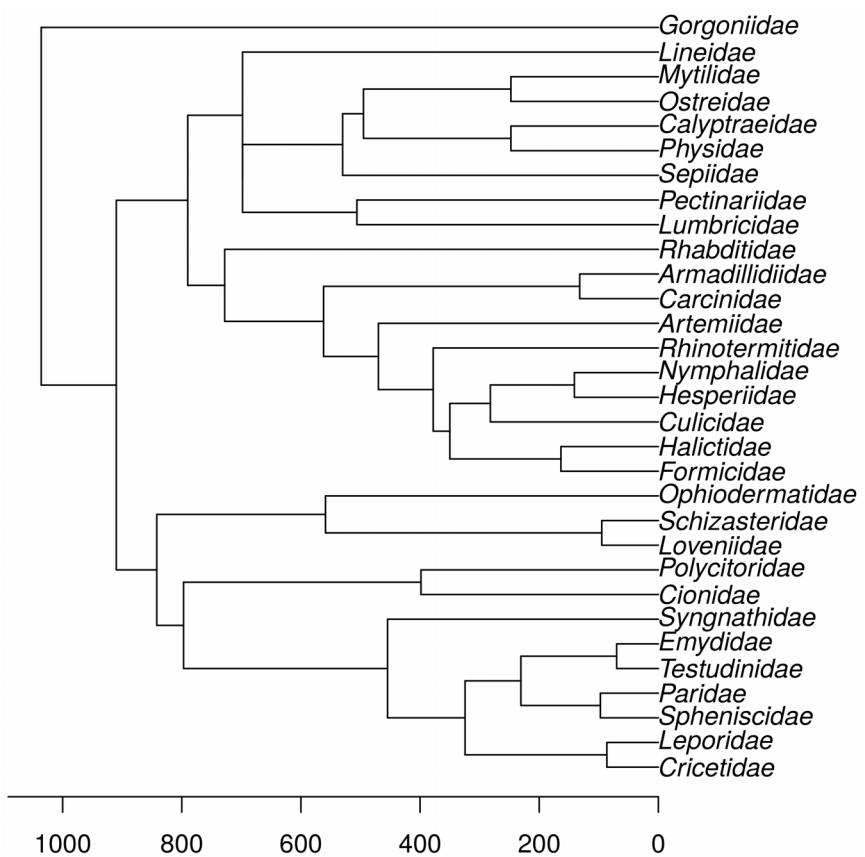


Extended Data Figure 1 | Phylogenetic tree of metazoan orders and the position of the taxa analysed in this study. The tree topology is consistent with the NCBI taxonomy. Red arrows identify 25 orders that were sampled. Five gastropod species from two distinct families (Calyptaeidae (*Crepidula*

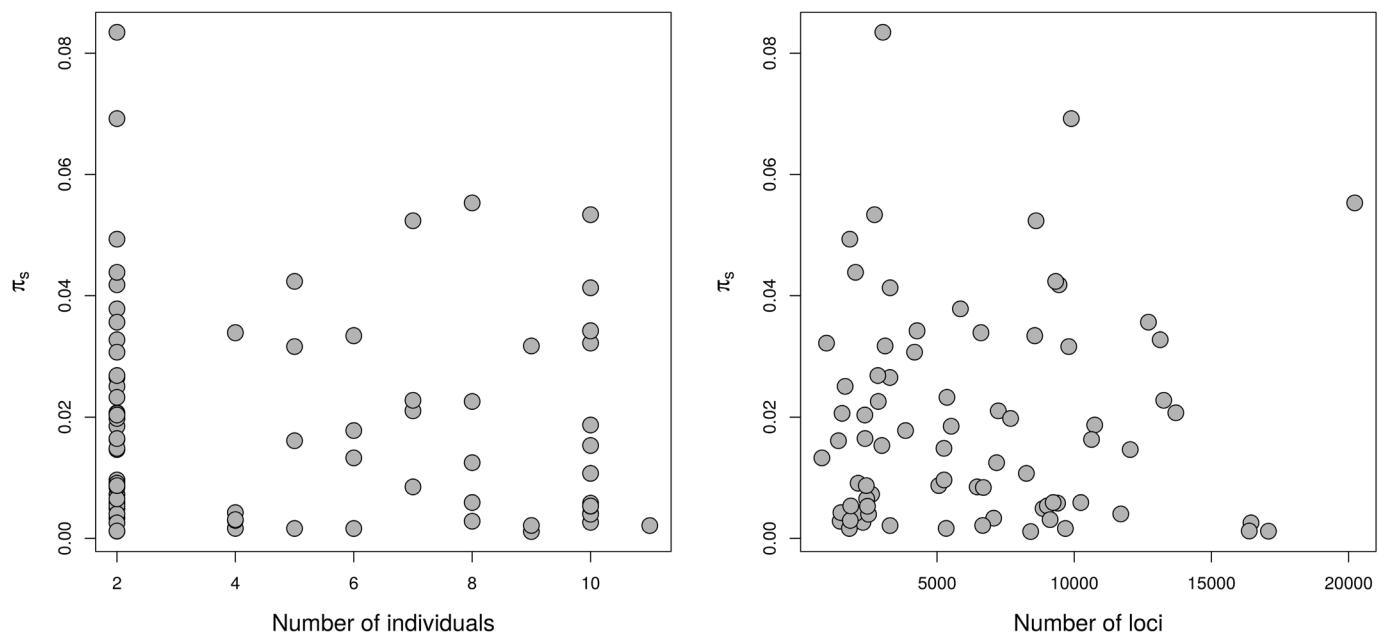
*fornicata, C. plana and *Bostrycapulus aculeatus*) and Physidae (*Physa acuta* and *P. gyrina*)) are not represented because they lacked any assignment to an order in current taxonomy.*



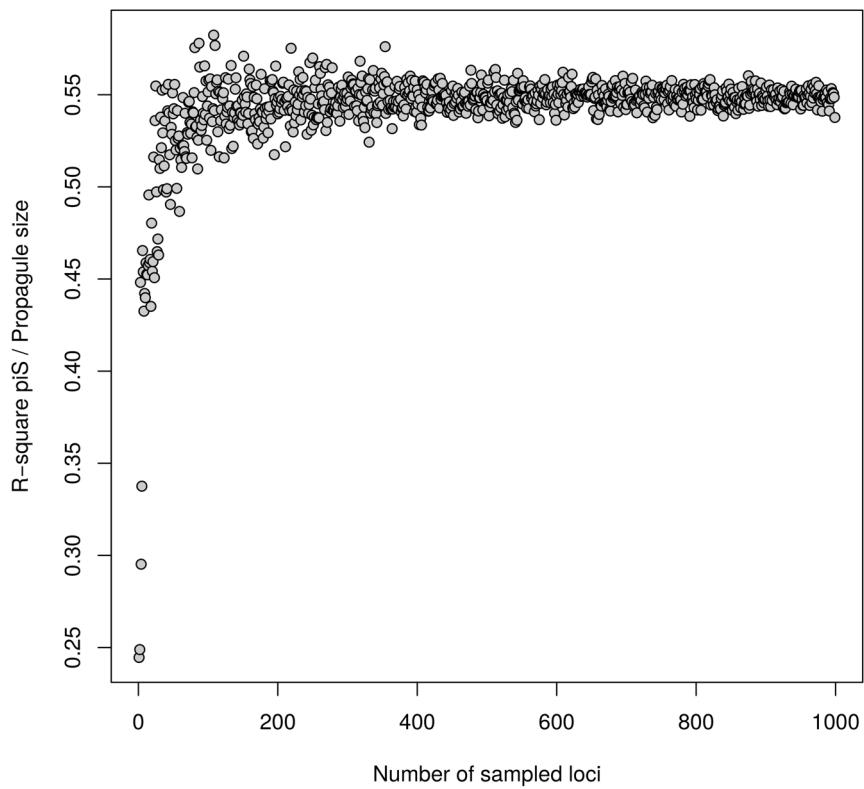
Extended Data Figure 2 | Correlations between genetic diversity and life history variables. Blue indicates a positive relationship, red a negative one; colour intensity is proportional to Pearson's correlation coefficient.



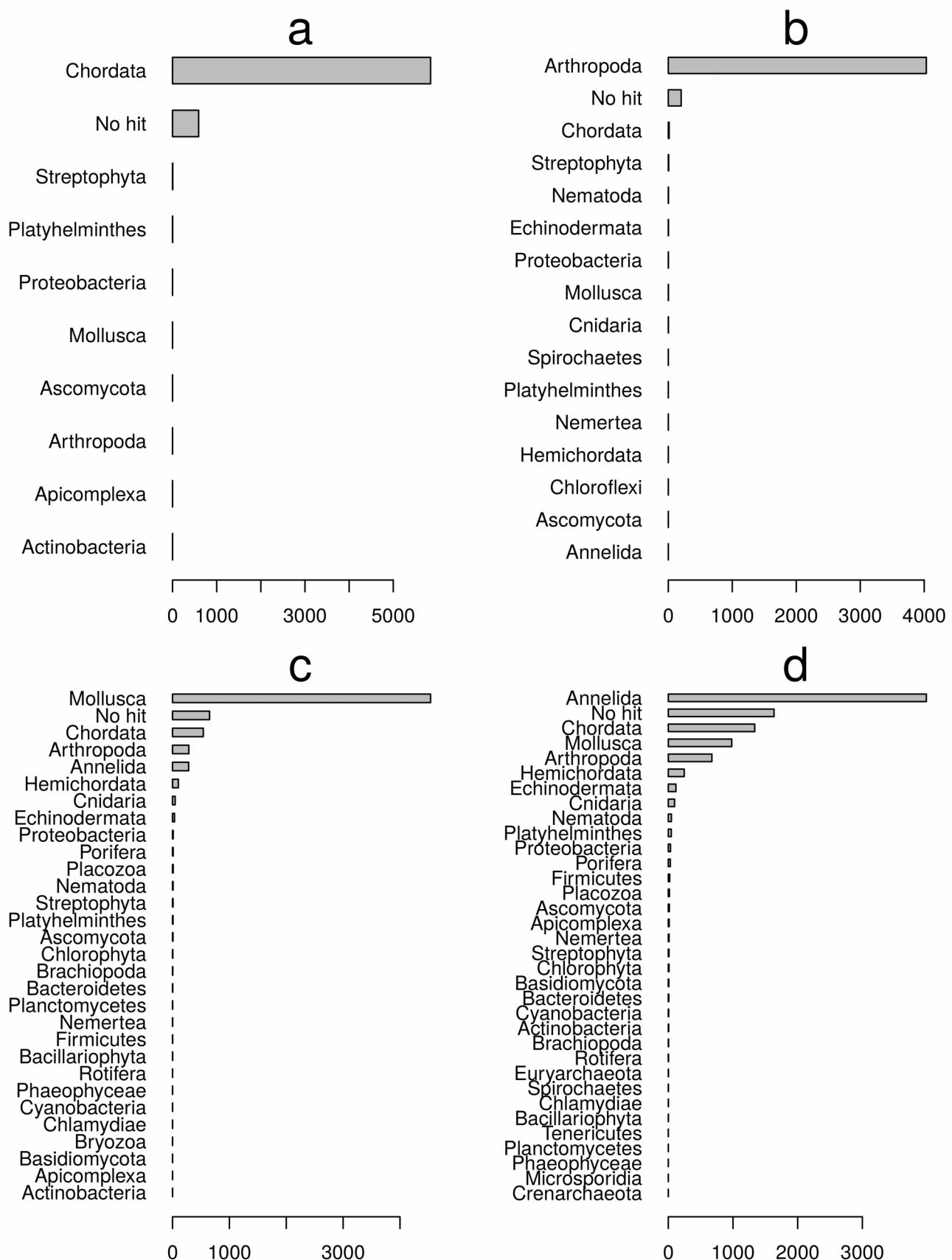
Extended Data Figure 3 | Family-level phylogenetic tree (31 families included). The scale is in million years of divergence.



Extended Data Figure 4 | Absence of significant correlation between species genetic diversity with individual sampling size ($P = 0.47, r^2 = 0.007$) and locus sampling size ($P = 0.78, r^2 = 0.001$).

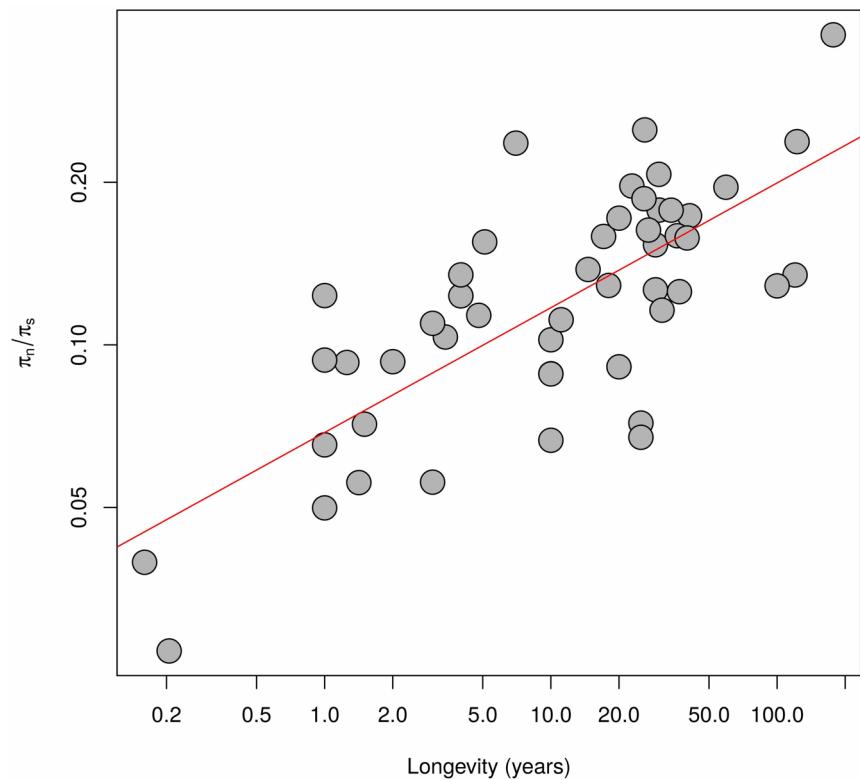


Extended Data Figure 5 | Relationship between the π_s /propagule-size r^2 and the number of sampled loci.



Extended Data Figure 6 | Phylum distribution of the first BLAST hit in four representative species. **a**, Common vole (*Microtus arvalis*). **b**, Glanville

fritillary butterfly (*Argynnis paphia*). **c**, Blue mussel (*Mytilus edulis*). **d**, Earthworm (*Allolobophora chlorotica*).



Extended Data Figure 7 | Correlation between π_n/π_s and maximum longevity ($P < 10^{-8}$, $r^2 = 0.54$). Only species with at least four individuals are included.

Convergent evolution of social hybridogenesis in *Messor* harvester ants

JONATHAN ROMIGUIER, AXEL FOURNIER, SZE HUEI YEK and LAURENT KELLER
Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland

Abstract

Sexual reproduction generally requires no more than two partners. Here, we show convergent evolution of social hybridogenesis, a reproductive system requiring three reproductive partners in harvester ants. In this unorthodox reproductive system, two distinct genetic lineages live in sympatry and queens have to mate with males of their own lineage to produce queens along with males of the alternative lineage to produce workers. Using a large transcriptomic data set of nine species, we show that social hybridogenesis evolved at least three times independently in the genus *Messor*. Moreover, a study of 13 populations of *Messor barbarus* revealed that this mode of reproduction is fixed in the whole range of this ecologically dominant species. Finally, we show that workers can produce males carrying genes of the two genetic lineages, raising the possibility of rare gene flow between lineages contributing to the long-term maintenance of pairs of interdependent lineages. These results emphasize the evolutionary importance of social hybridogenesis, a major transition possibly linked to the peculiar ecology of harvester ants.

Keywords: behavior/social evolution, bioinformatics/phylogenomics, evolution of sex, genomics/proteomics

Received 27 July 2016; revision received 26 September 2016; accepted 12 October 2016

Introduction

Several organisms have evolved unusual modes of reproduction, comprising both components of asexual and sexual reproduction. In some species of social insects, by example, queens are produced by parthenogenesis while workers are produced by normal sexual reproduction (Pearcy *et al.* 2004, 2011; Fournier *et al.* 2005, 2016; Kobayashi *et al.* 2008; Matsuura *et al.* 2009). Using alternative modes of reproduction for the queen and worker castes, queens can increase the transmission rate of their genes to their reproductive female offspring while maintaining genetic diversity and social cohesion in the worker force (Smith *et al.* 2008). The ant *Cataglyphis hispanica* has evolved a similar mode of reproduction whereby parthenogenesis also produces new queens but workers are the product of hybrid matings between two distinct genetic lineages that coexist within a single population (Leniaud *et al.* 2012). This reproductive system

relying on two dependent lineages to produce hybrid workers has been called social hybridogenesis and was first described in *Pogonomyrmex* harvester ants (Helms Cahan *et al.* 2002; Julian *et al.* 2002; Volny & Gordon 2002; Helms Cahan & Keller 2003). Contrary to the *Cataglyphis* case, social hybridogenesis in *Pogonomyrmex* does not involve parthenogenesis and the ensuing short-term advantage in terms of higher transfer of genes to the reproductive females. In some populations of *Pogonomyrmex*, two distinct genetic lineages coexist and queens mate multiply with males of their own and the alternate lineage; offspring produced from same-lineage matings always develop into queens, whereas interlineage hybrids develop into workers (Fig. 1A).

A remarkable feature of the *Pogonomyrmex* social hybridogenesis system is that the production of new fertile colonies requires the involvement of at least three mating partners – one female and two males, one from each lineage. Queens that only mate with males of different lineage fail to produce new queens, while queens that only mate with males of their own lineage fail to produce workers and the colony is therefore not viable. Understanding the origin and

Correspondence: Jonathan Romiguier and Laurent Keller, Fax: +41(0)21 692 41 73; Emails: jonathan.romiguier@gmail.com; laurent.keller@gmail.com

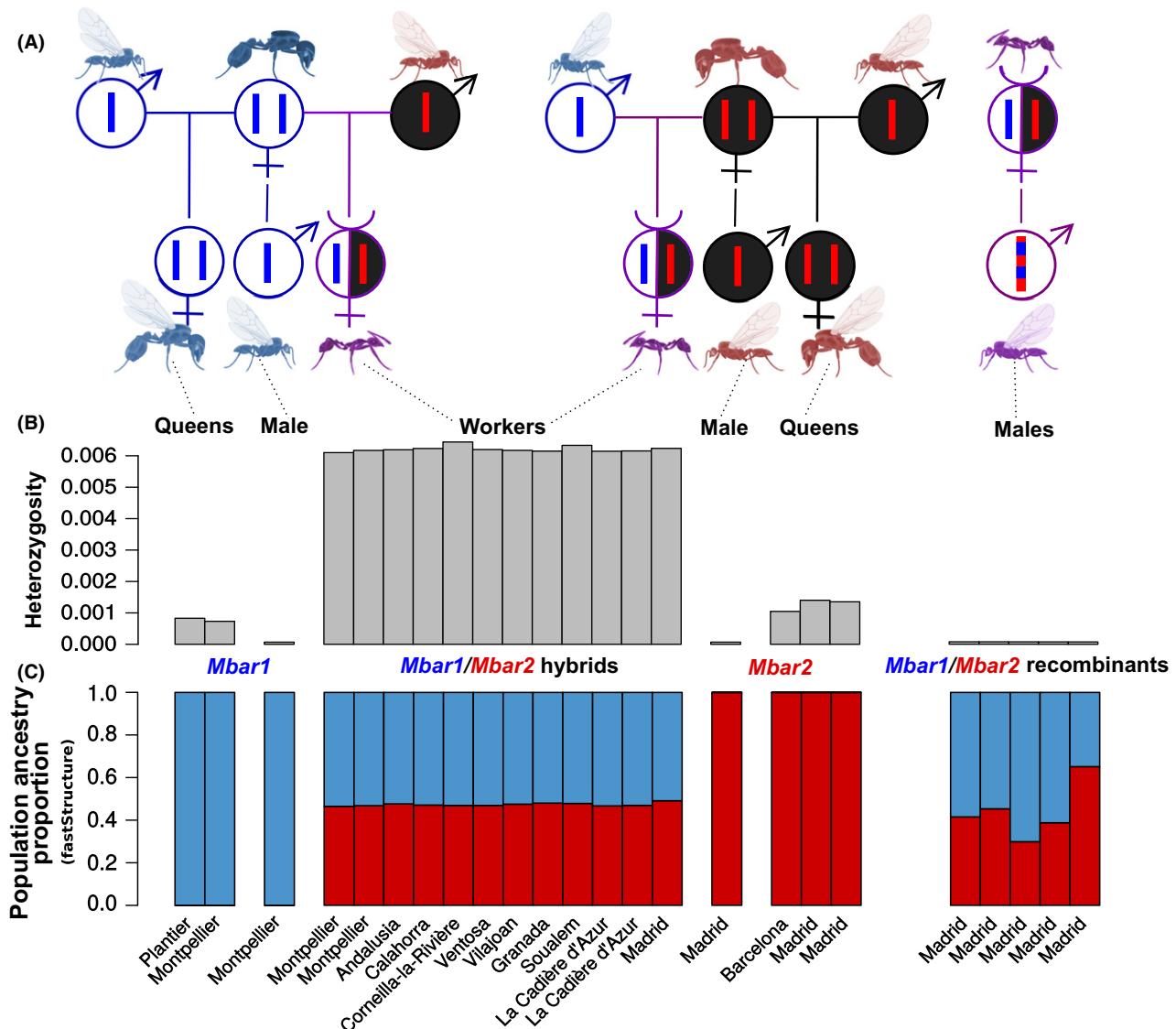


Fig. 1 Social hybridogenesis in *Messor barbarus*. (A) Social hybridogenesis reproductive system where two genetic lineages (blue and red) coexist within each population. Queens mate with males from both lineages. Pure lineage matings produce new queens while interlineage matings produce workers. (B) Workers of *M. barbarus* have higher genomewide heterozygosity levels (based on 90 951 SNPs) than queens (all pairwise comparisons significant, $P < 0.0001$, Wilcoxon tests). (C) Population ancestry proportions estimated by *fastStructure* in 24 individuals. Each bar corresponds to individuals from (A) and (B). Bars are named according to the population of origin of the individuals. Queens are pure *Mbar1* (blue) or *Mbar2* (red) lineages; workers are *Mbar1/Mbar2* hybrids. Males laid by queens are pure lineages, and males laid by workers are *Mbar1/Mbar2* recombinants.

maintenance of this unique system requiring three different mating partners has attracted considerable interest (Anderson *et al.* 2006, 2008; Linksvayer *et al.* 2006; Schwander *et al.* 2006, 2007a, b; Volny *et al.* 2006; Yamauchi & Yamamura 2006; Curry *et al.* 2009; Sirviö *et al.* 2010; Suni & Gordon 2010; Gordon *et al.* 2013; Mott *et al.* 2015). However, no new cases have been found since its discovery (Julian *et al.* 2002). So far, this reproductive system is restricted to populations thought to result from ancestral hybridization between *P. rugosus* and *P. barbatus*, two species with a standard

mode of reproduction (Schwander *et al.* 2007a). It is unclear whether such systems are oddities resulting from rare historical events or whether they can be favoured under certain ecological conditions. Another important issue inherent to social hybridogenesis is the long-term maintenance of interdependent lineages (Yamauchi & Yamamura 2006). As interlineage matings give rise only to workers, interlineage gene flow is suppressed, possibly leading to genetic incompatibilities between the increasingly divergent lineages (Anderson *et al.* 2008).

In this study, we show that the evolution of social hybridogenesis is not as rare as initially thought. Our genetic analyses reveal that three species of the genus *Messor* have such a mode of reproduction. Moreover, we show that this mode of reproduction evolved at least four times independently in harvester ants, suggesting that the peculiar ecology of these seed-eating ants might facilitate such a transition. Finally, we demonstrate that workers can reproduce by parthenogenesis and produce males with a chimeric composition of the two genetic lineages, paving the way to long-term maintenance of social hybridogenesis.

Methods

Sampling for transcriptome sequencing

To uncover the reproductive system of *Messor* species, we sampled 52 individuals across 29 populations from nine species (*M. barbarus*, *M. capitatus*, *M. structor*, *M. cf decipiens*, *M. aciculatus*, *M. cf hellenius*, *M. minor hesperius*, *M. wasmanni* and *M. ebeninus*) for RNA sequencing (see Table S1, Supporting information for details). To determine whether queens are produced parthenogenetically or sexually in *M. barbarus*, we sampled the mother queen, one of her sons and two daughters (a new queen and a worker) from a monogyne colony raised in laboratory conditions since 15 years (colony noted *Madrid1* in Table S1, Supporting information). To check whether the genomes of the dependent lineages (hereafter named *Mbar1* and *Mbar2*) can recombine through worker parthenogenesis, we sequenced five males laid by orphan workers from the same *M. barbarus* laboratory colony. For all species except *M. barbarus*, we sampled at least one mated queen after a mating flight and waited for the emergence of the first workers in the laboratory to sequence a mother queen and one of her daughter worker.

Transcriptome sequencing and assembly

The whole body of each individual was flash-frozen before RNA extraction. Total RNA was extracted using specific protocols for ants (Gayral *et al.* 2011). Complementary libraries were prepared using an Illumina TrueSeq preparation kit. We sequenced these libraries on a HiSeq 2000 (Illumina) to produce 100-base pairs (bp) paired-end reads. We used *Trimmomatic* to remove adapters and reads with length less than 60 bp and average quality less than 30 (Bolger *et al.* 2014). *De novo* transcriptome assemblies of each species were performed using a combination of *Abyss* and *Cap3* to treat reads of all individuals of a given species, following previously described methods (Romiguier *et al.* 2014a).

Computing population genetic statistics

We followed the pipeline of Romiguier *et al.* (2014a). Illumina reads of each individual were mapped to the *de novo* transcriptome assembly of its corresponding species using the *BWA* program (Li & Durbin 2010), and we discarded contigs with a per-individual average coverage below $\times 2.5$. We used the *TRINITY* package to predict open reading frames (ORFs). We discarded contigs carrying ORFs less than 200 bp. In contigs with ORFs longer than 200 bp, we deleted 5' and 3' flanking noncoding sequences, thus producing predicted coding sequences that are hereafter referred to as loci. We used the *READ2SNP* program and followed previously used methods to call genotypes (Romiguier *et al.* 2014a). To infer the reproductive system of *Messor* species, we calculated *Fis* and heterozygosity values for each locus and each species/individual using the tool *DNDSPINPIS* (<http://kimura.univ-montp2.fr/PopPhyl/index.php?section=tools>). In case of social hybridogenesis, we expect an excess of heterozygosity (high *Fis*) in workers which are hybrids between lineages compared to queens which are of pure lineage. *Fis* calculation was corrected for small sample size using the Weir–Cockerham correction (Weir & Clark Cockerham 1984). Confidence intervals were obtained by bootstrapping loci. We compared the average locus heterozygosity and the average locus *Fis* among individuals using R and non-parametric Wilcoxon tests. Finally, we conducted population structure analyses using *fastStructure* (Raj *et al.* 2014) with default parameters and a number of populations of $K = 2$.

Sampling and DNA sequencing

To quantify the dependent lineage distribution (hereafter called *Mbar1* and *Mbar2*) across colonies of two populations (Montpellier and Lunel, France), we sampled six colonies in each population during a mating flight in September 2013. We performed Sanger sequencing on a nuclear gene (*hsc70-4*) identified from transcriptomic data as carrying six fixed mutations between the *Mbar1* and *Mbar2* lineages. We sequenced the *hsc70-4* nuclear gene in 56 males, 37 alate queens and 52 workers along with the *cox1* mitochondrial gene in 27 males and 21 alate queens. We extracted DNA using the 'BS96 DNA Tissue' protocol of the *BioSprint 96* extractor robot. We purified PCR products with the 'Wizzard Genomic DNA Purification' Promega kit.

Phylogenetic analyses

We used *OrthoMCL* (Li 2003) to retrieve 2070 1-1 ortholog genes among the nine *de novo* assemblies of *Messor*

species + the official gene set (version 1.2) of *Pogonomyrmex barbatus*. To build a *Messor* phylogeny, we divided species featuring social hybridogenesis (*M. barbarus*, *M. structor* and *M. ebeninus*) in two dependent lineages defined from the *fastStructure* analyses (hereafter called *Mbar1/Mbar2*, *Mstr1/Mstr2* and *Mebe1/Mebe2*). For this, we built consensus sequences of queens grouped according to their respective lineages (2 *Mbar1*, 3 *Mbar2*, 1 *Mstr1*, 3 *Mstr2* and 1 *Mebe1*) or species (1 *M. cf. decipiens*, 2 *M. capitatus*, 1 *M. wasmanni*, 1 *M. minor*, 2 *M. cf. hellenius* and 1 *M. aciculatus*). We used the paternal alleles of the *M. ebeninus* hybrid worker (daughter of the *Mebe1* queen) as *Mebe2* sequences. The resulting 2070 1-1 ortholog genes were aligned using MACSE (Ranwez *et al.* 2011) and then concatenated in a supermatrix that was cleaned using the automated method included in *trimal* (Capella-Gutierrez *et al.* 2009). All phylogenetic reconstructions were performed using RAXML (GTR + GAMMA model, 500 bootstrap replicates, *P. barbatus* used as the outgroup; Ranwez *et al.* 2011; Stamatakis 2014).

Results

Messor barbarus workers are hybrids between two distinct genetic lineages

The analysis of 19 *M. barbarus* transcriptomes (two males, five queens and 12 workers) from 12 different populations across the Mediterranean Basin (Table S1, Supporting information) revealed that workers are produced by social hybridogenesis. The workers exhibited extremely high levels of heterozygosity ($F_{IS} = -0.666$ [-0.671 , -0.662], 12 532 genes) in stark contrast to queens which exhibited a strong excess of homozygosity ($F_{IS} = 0.746$ [0.738, 0.753], 5597 genes; Fig. 1B). An analysis of 126 967 SNPs with *fastStructure* (Raj *et al.* 2014) revealed that two queens and one male belonged to the same lineage (hereafter called *Mbar1*, synonymous nucleotide diversity of 0.0023), three queens and one male to another lineage (hereafter called *Mbar2*, synonymous nucleotide diversity of 0.0039) while all workers were inferred to be a 50/50 mix of these two genetic groups (Fig. 1C). This unusual pattern suggests social hybridogenesis, where queens belong to two distinct genetic lineages and workers are hybrids of both lineages.

To quantify the *Mbar1/Mbar2* lineage distribution across colonies of two populations (Montpellier and Lunel, France), we sampled males ($n = 56$), queens ($n = 37$) and workers ($n = 52$) from 12 colonies (six from each population) and sequenced a nuclear gene (*hsc70-4*) identified from transcriptomic data as carrying six fixed mutations between the *Mbar1* and *Mbar2* lineages.

All reproductive individuals (23 males and 18 queens) from six colonies (three in each population) carried the six fixed mutations corresponding to the *Mbar1* lineage, while all reproductive individuals (33 males and 19 queens) from the six remaining colonies (three in each population) carried the six fixed mutations corresponding to the *Mbar2* lineage. Remarkably, all workers ($n = 52$) were *Mbar1/Mbar2* hybrids, with the six *hsc70-4* positions that carry *Mbar1/Mbar2* substitutions being heterozygous for every worker (Table S2, Supporting information). We also sequenced a mitochondrial gene (*cox1*) in a subset of reproductive individuals of the 12 colonies (27 of the 57 males and 21 of the 37 queens) to build a phylogenetic tree. This analysis confirmed the presence of two clear monophyletic clades (bootstrap of 100), grouping the individuals according to their *Mbar1/Mbar2* lineage assignation (Fig. 2).

Sexual origin of *Messor barbarus* queens and worker reproduction

To determine whether queens are parthenogenetically or sexually produced, we sequenced the whole transcriptome of a mother queen and a daughter queen. While the mother and daughter queens had similar heterozygosity levels (0.00140 and 0.00135, respectively), 50.1% of the heterozygous positions of the daughter (368/734 positions) were homozygous in the mother, indicating that the daughter was sexually produced. Sexual production of queens indicates that social hybridogenesis uncovered in *M. barbarus* is similar to the system of social hybridogenesis described in *Pogonomyrmex* (Fig. 1A; Volny & Gordon 2002).

To investigate whether orphan workers can lay eggs, we isolated in the laboratory 100 *M. barbarus* workers from a *Mbar2* colony after removing all the brood from the nest. We observed 16 eggs that gave rise to 16 males, hatching 25 days after the isolation. To check whether the genomes of dependent lineages *Mbar1* and *Mbar2* can recombine through worker parthenogenesis, we sequenced the transcriptome of five of the 16 males. These five males were all haploid and had a mix of alleles coming from the *Mbar1* and *Mbar2* lineages (*fastStructure* revealed that these males had 42%, 45%, 30%, 39% and 65% of their alleles corresponding to the *Mbar2* lineage, that is the lineage of the queen in their colony, Fig. 1C).

At least three independent origins of social hybridogenesis in *Messor*

To investigate whether other cases of social hybridogenesis occur in the genus *Messor*, we sequenced the transcriptome of a mother queen and a daughter worker in

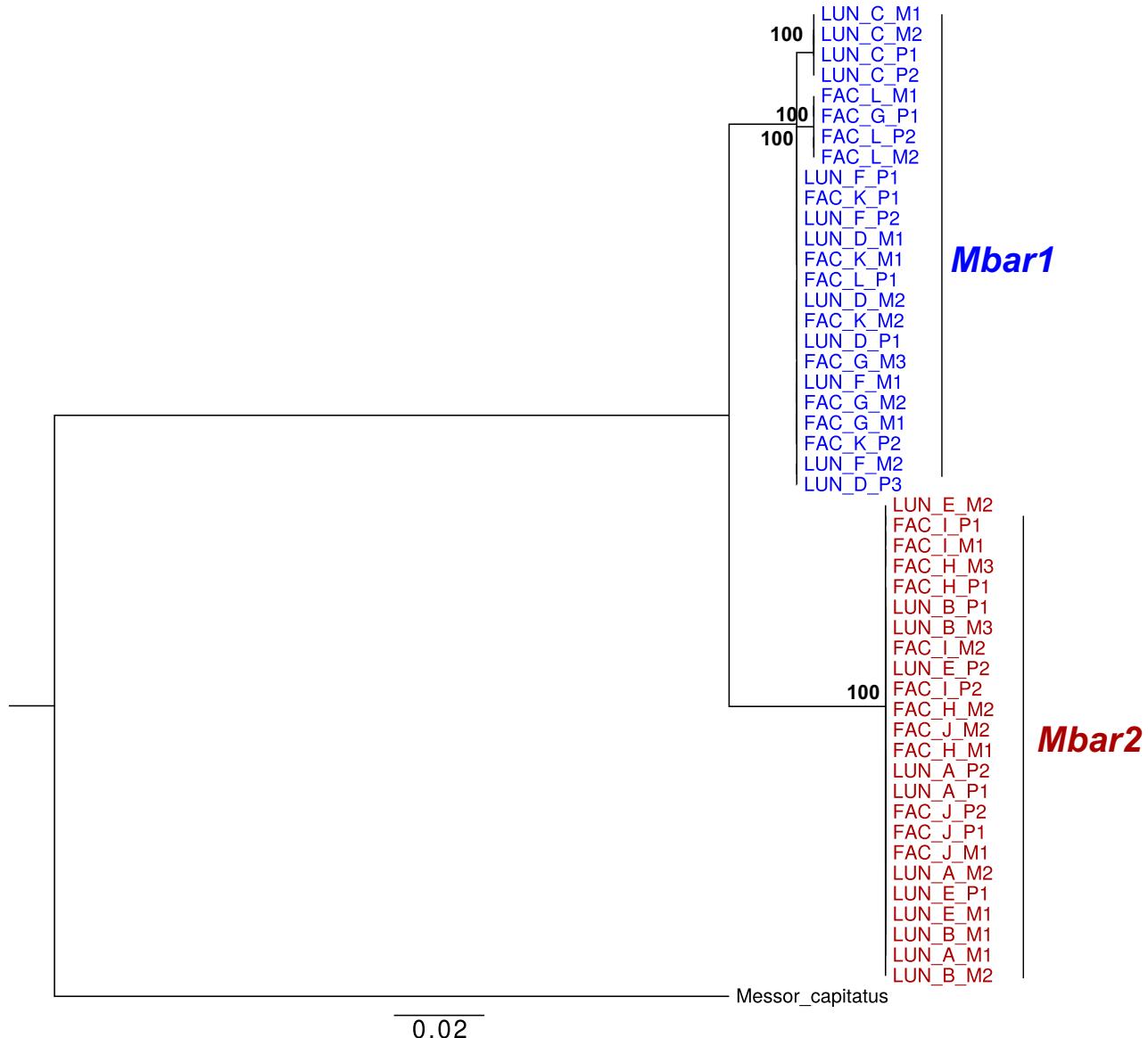


Fig. 2 Phylogenetic tree of 48 *M. barbarus* males and queens based on the *cox1* mitochondrial gene. Tip names correspond to the ID field in Table S2 (Supporting information). First three letters correspond to the population of origin (either FAC or LUN), 4th letter corresponds to the ID of one of the 12 colonies (from A to L) and the 5th letter corresponds to the caste (P for virgin queen and M for male). Individuals assigned to the *Mbar1* lineage are in blue, and individuals assigned to the *Mbar2* lineage are in red (assignment based on *hsc70-4* mutations, see Table S2, Supporting information). [Colour figure can be viewed at wileyonlinelibrary.com]

nine *Messor* species (including *M. barbarus*). Of the nine species, three (*M. barbarus*, *M. structor* and *M. ebeninus*) exhibited significant differences in the heterozygosity levels of the queen and the worker (Wilcoxon test, $P < 0.0001$ in the three cases), suggesting hybrid workers and pure lineage queens as typically seen in social hybridogenesis. The heterozygosity level of the queen and the worker was similar in the six remaining species (Wilcoxon test, NS in the six species, Fig. 3). To further explore the case of *M. structor*, we sequenced four additional queens and four additional workers from six

different populations. Similar to *M. barbarus*, *fastStructure* revealed that queens belonged to two distinct genetic lineages (*Mstr1* with one queen, *Mstr2* with three queens) while workers were all hybrids between these two lineages (Fig. S1, Supporting information).

To determine the number of origins of social hybridogenesis, we built a molecular phylogeny of the *Messor* genus separating each hybridogenetic species (*M. barbarus*, *M. structor* and *M. ebeninus*) in two dependent lineages (respectively, *Mbar1/Mbar2*, *Mstr1/Mstr2* and *Mebe1/Mebe2*). Each pair of dependent

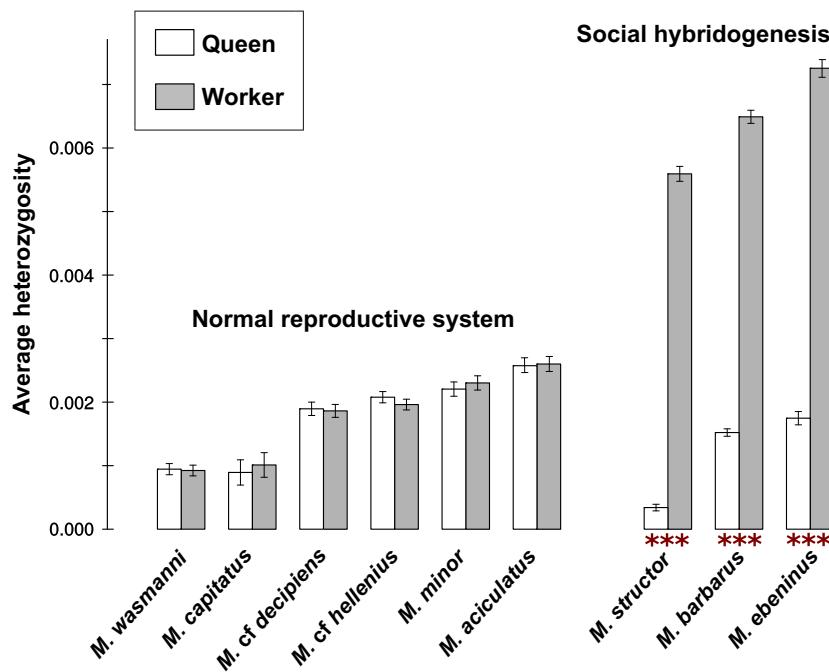


Fig. 3 Average locus heterozygosity comparisons between a mother queen and her daughter worker in nine *Messor* species. Three species have a significant heterozygosity difference characteristic of social hybridogenesis. [Colour figure can be viewed at wileyonlinelibrary.com]

lineages belonged to three distinct clades (Fig. 4), which suggests that social hybridogenesis (i.e. the occurrence of dependent genetic lineages) evolved three times independently. The two lineages of *M. barbarus* (*Mbar1* and *Mbar2*) clustered together (bootstrap support of 100). In contrast, one of the *M. structor* lineages (*Mstr2*) clustered with *M. cf hellenius* (bootstrap support of 100) rather than the other *M. structor* lineage (*Mstr1*). Similarly, the *M. ebeninus* lineage *Mebe2* clustered with *M. wasmanni* (bootstrap support of 100) instead of *Mebe1*.

Because social hybridogenesis results in a strict system of genetic caste determination (Schwander *et al.* 2010), we searched for genes where one given nucleotide was heterozygous in all hybrid workers ($n = 17$) of the three species with social hybridogenesis (*M. barbarus*, *M. structor* and *M. ebeninus*) and homozygous in queens of each of the six lineages ($n = 11$). There were only two such genomic regions, one for which the substitution induced a protein change. This was in *catalase* (position 1021 in the alignment available as Supplementary Material), a gene known to affect honeybee caste differentiation during early larval development (Cameron *et al.* 2013). The codon position varies between TTC and CTC across the phylogeny of *Messor*, which code, respectively, for the Phenylalanine and Leucine amino acids (Fig. 5). Interestingly, the *Messor* species with normal reproductive system (*M. cf decipiens*, *M. capitatus*, *M. wasmanni*, *M. minor*, *M. cf hellenius* and *M. aciculatus*) were always fixed for one of the two variants with all queens ($n = 8$) and workers ($n = 10$) being homozygous (Fig. 5).

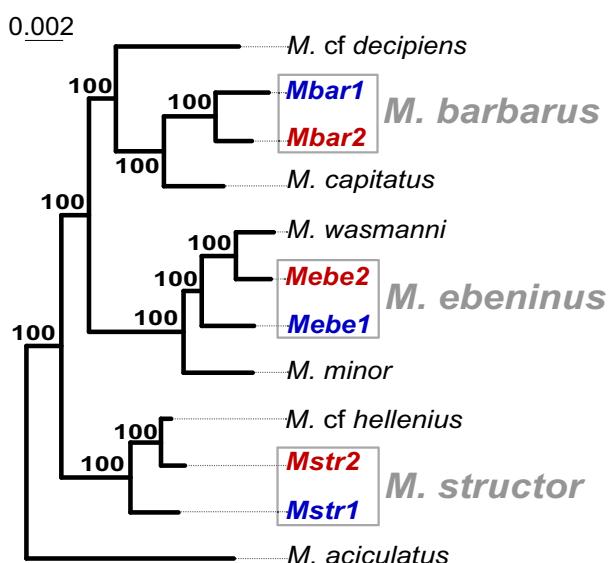


Fig. 4 Phylogenetic relationships among six *Messor* species and three pairs of dependent lineages (*Mbar1/Mbar2*, *Mstr1/Mstr2* and *Mebe1/Mebe2*) coexisting in species featuring social hybridogenesis (respectively, *M. barbarus*, *M. structor* and *M. ebeninus*). The phylogenetic tree is built from a supermatrix of 2070 orthologs genes, with sequences obtained from a consensus of all queens of a given lineage (for *M. barbarus*, *M. ebeninus* and *M. structor*) or all queens of a given species for *M. cf decipiens*, *M. capitatus*, *M. wasmanni*, *M. minor*, *M. cf hellenius* and *M. aciculatus*. [Colour figure can be viewed at wileyonlinelibrary.com]

Discussion

This study demonstrates that the reproductive system uncovered in *Pogonomyrmex* is not merely an

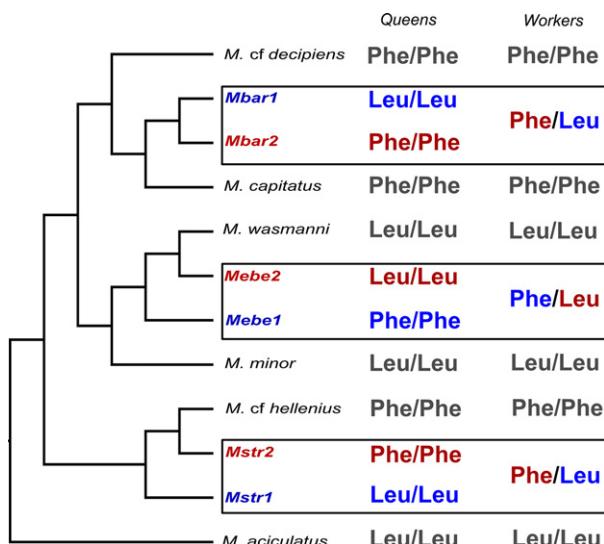


Fig. 5 Amino acid identity at position 1021 of the *catalase* gene across the *Messor* phylogeny. [Colour figure can be viewed at wileyonlinelibrary.com]

exceptional outcome of the complex and unusual hybridization history between *P. rugosus* and *P. barbatus*. A similar reproductive system has evolved independently in at least three species of the genus *Messor*, even if it is still unclear whether social hybridogenesis in two of these species (*M. ebeninus* and *M. structor*) is associated to parthenogenesis (as in *Cataglyphis*) or multiple matings with males from distinct genetic lineages (as in *Pogonomyrmex* and *M. barbarus*). Our data on *M. barbarus* reveal that this mode of reproduction can be the rule in a well-characterized morphospecies. The European harvester ant *M. barbarus* is one of the most dominant species of the western Mediterranean Basin, particularly in the Iberian Peninsula and Mediterranean France. The 13 populations that we investigated cover the whole distribution range of the species (Southern France, Iberian Peninsula and North Africa; Lebas *et al.* 2016) and they were all characterized by the presence of workers being hybrids between the two same distinct genetic lineages. This suggests the long-term persistence of this reproductive mode, possibly due to evolutionary advantages such as hybrid vigour of workers (Houle 1989; Cahan *et al.* 2010).

A surprising finding of this study is that three of the nine *Messor* species investigated features social hybridogenesis. Moreover, our analyses revealed that this mode of reproduction evolved independently in the three species, suggesting that the *Messor* genus is prone for such transitions. Interestingly, *Messor* and *Pogonomyrmex* have important convergences in their ecology. Species of these two genera live in dry climates and are obligate granivores (i.e. seed predators), which is unusual in ants. We suggest two explanations for the multiple

origins of social hybridogenesis in ants with such an ecology.

The first is associated with the fact that species of these two genera live in environments favouring large synchronous mating flights. To successfully initiate a colony and produce reproductive females, queens have to mate with males of the two distinct genetic lineages. The likelihood of encountering males coming from colonies headed by queens of both lineages increases when many colonies participate in massive and synchronous mating flights instead of small mating flights spread over a long period of time. Such massive and synchronous mating flights are common after summer rainfalls (when the soil is moistened and new nests easy to excavate), which is the most typical environmental cue triggering mating flights in ants (Hölldobler & Wilson 1990). Both *Pogonomyrmex* and *Messor* ants occur mainly in Mediterranean climates where summer rainfalls are concentrated over a few days. Consequently, *Pogonomyrmex* species and *M. barbarus* are known to have highly synchronous mating flights (Hölldobler 1976; Markl *et al.* 1977; Davidson 1982; Gómez & Abril 2012), increasing the opportunity for reproductive females of these species to encounter males of both lineages during mating flights.

An alternative explanation considers the granivorous diet of *Messor* and *Pogonomyrmex*. In many ants, caste determination is influenced both by genetic and environmental factors (Schwander *et al.* 2010). Diet and more particularly larval nutrition seem to be key environmental factors for caste determination (Haydak 1943; Kapheim *et al.* 2011). In an ant species with a mixed seed/insect diet, larvae consuming higher proportions of insect/proteins compared to seeds tend to develop into queens (Smith & Suarez 2010). Strict granivory, as is typical for most harvester ants, might decrease the ability of workers to efficiently control the queen/worker ratio through larval feeding, thus paving the way for the evolution of genetic factors for caste determination and social hybridogenesis. Interestingly, another type of social hybridogenesis was found in hybrid zones between two fire ants, whereby *Solenopsis xyloni/Solenopsis geminata* hybrids develop into workers while pure *S. xyloni* offspring develop into queens (Cahan & Bradleigh Vinson 2003). While these fire ants do not consume seeds exclusively, both are classified as granivorous and seasonally store large quantities of seeds (Valone & Michael 2005; Tschingel 2006), further supporting the view that granivory might be conducive of social hybridogenesis.

The exact evolutionary history of *Messor*-dependent lineages is difficult to assess. The two lineages of a pair may have split from a single ancestral gene pool or resulted from hybridization between two differentiated

species, as discussed in several studies on the *Pogonomyrmex* case (Anderson *et al.* 2006; Schwander *et al.* 2007a). In *M. barbarus*, the two dependent lineages are sister taxa (Fig. 4), which suggests that they derived from a single ancestral gene pool or from a closely related species that is extinct or not included in our analyses. In both *M. ebeninus* and *M. structor*, one of the dependent lineages is the closest relative of a species with a normal reproductive system. This could be explained by two hypothesis: (i) the closely related species with a normal reproductive system (*M. wasmanni* and *M. cf. hellenius*) may have been derived from a dependent lineage (respectively, *Mebe2* and *Mstr2*) and reverted back to a classical reproductive system. Alternatively, (ii) the dependent lineage *Mebe2* may have arisen from ancestral hybridization between *M. ebeninus* and *M. minor* while *Mstr2* might stem from ancestral hybridization between *M. structor* and *M. cf. hellenius*.

The analysis of males produced by workers also has important implications for our understanding of the maintenance of social hybridogenesis. A major cost associated with such a reproductive system is that dependent genetic lineages evolve independently, which could lead to an inability of colonies to produce workers in case of excessive divergence (Volny & Gordon 2002; Anderson *et al.* 2008). While inability to produce workers should theoretically be prevented by purifying selection, purifying selection is known to be particularly inefficient in eusocial insects due to their small effective population sizes (Romiguier *et al.* 2014b). Our finding that hybrid workers can produce viable males and that these males are the product of recombination between the hybrid genomes of workers raises the possibility of gene flow between the genetic lineages. Rare gene flow events between dependent lineages may prevent lineages from becoming too divergent, which could greatly contribute to the long-term persistence of these unusual reproductive systems, as it has been suggested in an analogous case for *Cataglyphis hispanica* (Darras & Aron 2015). On the other hand, introgression between lineages may also break the genetic caste determination system inherent to such reproductive systems, suggesting that gene flow must be rare and limited to genomic regions not involved in caste determination. Such rare gene flow limited to some parts of the genome has been recently reported between two highly divergent cryptic species (Roux *et al.* 2013).

Interestingly, the only heterozygous protein site common to all workers produced by social hybridogenesis (in *M. barbarus*, *M. structor* and *M. ebeninus*) belongs to the gene coding for *catalase*, an important enzyme in protecting the cell from oxidative damage by reactive oxygen species (ROS). This gene has higher expression in early larval development of honeybee queens

compared to workers (Cameron *et al.* 2013), possibly to sustain the increased respiration rate of queen larvae (Melampy & Willis 1939). The same *catalase* position is always homozygous in *Messor* species with a normal reproductive system, both for workers and queens (Fig. 5), suggesting that homozygous females can develop into either workers or queens. *Catalase* is a tetramer of four polypeptide chains, raising the possibility that heterodimeric *catalase* proteins may compromise queen development in *Messor*, either because the enzyme is less efficient or has tightly co-evolved with another gene. Such an interaction between two loci has been proposed as the simplest possible model to explain genetic caste determination in *Pogonomyrmex* ants (Helms Cahan & Keller 2003). However, more work is required to determine the number and identity of the genes implicated in the strict mechanism of caste determination in social hybridogenesis in *Messor*.

In conclusion, this study reveals that social hybridogenesis is not as rare as previously thought. On the basis of our survey, as many as one-third of the species of the genus *Messor* (a widespread genus comprising 160 species) may be characterized by such a reproductive system. Our phylogenetic analyses revealed that social hybridogenesis can readily evolve several times from normal reproductive systems. Our study also suggests that there might be a link between unusual modes of reproduction and ecology, the currently known social hybridogenesis cases requiring three reproductive partners being restricted to two granivorous genera living in a dry climate. Finally, the analyses of males produced by hybrid workers demonstrate that recombination between lineages is possible, raising the possibility of rare but important gene flow between genetic lineages. This may contribute to long-term persistence of social hybridogenesis, as seems to be the case in *M. barbarus*.

Acknowledgements

We thank Nicolas Galtier and Camille Roux for useful discussions and P. Gayral for help during the sampling. We thank Tanja Schwander and Sean McGregor for comments on this manuscript. Some computations were performed at the Vital-IT (<http://www.vital-it.ch>) Center for high-performance computing (HPC) of the SIB Swiss Institute of Bioinformatics and the EPSRC-funded MidPlus HPC center. This work was supported by a Federation of European Biochemical Societies' (FEBS) long-term fellowship to J.R., the Swiss NSF, and an ERC Advanced Grant. SHY was supported by Swiss NSF to Prof. Michel Chapuisat (grant number: 31003A-146641).

References

- Anderson KE, Gadau J, Mott BM *et al.* (2006) Distribution and evolution of genetic caste determination in *Pogonomyrmex* seed-harvester ants. *Ecology*, **87**, 2171–2184.

- Anderson KE, Linksvayer TA, Smith CR (2008) The causes and consequences of genetic caste determination in ants (Hymenoptera: Formicidae). *Myrmecological News/Oesterreichische Gesellschaft fur Entomofaunistik*, **11**, 119–132.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
- Cahan SH, Bradleigh Vinson S (2003) Reproductive division of labor between hybrid and nonhybrid offspring in a fire ant hybrid zone. *Evolution; International Journal of Organic Evolution*, **57**, 1562–1570.
- Cahan SH, Daly AB, Tanja S, Arthur Woods H (2010) Genetic caste determination does not impose growth rate costs in *Pogonomyrmex* harvester ants. *Functional Ecology*, **24**, 301–309.
- Cameron RC, Duncan EJ, Dearden PK (2013) Biased gene expression in early honeybee larval development. *BMC Genomics*, **14**, 903.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
- Curry MM, Wheeler DE, Yang K, Anderson KE (2009) The potential for gene Ffow in a dependent lineage system of a harvester ant: fair meiosis in the F1 generation. *The Journal of Heredity*, **101**, 378–384.
- Darras H, Aron S (2015) Introgression of mitochondrial DNA among lineages in a hybridogenic ant. *Biology Letters*, **11**, 20140971.
- Davidson DW (1982) Sexual selection in harvester ants (Hymenoptera: Formicidae: *Pogonomyrmex*). *Behavioral Ecology and Sociobiology*, **10**, 245–250.
- Fournier D, Estoup A, Orivel J et al. (2005) Clonal reproduction by males and females in the little fire ant. *Nature*, **435**, 1230–1234.
- Fournier D, Hellemans S, Hanus R, Roisin Y (2016) Facultative asexual reproduction and genetic diversity of populations in the humivorous termite *Cavitermes tuberosus*. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **283**, pii: 20160196. doi: 10.1098/rspb.2016.0196.
- Gayral P, Philippe G, Lucy W et al. (2011) Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. *Molecular Ecology Resources*, **11**, 650–661.
- Gómez C, Abril S (2012) Nuptial flights of the seed-harvester ant *Messor barbarus* (LINNAEUS, 1767) (Hymenoptera: Formicidae) in the Iberian Peninsula: synchrony, spatial scale and weather conditions. *Myrmecological News/Oesterreichische Gesellschaft fur Entomofaunistik*, **16**, 25–29.
- Gordon DM, Anna P, De Bortoli N, Ingram KK (2013) Does an ecological advantage produce the asymmetric lineage ratio in a harvester ant population? *Oecologia*, **173**, 849–857.
- Haydak MH (1943) Larval food and development of castes in the honeybee. *Journal of Economic Entomology*, **36**, 778–792.
- Helms Cahan S, Keller L (2003) Complex hybrid origin of genetic caste determination in harvester ants. *Nature*, **424**, 306–309.
- Helms Cahan S, Parker JD, Rissing SW et al. (2002) Extreme genetic differences between queens and workers in hybridizing *Pogonomyrmex* harvester ants. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **269**, 1871–1877.
- Hölldobler B (1976) The behavioral ecology of mating in harvester ants (Hymenoptera: Formicidae: *Pogonomyrmex*). *Behavioral Ecology and Sociobiology*, **1**, 405–423.
- Hölldobler B, Wilson EO (1990) *The Ants*. Springer, Berlin.
- Houle D (1989) Allozyme-associated heterosis in *Drosophila melanogaster*. *Genetics*, **123**, 789–801.
- Julian GE, Fewell JH, Gadau J, Johnson RA, Larrabee D (2002) Genetic determination of the queen caste in an ant hybrid zone. *Proceedings of the National Academy of Sciences*, **99**, 8157–8160.
- Kapheim KM, Bernal SP, Smith AR, Peter N, Wcislo WT (2011) Support for maternal manipulation of developmental nutrition in a facultatively eusocial bee, *Megalopta genalis* (Halictidae). *Behavioral Ecology and Sociobiology*, **65**, 1179–1190.
- Kobayashi K, Kazuya K, Eisuke H, Kyohsuke O (2008) Clonal reproduction by males of the ant *Vollenhovia emeryi* (Wheeler). *Entomological Science*, **11**, 167–172.
- Lebas C, Galkowski C, Blatrix R, Patrice W (2016) *Fourmis d'Europe Occidentale*. Delachaux et Niestlé, Paris.
- Leniaud L, Darras H, Boulay R, Aron S (2012) Social hybridogenesis in the clonal ant *Cataglyphis hispanica*. *Current Biology*, **22**, 1188–1193.
- Li L (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, **13**, 2178–2189.
- Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **26**, 589–595.
- Linksvayer TA, Wade MJ, Gordon DM (2006) Genetic caste determination in harvester ants: possible origin and maintenance by cyto-nuclear epistasis. *Ecology*, **87**, 2185–2193.
- Markl H, Hölldobler B, Hölldobler T (1977) Mating behavior and sound production in harvester ants (*Pogonomyrmex*, Formicidae). *Insectes Sociaux*, **24**, 191–212.
- Matsuura K, Vargo EL, Kawatsu K et al. (2009) Queen succession through asexual reproduction in termites. *Science*, **323**, 1687.
- Melampy RM, Willis ER (1939) Respiratory metabolism during larval and pupal development of the female honeybee (*Apis mellifica* L.). *Physiological Zoology*, **12**, 302–311.
- Mott BM, Gadau J, Anderson KE (2015) Phylogeography of *Pogonomyrmex barbatus* and *P. rugosus* harvester ants with genetic and environmental caste determination. *Ecology and Evolution*, **5**, 2798–2826.
- Pearcy M, Aron S, Doums C, Keller L (2004) Conditional use of sex and parthenogenesis for worker and queen production in ants. *Science*, **306**, 1780–1783.
- Pearcy M, Goodisman MAD, Keller L (2011) Sib mating without inbreeding in the longhorn crazy ant. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **278**, 2677–2681.
- Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, **197**, 573–589.
- Ranwez V, Vincent R, Sébastien H, Frédéric D, Douzery EJP (2011) MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS ONE*, **6**, e22594.
- Romiguier J, Gayral P, Ballenghien M et al. (2014a) Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, **515**, 261–263.
- Romiguier J, Lourenco J, Gayral P et al. (2014b) Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *Journal of Evolutionary Biology*, **27**, 593–603.
- Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species barrier: genomic hotspots of introgression

- between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, **30**, 1574–1587.
- Schwander T, Cahan SH, Keller L (2006) Genetic caste determination in *Pogonomyrmex* harvester ants imposes costs during colony founding. *Journal of Evolutionary Biology*, **19**, 402–409.
- Schwander T, Cahan SH, Keller L (2007a) Characterization and distribution of *Pogonomyrmex* harvester ant lineages with genetic caste determination. *Molecular Ecology*, **16**, 367–387.
- Schwander T, Keller L, Cahan SH (2007b) Two alternate mechanisms contribute to the persistence of interdependent lineages in *Pogonomyrmex* harvester ants. *Molecular Ecology*, **16**, 3533–3543.
- Schwander T, Lo N, Beekman M, Oldroyd BP, Keller L (2010) Nature versus nurture in social insect caste differentiation. *Trends in Ecology & Evolution*, **25**, 275–282.
- Sirviö A, Anu S, Pekka P et al. (2010) Origin and evolution of the dependent lineages in the genetic caste determination system of *Pogonomyrmex* ants. *Evolution; International Journal of Organic Evolution*, **65**, 869–884.
- Smith CR, Suarez AV (2010) The trophic ecology of castes in harvester ant colonies. *Functional Ecology*, **24**, 122–130.
- Smith CR, Toth AL, Suarez AV, Robinson GE (2008) Genetic and genomic analyses of the division of labour in insect societies. *Nature Reviews. Genetics*, **9**, 735–748.
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.
- Suni SS, Gordon DM (2010) Fine-scale genetic structure and dispersal distance in the harvester ant *Pogonomyrmex barbatus*. *Heredity*, **104**, 168–173.
- Tschinkel WR (2006) *The Fire Ants*. Harvard University Press, Cambridge, Massachusetts.
- Valone TJ, Michael K (2005) Interactions between granivorous and omnivorous ants in a desert grassland: results from a long-term experiment. *Ecological Entomology*, **30**, 116–121.
- Volny VP, Gordon DM (2002) Genetic basis for queen-worker dimorphism in a social insect. *Proceedings of the National Academy of Sciences*, **99**, 6108–6111.
- Volny VP, Greene MJ, Gordon DM (2006) Brood production and lineage discrimination in the red harvester ant (*Pogonomyrmex barbatus*). *Ecology*, **87**, 2194–2200.
- Weir BS, Clark Cockerham C (1984) Estimating F-Statistics for the analysis of population structure. *Evolution; International Journal of Organic Evolution*, **38**, 1358–1370.
- Yamauchi A, Yamamura N (2006) Persistence conditions of symmetric social hybridogenesis in haplo-diploid hymenoptera. *Journal of Theoretical Biology*, **240**, 24–31.

Data accessibility

Raw ILLUMINA reads are freely available from the SEQUENCE READ ARCHIVE (SRA) database (<http://www.ncbi.nlm.nih.gov/sra>). See Table S1 (Supporting information) for each accession ID. Sequence alignments (*hsc70-4*, *cox1* and *catalase*) are available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.6gr08>.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 List of *Messor* individuals with transcriptomic data sequenced. Species, caste, queen status (when relevant), colony/population/country of origin and ncbi SRA accession are provided.

Table S2 List of *M. barbarus* with the *hsc70-4* gene sequenced. Caste, population/colony of origin, sampling date, lineage, genotype at each diagnostic position of *hsc70-4* and presence/absence of the sequencing of the mitochondrial gene *cox1* are provided.

Table S3 Population genetic statistics from transcriptomic data. Each sheet corresponds to one different species. For each ORF (noted contigs), the number of bi-allelic/tri-allelic/quadr-allelic SNPs, the number of sequences (two per individuals), the number of complete sites, the GC3%, the non-synonymous nucleotidic diversity (piN), the synonymous nucleotidic diversity (piS), Fis values (with and without Weir-Cockerman correction) and individual heterozygosity are provided. Heterozygosity of each individuals is noted as H (*individual-id*), with *individual-id* corresponding to the ID field of Table S1.

Fig. S1 (A) Workers of *M. structor* have higher genome-wide heterozygosity levels (based on 90951 SNPs) than queens (all pairwise comparisons significant, $P < 0.0001$, Wilcoxon tests). (B) Population ancestry proportions estimated by *fastStructure* in 8 individuals. Each bar corresponds to individuals from (A). Bars are named according to the population of origin the individual. Queens are pure *Mstr1* (blue) or *Mstr2* (red) lineages, workers are *Mstr1/Mstr2* hybrids.

J.R. and L.K. conceived the project. J.R., A.F. and S.H.Y. performed sampling and laboratory work. J.R. developed the data analysis pipeline. J.R. and A.F. analysed the data. J.R. and L.K. wrote the paper.

LETTER

doi:10.1002/evl3.253



Open Access

Hybridization enables the fixation of selfish queen genotypes in eusocial colonies

Arthur Weyna,¹ Jonathan Romiguier,^{1,2,*} and Charles Mullon^{3,4,*}¹*Institut des Sciences de l'Evolution (UMR 5554), University of Montpellier, CNRS, Montpellier 34000, France*²*E-mail: jonathan.romiguier@umontpellier.fr*³*Department of Ecology and Evolution, University of Lausanne, Lausanne 1015, Switzerland*⁴*E-mail: charles.mullon@unil.ch*

Received June 9, 2021

Accepted August 9, 2021

A eusocial colony typically consists of two main castes: queens that reproduce and sterile workers that help them. This division of labor, however, is vulnerable to genetic elements that favor the development of their carriers into queens. Several factors, such as intracolonial relatedness, can modulate the spread of such caste-biasing genotypes. Here we investigate the effects of a notable yet understudied ecological setting: where larvae produced by hybridization develop into sterile workers. Using mathematical modeling, we show that the coevolution of hybridization with caste determination readily triggers an evolutionary arms race between nonhybrid larvae that increasingly develop into queens, and queens that increasingly hybridize to produce workers. Even where hybridization reduces worker function and colony fitness, this race can lead to the loss of developmental plasticity and to genetically hard-wired caste determination. Overall, our results may help understand the repeated evolution toward remarkable reproductive systems (e.g., social hybridogenesis) observed in several ant species.

KEY WORDS: Ant, caste determination, eusociality, genetic conflicts, hybridization, Hymenoptera, parasitism, reproductive system, social hybridogenesis.

Eusociality is characterized by a striking division of reproductive labor between two castes: queens and workers (Crespi and Yanega 1995). Queens monopolize reproduction, while typically sterile workers specialize on other colony tasks such as foraging and tending to the brood. The sterility of workers initially seemed so inconsistent with natural selection that Darwin referred to eusociality as his “one special difficulty” (Darwin 1859, ch. 7). This apparent paradox was resolved in the 1960s with Hamilton’s theory of kin selection (Hamilton 1964). Hamilton demonstrated that natural selection can favor eusociality when workers preferentially help relatives (who can transmit the same genetic material). In addition to laying the theoretical basis for the evolution of eusociality, Hamilton’s work led to the insight that caste determination should be plastic to allow identical gene copies to be in workers and in the queen they help (Seger 1981). In line with this notion, the developmental fate of female larvae in many eusocial insects depends on environmental factors (Trible and Kronauer

2017), such as food quantity and quality (Brian 1956, 1973), temperature and seasonality (Brian 1974; Schwander et al. 2008) or signals emitted by adults of the colony (Penick and Liebig 2012; Libbrecht et al. 2013). Probably the most iconic example of such plasticity is found in honeybees where queens arise only from larvae reared in royal cells and fed with royal jelly. For long, this and many other empirical findings strengthened the idea that caste determination is under strict environmental control and largely free from genetic effects.

More recently, however, substantial genetic variation for caste determination has been described across a number of eusocial species (Winter and Buschinger 1986; Moritz et al. 2005; Hartfelder et al. 2006; Linksvayer 2006; Schwander and Keller 2008; Smith et al. 2008; Frohschammer and Heinze 2009; Schwander et al. 2010). This variation is thought to derive from caste-biasing genotypes that bias the development of their carrier toward a particular caste (Moritz et al. 2005; Hughes and Boomsma 2008). Those genotypes that favor larval development toward the reproductive caste have sometimes been

*These authors share senior authorship.

referred to as “royal cheats” as they cause the individuals that carry them to increase their own direct reproduction at the expense of other colony members (e.g., Anderson et al. 2008; Hughes and Boomsma 2008). The segregation of such royal cheats should depend on a balance between: (1) direct benefits from increased representation in the reproductive caste; and (2) indirect costs due to reduced worker production and colony productivity (Hamilton 1964). As highlighted by abundant theory, several factors can influence these benefits and costs and thus tip the balance for or against the evolution of royal cheats. For instance, low relatedness between larvae due to polyandry (when queens mate with multiple males) or polygyny (when colonies have multiple queens) increases competition between genetic lineages within colonies and thereby favors royal cheating (e.g., Reuter and Keller 2001). Conversely, selection against cheats is bolstered by low dispersal abilities and high within-group relatedness (e.g., Hamilton 1964; Lehmann et al. 2008; Boomsma 2009), bivoltinism and asymmetrical sex-ratio (e.g., Trivers and Hare 1976; Seger 1983; Alpedrinha et al. 2014; González-Forero 2015; Quiñones and Pen 2017), coercion (i.e., policing; Wenseleers et al. 2004; Dobata 2012), queen longevity and competition between queens (e.g., Queller 1994; Bourke and Chan 1999; Avila and Fromhage 2015), or where workers reproduce following queen death (Field and Toyozumi 2020).

One intriguing factor that has been proposed to influence the cost of royal cheating is sperm parasitism, a behavior consisting in queens using the sperm of another species or lineage to produce hybrid workers (Linksvayer 2006; Anderson et al. 2008). Both morphological and genetic data suggest that this behavior is common in many ant species (e.g., in multiple *Temnothorax* populations, the majority of queens were found to produce some hybrid workers; Douwes and Stille 1991; Umphrey 2006 and Feldhaar et al. 2008 for reviews). In these species, sperm parasitism results in hybrid larvae that rarely, if ever, develop as fertile queens and rather become sterile workers (presumably due to genetic incompatibilities between parental lineages; Feldhaar et al. 2008; Trible and Kronauer 2017). Such hybrids should therefore be impervious to genetic caste-biasing effects and thus provide a reliable source of workers. In principle, this alternative supply of workers may reduce the indirect cost of royal cheats and hence favor their evolution (Anderson et al. 2008). But beyond these broad-brush predictions, the effect of sperm parasitism on the segregation of royal cheats remains poorly understood.

Here, we develop a mathematical model to explore the evolution of genetic caste determination via royal cheats when queens can hybridize to produce workers. In particular, we assess the effects of key factors on the evolutionary dynamics of caste determination, such as polyandry and queen parthenogenesis (when queens have the ability to produce daughters asexually), as well as their interactions with potential costs and benefits

of hybridization, for instance, owing to hybrid incompatibilities or hybrid vigor.

The Model

We consider a large population of annual eusocial haplodiploids with the following life-cycle (Fig. 1). First, virgin queens mate with a fixed number $m \in \{1, 2, \dots\}$ of males. Each of these mates can either be an allo- (with probability η) or a con-specific male (with complementary probability $1 - \eta$). Once mated, queens found monogynous colonies (i.e., one queen per colony) and lay a large number of eggs. A proportion f of these eggs are diploid (and develop into females) and $(1 - f)$ are haploid (and develop into males). Assuming random egg fertilization, a queen therefore produces on average $f\eta$ hybrid and $f(1 - \eta)$ nonhybrid females. We assume that a hybrid female can only develop as a worker, while a nonhybrid female can either develop as a worker (with probability ω) or as a queen (with complementary probability $1 - \omega$). Overall, a colony thus consists of $f\eta$ hybrid and $f(1 - \eta)\omega$ nonhybrid sterile workers, as well as $f(1 - \eta)(1 - \omega)$ virgin queens and $(1 - f)$ males that are available for reproduction at the next generation.

If only virgin queens and males can reproduce, their reproductive success depends on the workforce of their colony of origin. Specifically, we assume that the probability that a sexual reaches the mating pool increases linearly with the total number of workers in the colony, combining hybrid and nonhybrid workers (we show later that our results do not change qualitatively when the increase is nonlinear). We nonetheless allow for differential contribution to the workload between hybrid and nonhybrid workers, with the contribution of hybrid workers weighted by a parameter $e \geq 0$ (so that the effective workforce of a colony is $ef\eta + f(1 - \eta)\omega$). When $e = 1$, hybrid workers have the same working efficiency as nonhybrid workers. By contrast, when $e < 1$, hybrid workers are less efficient, for instance, due to outbreeding depression. This can also reflect other potential costs associated with hybridization, such as the production of sterile or nonviable hybrid queens (Feldhaar et al. 2008). Conversely, when $e > 1$, hybrid workers outperform regular workers, due, for example, to hybrid vigor (Umphrey 2006).

Results

HYBRIDIZATION AND SPERM PARASITISM, EVEN COSTLY, CAN LEAD TO THE FIXATION OF ROYAL CHEATS AND THE COMPLETE LOSS OF INTRASPECIFIC WORKERS

We first investigate the evolution of caste determination by allowing the probability ω that a larva develops as a worker to

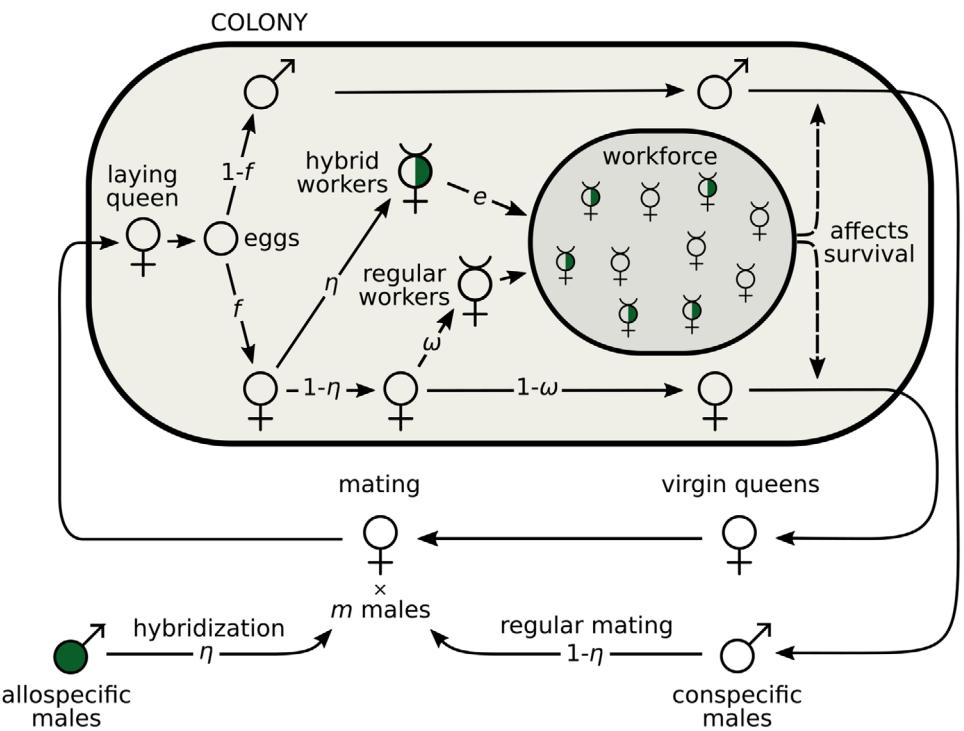


Figure 1. The life cycle of an annual eusocial with hybridization and sperm parasitism. At each generation, the life-cycle begins with virgin queens mating with m males, each of which has a probability η to be allo-specific and $1 - \eta$ to be conspecific. After mating, a queen founds a colony and starts producing eggs. Hybrid female eggs (with allo-specific paternal origin) all develop into workers. Regular female eggs (with conspecific paternal origin) develop into workers with probability ω and into queens otherwise. The variable η thus captures the tendency of queens to hybridize and parasitize sperm, while ω controls caste determination.

vary. We assume that this probability is under individual genetic control (i.e., the future caste of a female larva depends only on its own genotype) and that it evolves via random mutations with weak additive phenotypic effects (Appendix A for details on our methods). Mutational effects are unbiased so a new mutation is equally likely to increase or decrease the tendency ω of becoming a worker. Those mutations that decrease ω can be considered as more selfish as they increase the likelihood that their carriers develop into queens at the expense of other individuals of the same colony. Following the terminology of Hughes and Boomsma (2008), we thus refer to mutations decreasing ω as royal cheats. As a baseline, we consider the case where queens mate with a large number of males (i.e., $m \rightarrow \infty$) and where hybridization is fixed at a given level (e.g., η is the proportion of allo-specific males in the pool of mates from which females choose randomly).

Our analyses (Appendix B.1.1) reveal that the probability for a larva to develop as a worker evolves toward a unique and stable equilibrium,

$$\omega^* = \frac{1}{3} - e \frac{2\eta}{3(1-\eta)}. \quad (1)$$

To interpret this equation (1), consider first the case where hybridization is costless ($e = 1$). Equation (1) then tells that in the absence of hybridization ($\eta = 0$), a larva will develop into a worker with a probability of $1/3$ at equilibrium (in line with previous models that ignore hybridization, e.g., Reuter and Keller 2001, Appendix B.1.4 for connection). But as hybridization increases ($\eta > 0$), royal cheating is increasingly favored and larvae become increasingly likely to develop as queens rather than workers (i.e., $\omega^* < 1/3$, Fig. 2A). In fact past a threshold of hybridization ($\eta \geq 1/3$), the population evolves toward a complete loss of nonhybrid workers via the fixation of increasingly caste-biasing royal cheats alleles ($\omega \rightarrow 0$). In this case, nonhybrid females eventually all develop into queens that rely on sperm parasitism to produce workers.

Equation (1) also shows that the performance of hybrid workers relative to nonhybrids, e , modulates the effect of hybridization on the evolution of caste determination (Fig. 2B). As a result, royal cheating and worker-loss evolution are facilitated when hybrids outperform regular workers ($e > 1$) but hindered otherwise ($e < 1$). Nevertheless, even when hybridization is extremely costly ($0 < e \ll 1$), complete worker-loss can evolve (Fig. 2C).

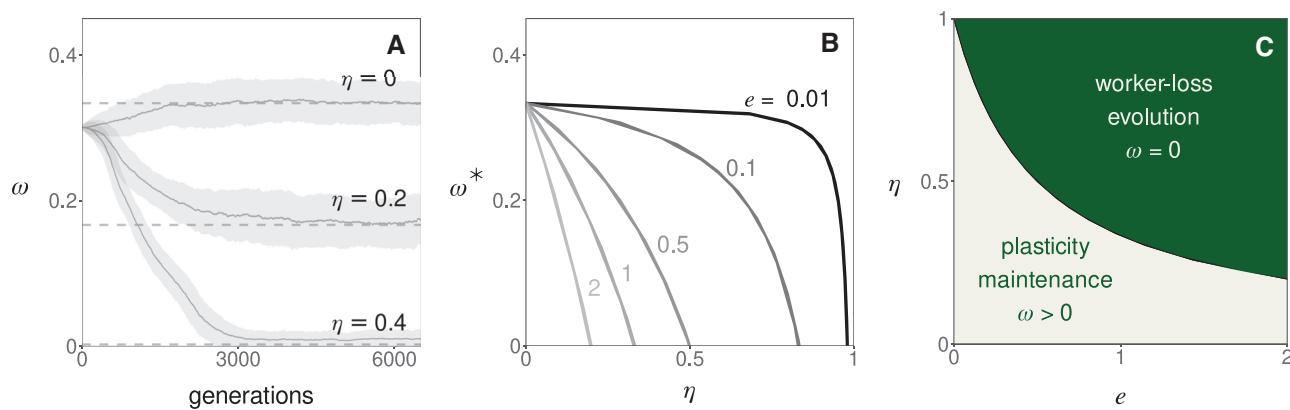


Figure 2. The fixation of royal cheats and evolution of intraspecific worker-loss. (A) Evolution of the probability ω that a female larva develops into a worker in a simulated population when queens mate with a large number of males (polyandry, $m \rightarrow \infty$) and the proportion of allospecific males η is fixed (top $\eta = 0$; middle $\eta = 0.2$, bottom $\eta = 0.4$; other parameters: $e = 1$, Appendix A.3 for details on simulations). Plain lines (and surrounding gray areas) show the population average ω (and its standard deviation). Dashed lines show the predicted equilibrium (from eq. 1). (B) Equilibrium of ω as a function of hybridization η and the efficiency of hybrid workers e (from eq. 1). (C) Parameter combinations leading to the evolution of complete worker-loss (i.e., $\omega \rightarrow 0$, in green, corresponding to $\eta \geq 1/(1+2e)$, which is found by substituting eq. 1 into $\omega^* \leq 0$).

WORKER-LOSS READILY EMERGES FROM THE COEVOLUTION OF GENETIC CASTE DETERMINATION AND SPERM PARASITISM, DRIVEN BY INTRACOLONIAL CONFLICT

The above analysis indicates that intraspecific worker-loss can evolve when queens have a sufficiently high tendency to hybridize. This raises the question of whether such tendency is also subject to selection. To answer this question, we allow the probability η that a queen's mate is allospecific to coevolve with caste determination (ω). We assume that this probability η is under individual queen control (i.e., it depends only on a queen's genotype) and like caste determination, evolves via rare mutations with weak additive phenotypic effects (Appendix A for details).

We find that depending on the efficiency e of hybrid workers, the coupled evolutionary dynamics of hybridization η and caste determination ω lead to an evolutionary arms race with one of two contrasted outcomes (Appendix B.1.2 for analysis). When e is small ($e \leq 1/4$, Fig. 3A gray region), the population evolves hybridization avoidance ($\eta \rightarrow 0$) while the probability ω to develop as a worker stabilizes for its baseline equilibrium ($\omega^* = 1/3$, Fig. 3B). By contrast, when hybrid workers are at least half as efficient as regular workers ($e \geq 1/2$, Fig. 3A, dark green region), intraspecific worker-loss evolves ($\omega \rightarrow 0$) and hybridization stabilizes at an intermediate equilibrium ($\eta^* = 2/3$, Fig. 3D). When hybrid worker efficiency is intermediate ($1/4 < e < 1/2$, Fig. 3A, light green region), the population evolves either hybridization avoidance or intraspecific worker-loss depending on initial conditions (Fig. 3C), with worker-loss favored by high initial tendency η of queens to hybridize. In sum, provided four hybrid workers are at least as good as one regular worker ($e > 1/4$),

the coevolution of genetic caste determination and hybridization can lead to worker-loss in our model.

To better understand the forces at play in the emergence of worker-loss, we further used a kin-selection approach to decompose the invasion fitness of mutant alleles into the sum of: (1) their direct fitness effects on the reproductive success of the individuals that express them; and (2) of their indirect fitness effects on other related individuals that can also transmit them (Taylor and Frank 1996, Appendix B.1.3 for details). Starting with a population at the baseline equilibrium in absence of hybridization ($\omega = 1/3$, $\eta = 0$), we tracked these different fitness effects along a typical evolutionary trajectory that leads to worker-loss (black arrow heads, Fig. 3D) for alleles that influence the tendency of a larva to develop as a worker (Fig. 3E) and of a queen to hybridize (Fig. 3F).

Our kin selection analysis reveals that alleles that increase hybridization in queens are selected because they allow queens to increase the number of sexuals produced by their colony (especially via males, blue curve, Fig. 3F). This is because the baseline tendency ω to develop as a worker that evolves is optimal from the point of view of a gene in a larva, but sub-optimal from the point of view of a gene in a queen who would benefit from a larger workforce. Hybridization by queens evolves to rectify this and align colony composition with the interests of the queen. Simultaneously, as queens evolve greater hybridization and augment their workforce with hybrids, genes in nonhybrid larva have an increasing interest for their carriers to develop as queens rather than workers (Fig. 3E). These two selective processes via queens and larvae fuel one another in an evolutionary arms race whose endpoint is complete intraspecific

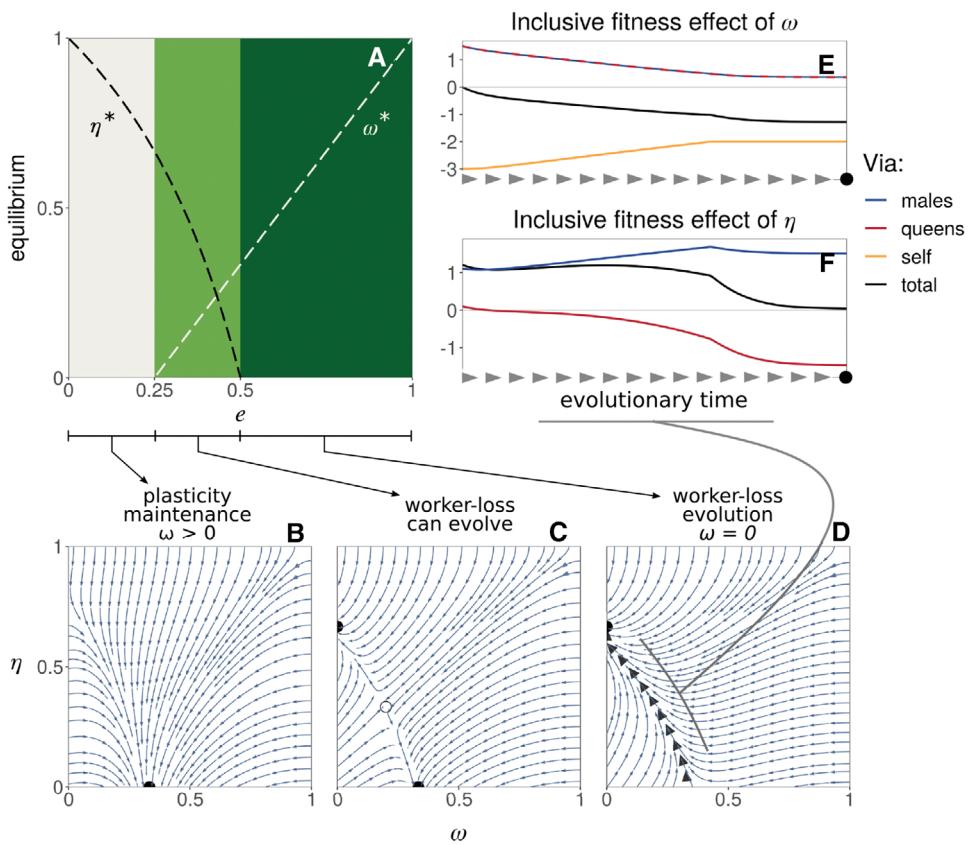


Figure 3. The coevolution of caste determination and sperm parasitism. (A) Evolutionary equilibria (for η in black and ω in white) as a function of hybrid worker efficiency e (eq. B6 in Appendix B.1.2 for details). These equilibria, however, are evolutionary repellors (eq. B7 in Appendix B.1.2). As a result, three types of coevolutionary dynamics are possible depending on e as illustrated in panels (B)–(D) (from eq. B5). These panels show examples of phenotypic trajectories when worker-loss: Panel (B) never evolves ($e = 0.1$); Panel (C) can evolve depending on initial conditions ($e = 0.4$); Panel (D) always evolves ($e = 0.7$). Black filled circles indicate the two evolutionary endpoints: hybridization avoidance with developmental plasticity ($\omega = 1/3$ and $\eta = 0$ in B and C) or worker-loss with hybridization ($\omega = 0$ and $\eta = 2/3$ in C and D). Empty circle in (C) shows the internal unstable equilibrium (eq. B6). Thick grey arrow heads in (D) represent the trajectory of a population starting from $\omega = 1/3$ and $\eta = 0$ and evolving to worker-loss. (E) Fitness effects of caste determination ω in a mutant larva via itself (in orange), related queens (red), and related males (blue) along the trajectory leading to worker-loss shown in panel (D) (total selection in black, Appendix B.1.3 for derivation). We see that negative fitness effects via self (orange line) lead to a total selection effect that is negative (black line). This indicates that mutant larvae with increasingly small values of ω are selected because these values increase larvae's direct fitness (by increasing the probability that they develop into queens). (F) Fitness effects of hybridization η in a mutant queen, via its sons (blue) and daughter queens (red) along the trajectory leading to worker-loss shown in panel (D) (total selection in black). Positive total selection (in black) is mostly due to an increase of fitness via males (in blue). This says that mutant queens with increasingly large values of η are selected because this increases their reproduction, especially via males.

worker-loss. Our decomposition of fitness effects thus shows that the loss of nonhybrid workers evolves in our model due to within-colony conflicts over colony composition. In fact, our results suggests that worker-loss emerges because hybridization allows queens to control the production of workers in their colony, while nonhybrid larvae lose their tendency to develop as workers to promote their own reproduction via the fixation of royal cheats.

WORKER-LOSS IS IMPAIRED BY LOW POLYANDRY BUT FACILITATED BY ASEXUAL REPRODUCTION

So far, we have assumed that queens mate with a large, effectively infinite, number of males. By increasing relatedness within the brood, low polyandry ($2 \leq m \ll \infty$), and monandry ($m = 1$) mediate within-colony conflicts and therefore should be relevant to the evolutionary arms race leading to worker-loss (Anderson et al. 2008; Schwander et al. 2010). To test this, we

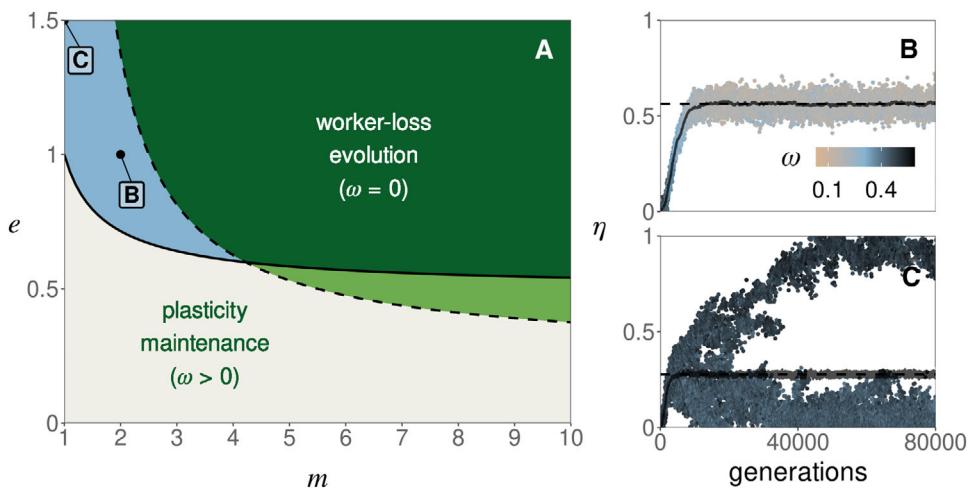


Figure 4. The effects of monandry and low polyandry. (A) Outcome of selection as a function of mate number m and hybrid worker efficiency e . Over the dashed line, worker-loss is a stable equilibrium (i.e., a population with traits $\omega = 0$ and $\eta = 2/3$ cannot be invaded, eq. B16 in Appendix B.2.1). Over the plain line, hybridization can invade when rare (i.e., $\eta = 0$ is unstable, eq. B18 in Appendix B.2.1). Below both lines (gray region), plasticity in caste determination is maintained (as in Fig. 3B). Over both lines (dark green region), hybridization and worker-loss evolve (as in Fig. 3D). In the light green region, worker-loss evolves for some initial conditions (as in Fig. 3C). In the blue region, there exists an internal attractor equilibrium (i.e., the population converges toward a phenotype $0 < \eta^* < 1$ and $0 < \omega^* < 1$) that is either uninvadable (for $2 \leq m \leq 4$, see, e.g., panel B) or invadable leading to polymorphism (for $m = 1$, see, e.g., panel C). (B) Evolution toward an uninvadable phenotype in a simulated population (when $e = 1$ and $m = 2$). Each dot represents the value of η of one of 20 haplotypes randomly sampled every 100 generation in a simulated population of 10,000 queens (Appendix A.3 for details on simulations). The color of each dot gives the value of ω of the associated haplotype (legend). The horizontal dashed line represents the predicted equilibrium (from Fig. S1). The gray line represents the mean value of η across the simulation. (C) Evolution toward an invadable phenotype and the emergence of polymorphism in a simulated population (when $e = 1.5$ and $m = 1$, other parameters and figure legend: same as B).

investigated the effect of mate number m on the coevolution of ω and η (Appendix B.2.1 for details).

We find that as the number m of mates decreases, the conditions for intraspecific worker-loss emergence become more restrictive. Specifically, the threshold of hybrid worker efficiency e above which worker-loss always evolves increases as polyandry decreases (as $m \rightarrow 1$, Fig. 4A, dark green region). In addition, when the number of mates is low ($m \leq 4$), evolutionary dynamics do not necessarily lead to either complete worker-loss or hybridization avoidance. For intermediate values of e (Fig. 4A, blue region) the population actually converges to an intermediate state where queens partially hybridize ($0 < \eta^* < 1$) and larvae retain developmental plasticity ($0 < \omega^* < 1$, Fig. 4B, Appendix B.2.1 and Fig. S1 for analysis). Under monandry ($m = 1$) the evolution toward such intermediate state always happens when hybrid workers outperform regular workers ($e > 1$, Fig. 4A, blue region).

In the special case of monandry and overperforming hybrid workers ($m = 1$ and $e > 1$), our mathematical analysis further shows that partial hybridization and larval plasticity is not evolutionary stable (Appendix B.2.1, Figs. S1 and S2). Rather, the population experiences disruptive selection that should favor the emergence of polymorphism. To test this, we performed

individual-based simulations under conditions predicted to lead to polymorphism (Fig. 4C). These show the emergence and long-term coexistence of two types of queens: one that hybridizes with low probability (and reproduces via both males and queens); and another that mates almost exclusively with allospecific males and thus reproduces mostly via males (because $m = 1$, these queens only produce hybrid workers and males). Beyond this special case, the evolution of worker-loss is impeded by low polyandry and impossible under monandry in our model. This is because with a low number of mates, a queen runs the risk of being fertilized by only one type of male. Under complete worker-loss (when the population is fixed for $\omega = 0$), a queen mated to only conspecific males produces only larvae destined to be queens but no workers to ensure their survival and thus has zero fitness.

Our finding that monandry inhibits the emergence of worker-loss contrasts with the observation that several ant species, notably of the genus *Cataglyphis*, lack nonhybrid workers and rely on sperm parasitism for workers in spite of being mostly monandrous (Kuhn et al. 2020). One potential mechanism that could have allowed such evolution is thelytokous parthenogenetic reproduction by queens, whereby queens can produce daughters clonally. This reproduction mode, which is common in eusocial Hymenoptera (Rabeling and Kronauer 2013) and in

particular in *Cataglyphis* (Kuhn et al. 2020), could allow queens fertilized exclusively by allospecific males to nevertheless produce queens via parthenogenesis. To investigate how thelytokous parthenogenesis influences the evolution of caste determination, we extend our model so that a fraction c of the female progeny of queens is produced parthenogenetically (Appendix B.2.2 for details). We assume that larvae produced in such a way are equivalent to nonhybrid larvae: they develop into workers with a probability ω determined by their own genotype (which in this case is the same as their mother's genotype) and if they develop into workers, they have the same working efficiency as nonhybrid workers (i.e., there is no direct cost or benefit to parthenogenesis).

The coevolutionary dynamics of caste determination and hybridization with parthenogenesis are in general too complicated to be tractable. We could nonetheless gain insights into worker-loss evolution by performing an invasion analysis, asking (1) when is worker-loss ($\omega = 0$) evolutionary stable (so that a population where intraspecific workers have been lost cannot be invaded by a genetic mutant with developmental plasticity)? And (2) when can hybridization evolve when absent in the population (i.e., when is $\eta = 0$ evolutionary unstable)? When these two conditions are met, evolution will tend to favor the emergence and maintenance of worker-loss (e.g., as in Fig. 3D). We thus studied when conditions (1) and (2) above are both true in terms of parthenogenesis c , as well as hybrid workers efficiency e and mate number m . This revealed that parthenogenesis has a nonmonotonic relationship with worker-loss evolution (Fig. 5A and B). As parthenogenesis increases from zero, worker-loss evolution is initially favored, especially under monandry (as expected; e.g., Fig. 5C; see eq. B26 in the Appendix for details). But past a threshold of parthenogenesis, the conditions leading to worker-loss become increasingly stringent until such evolution becomes impossible (see eq. B25 in the Appendix for details). This is because as parthenogenesis increases, the relatedness among a queen and larvae of the same colony also increases. The conflict between them, which fuels the evolution of worker-loss, therefore abates until it is no longer advantageous for a larva to preferentially develop as a queen.

We additionally computed the level of hybridization favored by selection when the population has evolved worker-loss (and this is an evolutionarily stable state). We find that hybridization increases as queens mate with fewer males and as parthenogenesis increases (Fig. 5D), so much so that selection can lead to complete hybridization ($\eta = 1$, e.g., Fig. 5C). As a result, there exists a range of intermediate values of parthenogenesis for which worker-loss evolves in association with a complete loss of intraspecific matings, that is, queens never mate with males of their own species or lineage. These males are nevertheless still

being produced in our model (as the primary sex ratio is such that $f < 1$).

Discussion

In sum, our analyses indicate that worker-loss readily evolves when queens can hybridize with a lineage of males by whom fertilization leads to the production of workers. This evolution in our model occurs through a sequence of substitutions of alleles that increasingly bias the development of their carrier toward the queen caste, that is, “royal cheats”. Hybridization, or sperm parasitism, allows royal cheats to fix in the population by providing a way for colonies to compensate for the reduced workforce. In fact, when queens are capable of recognizing genetic differences among males and when royal cheats are present in the population, selection favors hybridization by queens to regain control over caste allocation in their colony. This in turn promotes greater cheating by larvae, which favors greater hybridization by queens and so on. This evolutionary arms race, fueled by intracolonial conflicts, eventually leads to complete intraspecific worker-loss: a state where larvae have lost their developmental plasticity and develop as workers or queens depending only on whether they are the product of hybridization or not, respectively.

MODEL LIMITATIONS

Of course, our analyses are based on several idealized assumptions. In particular, we assumed that the probability for larvae to develop as workers is under complete larval genetic control. Typically the developmental fate of female larvae also depends on various environmental factors created by adult colony members, such as food quality and quantity (Brian 1956; Trible and Kronauer 2017), or mechanical (Penick and Liebig 2012) and chemical (Schwander et al. 2008; Penick et al. 2012) stimuli. The conclusions of our study apply as long as these environmental effects are held constant (or evolve more slowly than genetic caste determination). In this case, worker-loss would emerge via royal cheats that modify larval developmental reaction norm to environmental effects in such a way that their carriers are more likely to develop as queens (Hughes and Boomsma 2008; Wolf et al. 2018). We also assumed that caste determination and hybridization evolve via rare mutations with weak additive effects at a single locus. These assumptions, which are typical to adaptive dynamics and kin selection approaches, have been extensively discussed elsewhere in a general context (Frank 1998; Rousset 2004; Geritz and Gyllenberg 2005; Dercole and Rinaldi 2008). In particular, all our results extend to the case where traits are determined by many genes and/or many co-segregating alleles, provided genetic variance in the population remains small (e.g., Charlesworth 1990; Iwasa et al. 1991; Abrams et al. 1993; Mullon and Lehmann 2019). In cases where mutations have

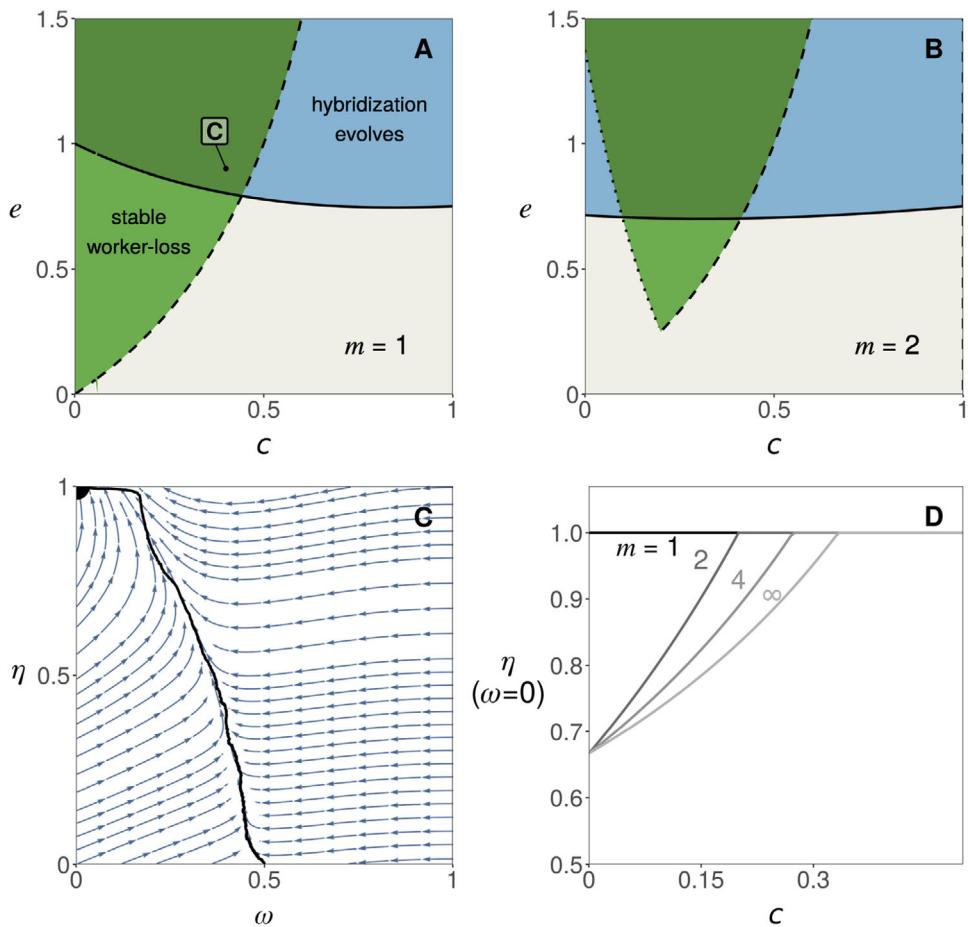


Figure 5. The influence of thelytokous parthenogenesis. (A) and (B) Invasion analysis as a function of parthenogenesis c and hybrid worker efficiency e (with $m = 1$ in A and $m = 2$ in B). In the region over the plain line, hybridization can invade when rare (i.e., $\eta = 0$ is unstable, eq. B23). In the region over the dashed line (in A) or framed by the dotted and dashed lines (in B), worker-loss is a stable equilibrium (i.e., a population at equilibrium for η and with $\omega = 0$ cannot be invaded, Appendix B.2.2, eqs. B25 and B26 for details). In the dark green region, selection thus favors both the evolution of hybridization and the maintenance of worker-loss (e.g., panel C). In the light green region, worker-loss can evolve only for some initial conditions (as in Fig. 3C). (C) Phenotypic trajectories leading to worker-loss (when $e = 0.9$, $c = 0.4$, and $m = 1$). Arrows show the direction of evolution favored by selection. Black filled circles indicate the evolutionary end-point. The black line shows the average trait values of a simulated population starting at $(\omega = 1/2, \eta = 0)$. In this example, selection leads to a state where worker-loss ($\omega = 0$) is coupled with complete hybridization ($\eta = 1$). (D) Level of hybridization η favored by selection when worker-loss has evolved ($\omega = 0$) as a function of parthenogenesis c . This shows that worker-loss is always associated to complete hybridization ($\eta = 1$) under monandry ($m = 1$) and if $c \geq (m - 1)/(3m - 1)$ under polyandry ($m > 1$) (Appendix B.2.2, eq. B24, for details).

large additive or dominance effects, we expect more complex evolutionary dynamics, such as genetic polymorphism. These dynamics can nonetheless be straightforwardly investigated with the recurrence equations we derived (eq. A4 in Appendix). However, our model cannot accommodate potential interaction effects among loci (i.e., epistasis). If a quantitative genetics analysis in *Temnothorax curvispinosus* supports that caste determination is influenced by additive effects in this species (Linksvayer 2006), only epistatic effects were found in *Pogonomyrmex rugosus* (Schwander and Keller 2008). It would therefore be relevant in the future to allow for a more complex genetic basis of caste determination, including epistasis (in particular, in the context of the

evolution of unorthodox reproductive systems, see next section). Another important assumption we made is that hybrid larvae do not develop into fertile queens, for instance owing to hybrid incompatibilities (Trible and Kronauer 2017). If fertile hybrid queens are produced regularly, evolution toward worker-loss like in our model is less likely to happen as hybrids no longer make a reliable source of workers. In ants at least, the idea that hybrid queens are rarely fertile is supported by the contrast between high frequency of interspecific mating on one hand, and weak genetic signals of interspecific gene flow on the other (Umphrey 2006; Feldhaar et al. 2008). Finally, we focused in the main text on the case where colony productivity increases linearly with

workers (i.e., the probability that a sexual survives until reproduction increases linearly with the number of workers). More realistically, the gain in productivity brought by one additional worker is likely to decrease with increasing workforce (Nonacs and Tobin 1992; Reuter and Keller 2001). Such diminishing returns tend to favor cheating because the indirect benefit of developing into a worker gets smaller as colony size increases (e.g., Reuter and Keller 2001; Field and Toyoizumi 2020). In line with this, we find that worker-loss evolves even more easily under diminishing compared to linear returns (Appendix B.2.3 and fig. S3).

AN ADAPTIVE PATH TO UNORTHODOX REPRODUCTIVE SYSTEMS?

Our result that sperm parasitism favors the emergence of worker-loss via the fixation of royal cheats may be relevant to unorthodox reproductive systems found in ants. Of particular interest is social hybridogenesis, whereby females produced through regular intralineage mating or thelytokous parthenogenesis develop into queens, while workers emerge from eggs fertilized by allospecific males (Helms Cahan et al. 2002; Helms Cahan and Vinson 2003; Anderson et al. 2006; Romiguier et al. 2017; Lacy et al. 2019; Kuhn et al. 2020). Such a striking system was first described just two decades ago in *Pogonomyrmex* harvester ants (Helms Cahan et al. 2002), and has since been found in several species spread across four genera (Helms Cahan et al. 2002; Helms Cahan and Vinson 2003; Romiguier et al. 2017; Lacy et al. 2019; Kuhn et al. 2020). If these observations suggest that social hybridogenesis has evolved independently multiple times, the evolutionary origins of this complex system remain poorly understood (Anderson et al. 2008; Schwander et al. 2010; Lavanchy and Schwander 2019). One early suggestion is based on the hypothesis that worker development requires the combination of co-adapted alleles at key loci (i.e., requires epistatic interactions; Helms Cahan and Keller 2003). According to this theory, worker-loss in hybridogenetic lineages would have originated in the random loss of such combinations during episodes of ancestral hybridization. Present hybridization would then have evolved to restore genetic combinations and epistatic interactions in F1-hybrids allowing for worker development.

Here, we have shown mathematically that social hybridogenesis could also result from additive genetic effects on caste development and queen-larvae conflicts within colonies. This theory, previously described verbally in Anderson et al. (2006, 2008), may help explain the multiple convergence toward social hybridogenesis because virtually every sexual eusocial species should experience queen-larvae conflicts over caste investment. Furthermore, because this path to social hybridogenesis does not depend on changes in the sympatric species whose sperm is parasitized, our model is relevant to both cases of asymmetrical (where the sympatric species produces workers through

regular sex, as, e.g., in *Solenopsis xyloni*; Helms Cahan and Vinson 2003) and symmetrical social hybridogenesis (where the sympatric species also produces workers via hybridization, as, e.g., in *Pogonomyrmex* harvester ants; Anderson et al. 2006).

Our model may also be relevant to other unorthodox systems of reproduction such as those found in populations of *Wasemannia auropunctata* (Fournier et al. 2005), *Vollenhovia emeryi* (Ohkawara et al. 2006), or *Paratrechina longicornis* (Pearcy et al. 2011). As with some forms of social hybridogenesis, queens of these systems produce their reproductive daughters via female parthenogenesis and their workers via sex with genetically distant males. In contrast to social hybridogenesis, however, these males belong to a divergent all-male lineage maintained by male clonality. This is further accompanied with a complete absence of arrhenotokous males (i.e., queens never make hemiconal haploid sons, as shown in *W. auropunctata*; Rey et al. 2013). When queens are able to produce daughters parthenogenetically in our model, evolution can lead to a state where worker-loss is coupled with a complete absence of intralineage mating (i.e., $\eta = 1$, Fig. 5C and D). In this state, arrhenotokous males represent a genetic dead-end, laying the basis for their disappearance. To investigate these systems in more detail, it would be interesting to extend our model to consider the evolution of female parthenogenesis and male clonality.

Our formal approach is especially useful in a context where hybrid vigor in workers has been raised to explain the evolutionary origin of social hybridogenesis and other hybridization-dependent systems (Julian and Cahan 2006; Umphrey 2006; Anderson et al. 2008; Feldhaar et al. 2008; Schwander et al. 2010). According to this argument, selection favored hybridization because hybrid workers are more efficient, more resilient, or better suited to exploit marginal habitats than regular workers. But in spite of much effort, empirical evidence supporting hybrid vigor in workers is still lacking (Ross and Robertson 1990; James et al. 2002; Julian and Cahan 2006; Feldhaar et al. 2008). Further challenging this view, we have shown here that hybrid vigor is not necessary to the evolution of hybridization-dependent reproductive systems. In fact, our results demonstrate that these systems can easily evolve even when hybridization is costly due to pre- and postzygotic barriers (i.e., when $e < 1$, e.g., because hybridization leads to an inefficient workforce due to hybrid incompatibilities in workers; or increased efforts in mate-finding and mating, Maroja et al. 2014; or the production of nonviable or infertile hybrid queens, Umphrey 2006; Feldhaar et al. 2008). In contrast to previous suggestions (Anderson et al. 2008), our model thus indicates that hybridization-dependent reproductive systems can emerge among species that have already substantially diverged, and can be maintained even with further accumulation of hybrid incompatibilities.

More generally, our results suggest that natural selection can lead to an association between hybridization and caste determination. To date, such associations have been reported in only 18 distinct ant species or populations (Helms Cahan et al. 2002; Helms Cahan and Vinson 2003; Fournier et al. 2005; Anderson et al. 2006; Ohkawara et al. 2006; Pearcy et al. 2011; Romiguier et al. 2017; Lacy et al. 2019; Kuhn et al. 2020). But this rarity may be due—at least partly—to the difficulty with describing these systems (which in particular requires sampling and genotyping both queens and workers of the same populations, Helms Cahan et al. 2002). For instance, studies specifically testing for social hybridogenesis discovered five new cases of this reproductive system in *Cataglyphis* (out of 11 species tested, Kuhn et al. 2020) and three in *Messor* (out of 9, Romiguier et al. 2017). These considerations, together with our results, support the notion that currently known cases likely represent only a small fraction of extant eusocial systems relying on hybridization (Helms Cahan et al. 2002; Lavanchy and Schwander 2019).

FACTORS PROMOTING THE EVOLUTION OF INTRASPECIFIC WORKER-LOSS

In addition to showing that hybrid vigor is not necessary for the emergence of intraspecific worker-loss, our model highlights several factors that can facilitate such evolution. The first of these is polyandry, which favors sperm parasitism and worker-loss by minimizing the risks associated with hybridization. Interestingly, even though polyandry is generally rare in social insects (Strassmann 2001; Hughes et al. 2008), meaningful exceptions are found in *Pogonomyrmex* (Rheindt et al. 2004) and *Messor* (Norman et al. 2016) harvester ants, two taxa where social hybridogenesis has evolved multiple times (Anderson et al. 2006; Romiguier et al. 2017). Although the number of males a queen mates with is fixed in our model, it is conceivable that this number also responds to hybridization, leading polyandry and hybridization to coevolve. Indeed as low levels of polyandry represent a risk for out-breeding queens, we can expect selection to favor queen behaviors that increase their number of mates. This would in turn allow for greater levels of hybridization, which would increase selection on polyandry and so on. We therefore expect that the coevolution between polyandry, hybridization, and caste determination further promotes the emergence of worker loss. For species that are fixed for strict (or close to) monandry, our model shows that worker-loss can evolve when queens have the ability to reproduce via thelytokous parthenogenesis as it allows interspecifically mated queens to nevertheless produce daughter queens. This supports the notion that thelytoky has been important for the convergent evolution of social hybridogenesis in the (mostly) monandrous *Cataglyphis* ants (Kuhn et al. 2020).

Although not considered in our study for simplicity, another factor that can minimize the risks associated with hybridization

in monandrous species is polygyny, whereby related queens form multi-queen nests. Such social organization allows both intra- and interspecifically mated queens to be part of the same colony, which can then produce both queens and workers. Polygyny should therefore further facilitate hybridization. Although this may have played a role in the evolution of social hybridogenesis in the polygynous *Solenopsis* species with this reproductive system (Helms Cahan and Vinson 2003; Lacy et al. 2019), we do not expect polygyny to be critical for the evolution of worker-loss as such loss has been described in both monogynic and polygynic species of the same genus (e.g., *Messor barbarus* and cf. *structor*; Romiguier et al. 2017). Beyond these considerations, any trait (e.g., polyandry, polygyny, or reproduction by workers) that influences kinship structure within colonies and thus modulates intracolonial conflicts has the potential to play a role in the evolution of worker-loss. Studying the evolution of such traits and its feedback on hybridization and caste determination therefore represents an interesting avenue for future research.

More important for the evolution of worker-loss in our model is that queens hybridize often enough. This readily happens when the propensity of queens to mate with allo- *versus* conspecific males evolves (Fig. 3). In this case, sperm parasitism, worker-loss, and social hybridogenesis emerge even in species that initially do not hybridize. Such evolution of hybridization is especially likely to occur where queens are able to recognize differences among males and choose their mates accordingly. There is, however, currently little, if any, evidence for such direct mate or sperm choice in eusocial insects (Strassmann 2001; Schwander et al. 2006; Umphrey 2006; Feldhaar et al. 2008). Alternatively, queens may be able to modulate the degree of hybridization via more indirect mechanisms, such as mating flight synchronization (Kaspari et al. 2001). Under completely random mating, hybridization can reach sufficient levels for worker-loss to evolve in our model as long as allo-specific males are sufficiently abundant (Fig. 2), for instance, because phenology is shared with an ecologically dominant species (Klein et al. 2017). In intermediate situations where allo-specific males are available but scarce, the evolution of caste determination under random mating leads to a situation where queens produce both hybrid and nonhybrid workers (Fig. 2A and B). Such a scenario may be relevant to species of ants where hybrid workers has been reported but where worker-loss has not evolved (e.g., in some North American *Solenopsis* or European *Temnothorax*; Feldhaar et al. 2008).

Whether it occurs randomly or not, hybridization requires pre-zygotic barriers to be sufficiently low. Various mechanisms, such as secondary contacts or high dispersal ability, are known to lower these barriers (de Aguiar et al. 2009). In particular, it has been proposed that the typically low phenotypic variation among males of different ant species facilitates hybridization in this taxa (Feldhaar et al. 2008). With these considerations in mind, it is

noteworthy that all known cases of social hybridogenesis have been found in ants that live in dry climates (Helms Cahan et al. 2002; Helms Cahan and Vinson 2003; Romiguier et al. 2017; Lacy et al. 2019; Kuhn et al. 2020), where the synchronicity of mating flights between species is highest due to shared dependence on punctual climatic events (Hölldobler and Wilson 1990; Feldhaar et al. 2008).

At a broader level, our results suggest that worker-loss can readily evolve when a source of workers that is impervious to royal cheats can be exploited by queens. Besides sperm parasitism, other forms of parasitism can provide such a source of workers and have been associated with worker-loss (Nonacs and Tobin 1992). In inquiline ants such as *Teleutomyrmex schneideri*, for instance, queens do not themselves produce workers but rather infiltrate the colony of a host and trick host workers into caring for their progeny (Hölldobler and Wilson 1990; Buschinger 2009). Like in our model, such social parasitism could be the endpoint of an arms race between queens and larvae of the same lineage, whereby increasingly caste-biasing cheats reduce colony workforce leading queens to increasingly rely on host workers.

CONCLUSIONS

Intracolonial conflicts are inevitably part of the social lives of nonclonal organisms. Here we have shown that such conflicts readily lead to an association between interspecific sperm parasitism and intraspecific worker-loss via the fixation of royal cheats. This association is especially relevant to the evolution of reproductive systems that like social hybridogenesis rely on hybridization. Beyond these unorthodox systems and sperm parasitism, the fixation of royal cheats and loss of intraspecific workers may be connected to other forms of antagonistic interspecific relationships such as social parasitism. More broadly, our model illustrates how the unique conflicts that are inherent to eusocial life can lead to evolutionary arms races, with implications for elaborate reproductive systems and novel ecological interactions between species.

ACKNOWLEDGMENTS

We thank Nicolas Galtier for useful discussions, Miya Qiaowei Pan, Laurent Keller, and Tanja Schwander for comments on an earlier version of our manuscript, the Swiss National Science Foundation (PCEFP3181243 to CM) and the french national research agency (ANR t-ERc RoyalMess) for funding.

AUTHOR CONTRIBUTIONS

AW, JR, and CM conceived the study. AW performed the analysis and wrote the first draft of the manuscript under the guidance of JR and CM. All authors contributed to the final version.

DATA ARCHIVING

A Mathematica notebook that reproduces our results and a R file implementing our simulations are available here: <https://zenodo.org/record/5167179>.

LITERATURE CITED

- Abrams, P. A., Y. Harada, and H. Matsuda. 1993. On the relationship between quantitative genetic and ESS models. *Evolution* 47:982–985.
- Alpedrinha, J., A. Gardner, and S. A. West. 2014. Haplodiploidy and the evolution of eusociality: worker revolution. *Am. Nat.* 184:303–317.
- Anderson, K. E., J. Gadau, B. M. Mott, R. A. Johnson, A. Altamirano, C. Strehl, et al. 2006. Distribution and evolution of genetic caste determination in *Pogonomyrmex* seed-harvester ants. *Ecology* 87:2171–2184.
- Anderson, K. E., T. A. Linksvayer, and C. R. Smith. 2008. The causes and consequences of genetic caste determination in ants (Hymenoptera: Formicidae). *Myrmecol. News* 11:119–132.
- Avila, P., and L. Fromhage. 2015. No synergy needed: ecological constraints favor the evolution of eusociality. *Am. Nat.* 186:31–40.
- J Boomsma, J. 2009. Lifetime monogamy and the evolution of eusociality. *Philos. Trans. R. Soc. B Biol. Sci.* 364:3191–3207.
- Bourke, A. F. G., and G. L. Chan. 1999. Queen-worker conflict over sexual production and colony maintenance in perennial social insects. *Am. Nat.* 154:417–426.
- Bränström, A., J. Johansson, and N. Von Festenberg. 2013. The hitchhiker's guide to adaptive dynamics. *Games* 4:304–328.
- V Brian, M. 1956. Studies of caste differentiation in *Myrmica Rubra* L. 4. Controlled larval nutrition. *Insect. Soc.* 3:369–394.
- . 1973. Temperature choice and its relevance to brood survival and caste determination in the ant *myrmica rubra* L. *Physiol. Zool.* 46:245–252.
- . 1974. Caste differentiation in *Myrmica rubra*: the role of hormones. *J. Insect Physiol.* 20:1351–1365.
- Buschinger, A. 2009. Social parasitism among ants: a review (Hymenoptera: Formicidae). *Myrmecol. News* 12:219–235.
- Caswell, H. 2000. Matrix population models. Sinauer Associates Inc, Sunderland, MA.
- Charlesworth, B. 1990. Optimization models, quantitative genetics, and mutation. *Evolution* 44:520–538.
- Crespi, B. J., and D. Yanega. 1995. The definition of eusociality. *Behav. Ecol.* 6:109–115.
- Darwin, C. 1859. On the origin of species. J. Murray, Lond., U.K.
- de Aguiar, M., M. Baranger, E. M. Baptestini, and L. Kaufman. 2009. Global patterns of speciation and diversity. *Nature* 460:384–387.
- Dercole, F., and S. Rinaldi. 2008. Analysis of evolutionary processes: The adaptive dynamics approach and its applications. Princeton Univ. Press, Princeton, NJ.
- Dobata, S. 2012. Arms race between selfishness and policing: two-trait quantitative genetic model for caste fate conflict in eusocial hymenoptera. *Evolution* 66:3754–3764.
- Douwes, P., and B. Stille. 1991. Hybridization and variation in the *Leptothorax tuberum* group (Hymenoptera: Formicidae). *Z. Zool. Syst. Evol.* 29:165–175.
- Feldhaar, H., S. Foitzik, and J. Heinze. 2008. Lifelong commitment to the wrong partner: hybridization in ants. *Philos. Trans. R. Soc. B Biol. Sci.* 363:2891–2899.
- Field, J., and H. Toyoizumi. 2020. The evolution of eusociality: no risk-return tradeoff but the ecology matters. *Ecol. Lett.* 23:518–526.
- Fournier, D., A. Estoup, J. Orivel, J. Foucaud, H. Jourdan, J. Le Breton, et al. 2005. Clonal reproduction by males and females in the little fire ant. *Nature* 435:1230–1234.

- Frank, S. A. 1998. Foundations of social evolution. Princeton Univ. Press, Princeton, NJ.
- Frohschammer, S., and J. Heinze. 2009. A heritable component in sex ratio and caste determination in a *Cardiocondyla* ant. *Front. Zool.* 6:1–6.
- Geritz, S. A. H., and M. Gyllenberg. 2005. Seven answers from adaptive dynamics. *J. Evol. Biol.* 18:1174–1177.
- Geritz, S. A. H., E. Kisdi, G. Meszéna, and J. A. J. Metz. 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* 12:35–57.
- Geritz, S. A. H., J. A. J. Metz, and C. Rueffler. 2016. Mutual invadability near evolutionarily singular strategies for multivariate traits, with special reference to the strongly convergence stable case. *J. Math. Biol.* 72:1081–1099.
- González-Forero, M. 2015. Stable eusociality via maternal manipulation when resistance is costless. *J. Evol. Biol.* 28:2208–2223.
- D Hamilton, W. 1964. The genetical evolution of social behaviour. I. *J. Theor. Biol.* 7:1–16.
- Hartfelder, K., G. R. Makert, C. C. Judice, G. A. G. Pereira, W. C. Santana, R. Dallacqua, et al. 2006. Physiological and genetic mechanisms underlying caste development, reproduction and division of labor in stingless bees. *Apidologie* 37:144–163.
- Helms Cahan, S., and L. Keller. 2003. Complex origin of a genetic system of caste determination in harvester ants. *Nature* 424:306–309.
- Helms Cahan, S., J. D. Parker, S. W. Rissing, R. A. Johnson, T. S. Polony, M. D. Weiser, et al. 2002. Extreme genetic differences between queens and workers in hybridizing *Pogonomyrmex* harvester ants. *Proc. R. Soc.* 269:1871–1877.
- Helms Cahan, S., and S. B. Vinson. 2003. Reproductive division of labor between hybrid and nonhybrid offspring in a fire ant hybrid zone. *Evolution* 57:1562–1570.
- Hölldobler, B., and E. Wilson. 1990. The ants. Harvard Univ. Press, Cambridge, MA.
- Hughes, W. H., B. P. Oldroyd, M. Beekman, and F. L. W. Ratnieks. 2008. Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* 320:1213–1216.
- Hughes, W. O. H., and J. J. Boomsma. 2008. Genetic royal cheats in leaf-cutting ant societies. *Proc. Natl. Acad. Sci.* 105.
- Iwasa, Y., A. Pomiankowski, and S. Nee. 1991. The evolution of costly mate preferences. II. The handicap principle. *Evolution* 45:1431–1442.
- James, S. S., R. M. Pereira, K. M. Vail, and B. H. Ownley. 2002. Survival of imported fire ant (Hymenoptera: Formicidae) species subjected to freezing and near-freezing temperatures. *Environ. Entomol.* 31:127–133.
- Julian, G. E., and S. H. Cahan. 2006. Behavioral differences between *Pogonomyrmex rugosus* and dependent lineages (H1/H2) harvester ants. *Ecol.logy* 87:2207–2214.
- Kaspari, M., J. Pickering, J. T. Longino, and D. Windsor. 2001. The phenology of a neotropical ant assemblage: evidence for continuous and overlapping reproduction. *Behav. Ecol. Sociobiol.* 50:382–390.
- Klein, E. K., L. Lagache-Navarro, and R. J. Petit. 2017. Demographic and spatial determinants of hybridization rate. *J. Ecol.* 105:29–38.
- Kuhn, A., H. Darras, O. Paknia, and S. Aron. 2020. Repeated evolution of queen parthenogenesis and social hybridogenesis in *Cataglyphis* desert ants. *Mol. Ecol.* 29:549–564.
- Lacy, K. D., D. Shoemaker, and K. G. Ross. 2019. Joint evolution of asexuality and queen number in an ant. *Curr. Biol.* 29:1394–1400.
- Lavanchy, G., and T. Schwander. 2019. Hybridogenesis. *Curr. Biol.* 29:R1–R15.
- Lehmann, L., C. Mullon, E. Akçay, and J. V. Cleve. 2016. Invasion fitness, inclusive fitness, and reproductive numbers in heterogeneous populations. *Evolution* 70:1689–1702.
- Lehmann, L., V. Ravigne, and L. Keller. 2008. Population viscosity can promote the evolution of altruistic sterile helpers and eusociality. *Proc. R. Soc. B Biol. Sci.* 275:1887–1895.
- Leimar, O. 2009. Multidimensional convergence stability. *Evol. Ecol. Res.* 11:191–208.
- Libbrecht, R., M. Corona, F. Wende, D. O. Azevedo, J. E. Serrão, and L. Keller. 2013. Interplay between insulin signaling, juvenile hormone, and vitellogenin regulates maternal effects on polyphenism in ants. *Proc. Natl. Acad. Sci.* 110:11050–11055.
- A Linksvayer, T. 2006. Direct, maternal and sibsocial genetic effects on individual and colony traits in an ant. *Evolution* 60:2552–2561.
- Maroja, L. S., Z. M. McKenzie, E. Hart, J. Jing, E. L. Larson, and D. P. Richardson. 2014. Barriers to gene exchange in hybridizing field crickets: the role of male courtship effort and cuticular hydrocarbons. *BMC Evol. Biol.* 14:1–10.
- Moritz, R. F. A., H. M. G. Lattorff, P. Neumann, F. B. Kraus, S. E. Radloff, and H. R. Hepburn. 2005. Rare royal families in honeybees. *Apis mellifera*. *Naturwissenschaften* 92:488–491.
- Mullon, C., L. Keller, and L. Lehmann. 2018. Social polymorphism is favoured by the co-evolution of dispersal with social behaviour. *Nat. Ecol. Evol.* 2:132–140.
- Mullon, C., and L. Lehmann. 2019. An evolutionary quantitative genetics model for phenotypic (co)variances under limited dispersal, with an application to socially synergistic traits. *Evolution* 4:166–34.
- Nonacs, P., and J. E. Tobin. 1992. Selfish larvae: Development and the evolution of parasitic behavior in the Hymenoptera. *Evolution* 46:1605–1620.
- Norman, V., H. Darras, C. Tranter, S. Aron, and W. O. H. Hughes. 2016. Cryptic lineages hybridize for worker production in the harvester ant *Messor barbarus*. *Biol. Lett.* 12:1–5.
- Ohkawara, K., M. Nakayama, A. Satoh, A. Trindl, and J. Heinze. 2006. Clonal reproduction and genetic caste differences in a queen-polymorphic ant, *Vollenhovia emeryi*. *Biol. Lett.* 5:359–363.
- Parker, G. A., and J. Maynard Smith. 1990. Optimality theory in evolutionary biology. *Nature* 348:27–33.
- Pearcy, M., M. A. D. Goodisman, and L. Keller. 2011. Sib mating without inbreeding in the longhorn crazy ant. *Proc. R. Soc.* 278:2677–2681.
- Penick, C. A., and J. Liebig. 2012. Regulation of queen development through worker aggression in a predatory ant. *Behav. Ecol.* pp. 992–998.
- Penick, C. A., S. S. Prager, and J. Liebig. 2012. Juvenile hormone induces queen development in late-stage larvae of the ant *Harpegnathos saltator*. *J. Insect Physiol.* 58:1643–1649.
- Phillips, P., and S. J. Arnold. 1989. Visualizing multivariate selection. *Evolution* 43:1209–1222.
- C Queller, D. 1994. Extended parental care and the origin of eusociality. *Proc. R. Soc. B Biol. Sci.* 256:105–111.
- Quiñones, A. E., and I. Pen. 2017. A unified model of Hymenopteran preadaptations. *Nat. Commun.* 8:1–13.
- Rabeling, C., and D. J. C. Kronauer. 2013. Thelytokous parthenogenesis in eusocial hymenoptera. *Annu. Rev. Entomol.* 58:273–292.
- Reuter, M., and L. Keller. 2001. Sex ratio conflict and worker production in eusocial Hymenoptera. *Am. Nat.* 158:166–177.
- Rey, O., B. Facon, J. Foucaud, A. Loiseau, and A. Estoup. 2013. Androgenesis is a maternal trait in the invasive ant *Wasmannia auropunctata*. *Proc. R. Soc. B Biol. Sci.* 280:1–7.
- Rheindt, F. E., J. Gadau, C.-P. Strehl, and B. Hölldobler. 2004. Extremely high mating frequency in the Florida harvester ant (*Pogonomyrmex badius*). *Behav. Ecol. Sociobiol.* 56:472–481.
- Ross, K. G., and J. L. Robertson. 1990. Developmental stability, heterozygosity, and fitness in two introduced fire ants (*Solenopsis invicta* and *S. richteri*) and their hybrid. *Heredity* 64:93–103.

- Romiguier, J., A. Fournier, S. H. Yek, and L. Keller. 2017. Convergent evolution of social hybridogenesis in *Messor* harvester ants. *Mol. Ecol.* 26:1108–1117.
- Rousset, F. 2004. Genetic structure and selection in subdivided populations. Princeton Univ. Press, Princeton, NJ.
- Rousset, F., and O. Ronce. 2004. Inclusive fitness for traits affecting metapopulation demography. *Theoret. Popul. Biol.* 65:127–141.
- Schwander, T., S. H. Cahan, and L. Keller. 2006. Genetic caste determination in *Pogonomyrmex* harvester ants imposes costs during colony founding. *J. Evol. Biol.* 19:402–409.
- Schwander, T., J.-y. Humbert, C. S. Brent, S. H. Cahan, L. Chapuis, E. Renai, et al. 2008. Maternal effect on female caste determination in a social insect. *Curr. Biol.* 18:265–269.
- Schwander, T., and L. Keller. 2008. Genetic compatibility affects queen and worker caste determination. *Science* 322:552.
- Schwander, T., N. Lo, M. Beekman, B. P. Oldroyd, and L. Keller. 2010. Nature versus nurture in social insect caste differentiation. *Trends Ecol. Evol.* 25:275–282.
- Seger, J. 1981. Kinship and covariance. *J. Theor. Biol.* 91:191–213.
- . 1983. Partial bivoltinism may cause alternating sex-ratio biases that favour eusociality. *Nature* 301:59–62.
- Smith, C. R., K. E. Anderson, C. V. Tillberg, J. Gadau, and A. V. Suarez. 2008. Caste determination in a polymorphic social insect: nutritional, social, and genetic factors. *Am. Nat.* 172:497–507.
- Strassmann, J. 2001. The rarity of multiple mating by females in the social Hymenoptera. *Insect. Soc.* 48:1–13.
- Taylor, P. D., and S. A. Frank. 1996. How to make a kin selection model. *J. Theor. Biol.* 180:27–37.
- Tribble, W., and D. J. C. Kronauer. 2017. Caste development and evolution in ants: it's all about size. *J. Exp. Biol.* 53:53–62.
- Trivers, L. R., and H. Hare. 1976. Haplodiploidy and the evolution of the social insects. *Science* 191:249–263.
- J Umphrey, G. 2006. Sperm parasitism in ants: selection for interspecific mating and hybridization. *Ecology* 87:2148–2159.
- Wenseleers, T., A. G. Hart, and F. L. W. Ratnieks. 2004. When resistance is useless: policing and the evolution of reproductive acquiescence in insect societies. *Am. Nat.* 164:E154–E167.
- Winter, U., and A. Buschinger. 1986. Genetically mediated queen polymorphism and caste determination in the slave-making ant, *Harpagoxenus sublaevis* (Hymenoptera: Formicidae). *Entomol. Gen.* 11:125–137.
- Wolf, J. I., P. Punttila, and P. Seppä. 2018. Life-history trait variation in a queen-size dimorphic ant. *Ecol. Entomol.* 43(6): 763–773.
- Wolfram Research, Inc. 2020. Mathematica.

Associate Editor: C. Moreau

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1: Colonial investment in males, queens and workers.

Figure S1: Properties of the internal singular strategy under monoandry and low polyandry.

Figure S2: Polymorphism under monandry is due to positive correlational selection. A.

Figure S3: Non-linear effects of investment in workers.

Detection of F1 Hybrids from Single-genome Data Reveals Frequent Hybridization in Hymenoptera and Particularly Ants

Arthur Weyna,* Lucille Bourouina, Nicolas Galtier  and Jonathan Romiguier*,†

Institut des Sciences de l'Evolution (UMR 5554), University of Montpellier, CNRS Montpellier, France

*Corresponding author: E-mail: arthur.weyna@umontpellier.fr.

†These authors share senior authorship.

Associate editor: Keith Crandall

Abstract

Hybridization occupies a central role in many fundamental evolutionary processes, such as speciation or adaptation. Yet, despite its pivotal importance in evolution, little is known about the actual prevalence and distribution of current hybridization across the tree of life. Here we develop and implement a new statistical method enabling the detection of F1 hybrids from single-individual genome sequencing data. Using simulations and sequencing data from known hybrid systems, we first demonstrate the specificity of the method, and identify its statistical limits. Next, we showcase the method by applying it to available sequencing data from more than 1,500 species of Arthropods, including Hymenoptera, Hemiptera, Coleoptera, Diptera, and Archnida. Among these taxa, we find Hymenoptera, and especially ants, to display the highest number of candidate F1 hybrids, suggesting higher rates of recent hybridization between previously isolated gene pools in these groups. The prevalence of F1 hybrids was heterogeneously distributed across ants, with taxa including many candidates tending to harbor specific ecological and life-history traits. This work shows how large-scale genomic comparative studies of recent hybridization can be implemented, uncovering the determinants of first-generation hybridization across whole taxa.

Key words: hybridization, coalescent, F1 hybrids detection, arthropods, hymenoptera, ants.

Introduction

Hybridization, whereby members of genetically distinct populations mate and produce offspring of mixed ancestry (Barton and Hewitt 1985; Abbott et al. 2013), has received much attention since the early days of evolutionary biology. From the onset, Darwin and his contemporaries spent a great deal of time studying hybrids and their fitness, which they recognized as a challenge to a discrete definition of species (Roberts 1919). But the crucial importance of hybridization to biological evolution was fully realized only with the development of genetics in the following century. Formal studies of hybridization genetics led to the formulation of the biological species concept, and to the fundamental insight that speciation is generally driven by the evolution of isolating mechanisms in response to hybridization (Dobzhansky 1940; Mayr 1942; Smadja and Butlin 2011; Abbott et al. 2013). The advent of genetic data also revealed the role of hybridization and introgression as important contributors to genetic variation and adaptation in many existing species (Anderson 1953; Harrison and Larson 2014), especially in the contexts of changing environments (Hamilton and Miller 2016) and biological invasion (Prentis et al. 2008). Additionally, while hybridization was thought by many biologists to be

relevant only for a few taxa such as plants (Barton 2001), the accumulation of molecular data has continuously revealed its presence in many groups, including mammals, birds, fish, fungi, and insects (Taylor and Larson 2019), with Mallet (2005) estimating that at least 10% of animal species frequently hybridize. These findings have further underlined the importance of hybridization in understanding many micro- and macro-evolutionary patterns across the tree of life (Abbott et al. 2013).

The same findings, however, also corroborated the old intuition that taxa can differ greatly in their susceptibility to hybridize, fueling discussions about the determinants of such heterogeneity (see Mallet 2005 for a useful review). It was first understood that groups displaying a high number of sympatric species with low divergence, where the contact between compatible species is maximized, should be the most likely to hybridize (Edmands 2002; Price and Bouvier 2002). But sympathy and divergence are by themselves incomplete predictors of hybridization frequency, as strong reproductive barriers can arise from discrete evolutionary events (e.g., chromosome rearrangements or cytoplasmic incompatibilities; Bordenstein et al. 2001; Fishman et al. 2013), and can be rapidly selected for (i.e., reinforcement) or against depending on the relative fitness of hybrids (Smadja and Butlin 2011). To understand heterogeneity in hybridization

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

rates, it is thus important to also consider these ecological and phenotypic features of species that influence hybrid fitness, and more generally that influence the cost or gain in producing hybrids (Mallet 2005). For instance, hybridization has been found to be more frequent in populations of spadefoot toads inhabiting ephemeral environments where hybrids outperform (Pfennig 2007), or in rare species of birds where allosppecific mates are easier to come by Randler (2002). A similar point was made by Mayr (1963), who suggested that polygamous species of birds should be the most likely to hybridize, because males with low parental investment should be more likely to accept interspecific mates. This early hypothesis is particularly significant in that it emphasizes on the idea that among characteristics of species relevant to hybridization, their life-history and mating system are of central importance.

One specific taxon in which relations between hybridization, mating systems and life-history have been extensively discussed is ants (Formicidae). Some ant genera are known to display unusually high rates of hybridization, based on both morphological and molecular data (Nonacs 2006; Umphrey 2006; Feldhaar et al. 2008). The first key trait of ants invoked to explain this pattern is haplodiploidy, a trait common to all Hymenoptera. Because males of Hymenoptera are haploids produced without fecundation, it is likely that hybrid sterility does not nullify the fitness of female Hymenoptera, which can still produce males after hybridizing (Nonacs 2006; Feldhaar et al. 2008). This particularity of haplodiploids would hinder selection against hybridization and limit the formation of strict barriers to interspecific mating. A second important ancestral trait of ants is eusociality, whereby reproductive females (i.e., queens) produce a large number of sterile helper individuals (i.e., workers) to form colonies. It was hypothesized that selection against hybridization is weaker in eusocial species because the fitness cost of hybrid sterility should be minimal in species producing a large majority of sterile individuals (Nonacs 2006; Umphrey 2006). This is especially likely in species in which queens mate multiply, and can combine inter- and intra-specific matings to ensure the production of a fraction of nonhybrid daughters (Cordonnier et al. 2020). Such interplay between hybridization, mating systems and life-history culminates in a handful of ant species that display unique hybridization-dependent reproductive systems, such as social hybridogenesis (Helms Cahan et al. 2002; Helms Cahan and Vinson 2003; Fournier et al. 2005; Anderson et al. 2006; Ohkawara et al. 2006; Pearcy et al. 2011; Romiguier et al. 2017; Lacy et al. 2019; Kuhn et al. 2020). In these species, the cost of hybrid sterility is fully avoided because strong genetic caste determination constrains the development of hybrids towards the worker caste, while reproductive females can only be produced through intra-specific mating or parthenogenesis. The prevalence of hybridization-dependent systems within ants is virtually unknown (Anderson et al. 2008), but because they maintain large cohorts of first-generation (F1) hybrid workers, they may help explain observations of high hybridization rates in ants.

While several hypotheses have been proposed to explain variation in hybridization rates across taxa, empirical comparative studies are still lacking, impeding any further understanding of its determinants. This is mainly due to the difficulties in evaluating the prevalence of hybridization at the group level. Methods to detect hybridization typically rely either on ambiguous morphological identification (which can lead to important ascertainment bias; Mallet 2005), or on the use of large population-scale genetic samples including data for potential parental species (Anderson and Thompson 2002; Payseur and Rieseberg 2016; Schubert et al. 2017). These methods are sensitive and reliable in inferring hybrid status, and can yield substantial information regarding both recent and ancient events of hybridization and introgression. The same methods however, also require large investments in time and money to produce results. To allow for comparative studies of hybridization at the level of entire taxa, it is necessary to implement methods that can be applied to many nonmodel species in parallel. In particular, methods applicable to the large volume of already published phylogenomic data (i.e., with one sequenced genome per species) would be especially desirable and cost-effective. For instance, phylogenomic data are available for more than 900 species of ants (223 represented genera), as the result of an extensive effort of Branstetter et al. (2017), who set a goal to sequence a large part of the diversity of Formicidae using standardized protocols (Faircloth et al. 2012). The same type of data has also been produced for many other Hymenoptera, and for other groups of Arthropods (including Hemiptera, Coleoptera, Diptera, and Arachnida), thus calling for a comparative study of hybridization prevalence across these taxa of interest. The main issue about such single-genome data is that they do not allow for the inference of complex histories of hybridization and introgression, which unavoidably requires population genetic data. Yet, heterozygosity distribution in a single-genome theoretically contains enough information to predict whether an individual is a first-generation hybrid or not. The frequency of F1 hybrids can thus be estimated from large phylogenetic datasets, and be used as a proxy for ongoing rates of recent hybridization. Such an exploratory approach has the potential to identify previously unknown hybridization hotspots, allowing for comparative studies and paving the way for more informative population genetic studies.

In this study, we implement a coalescent-based statistical method that allows for the detection of F1-hybrids using single diploid genomes. We first test this method and assess its efficiency using simulations and real data from identified F1 hybrid and nonhybrid individuals. We then apply the method to phylogenomic data, assessing the prevalence of F1 hybrids among five groups of Arthropods.

Materials and Methods

Model

In this section, we present the coalescent-based model of divergence which forms the basis of our F1 hybrid detection

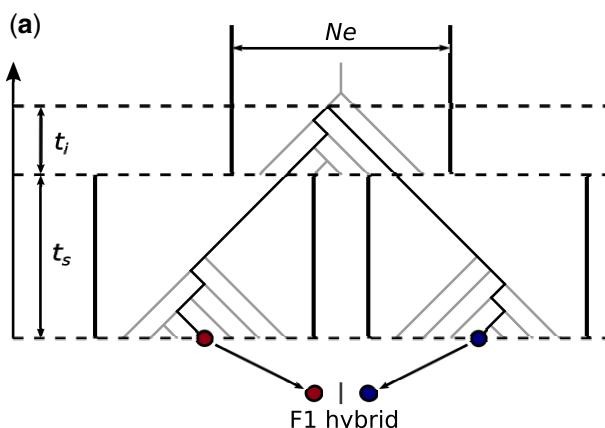
procedure. A F1 hybrid is the result of a cross between individuals from two different species. The heterozygosity of a such hybrid therefore reflects the divergence between its parental species, and can be modeled as shown in figure 1. This model describes the expected distribution of the number of differences between two alleles found in an F1 hybrid in terms of two main parameters: the divergence time between the two parental species t_s and the ancestral population size in their common ancestor Ne (fig. 1a). Briefly, if no migration occurred between the two parental lineages after their separation, and if t_s is large enough for lineage sorting to be complete, the total coalescence time of the two alleles is the sum of the divergence time t_s and the coalescence time in the ancestral population t_i . It is known from standard coalescence theory that the distribution of t_i is well approximated by an exponential distribution with mean $2Ne$ (Wakeley 2008). Consider a locus i of sequence length l_i at which the two alleles of an F1 hybrid individual have been sequenced. Assuming an infinite-site mutation model with constant per-site mutation rate μ , the number n_i of observed allelic differences follows a Poisson distribution with mean $2l_i\mu(t_i + t_s)$, that is

$$n_i \sim P[2l_i\mu(t_i + t_s)] \quad \text{with } t_i \sim E\left(\frac{1}{2Ne}\right). \quad (1)$$

where P and E denote the Poisson and exponential distributions, respectively. Equation (1) leads to an expression for the probability to observe a number k of allelic differences between alleles at any given locus i in a F1 hybrid (see supplementary Appendix A, Supplementary Material online for a complete derivation),

$$\Pr(n_i = k) = \frac{(l_i\theta)^k e^{(\gamma/\theta)}}{k!(l_i\theta + 1)^{k+1}} \int_{(l_i\gamma + \gamma/\theta)}^{\infty} t^k e^{-t} dt \quad (2)$$

with $\begin{cases} \theta = 4Ne\mu \\ \gamma = 2t_s\mu \end{cases}$



where θ is the ancestral population mutation rate, and γ is a measure of the heterozygosity acquired during the divergence process. Under the assumptions that μ , t_s , and Ne are constant across a set of j independent loci in a given diploid individual, each locus can be considered as a replicate of the same divergence scenario. In this case, the likelihood function of the set of observed numbers of differences between alleles is obtained by multiplying equation (2) across loci, and can be used to jointly estimate of θ and γ . To our knowledge, this model was first introduced by Takahata et al. (1995), and later refined by Yang (1997), in the context of phylogenomics and ancestral population size estimation, with equation (2) being the continuous equivalent to equation (8) given in Yang (1997).

Figure 1b illustrates the signal that is intended to be captured when estimating γ and θ . In a nonhybrid individual (i.e., whose parents belong to the same panmictic population), coalescence times between allele pairs are expected to follow an exponential distribution with mean and variance both determined by Ne (fig. 1b; top). In F1 hybrids, coalescence times are further increased by a fixed amount, which corresponds to the number of generations of divergence between the parental populations (fig. 1b; bottom, red bars). This uniform increase in coalescence times brought by divergence logically leads to an increased average coalescence time. This effect, however, is not by itself diagnostic of hybridization as it could be produced by an increase in Ne . Instead, what constitutes a unique signature of F1 hybrids is a decrease in the variance of coalescence times relative to the mean (fig. 1b, compare bottom to top). The relative variance in coalescence times is expected to be highest in nonhybrids, and to approach zero in F1 hybrids as the divergence between parental populations increases. The γ parameter captures this effect, whereas both γ and θ monitor the mean coalescence time. In other words, a nonzero estimate of γ means that the observed divergence between alleles is more similar across loci than expected under the standard coalescent.

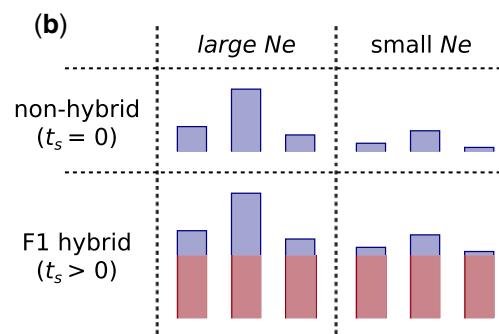


Fig. 1. Coalescent-based model of divergence. (a) The population history assumed in the model of divergence described in the main text (eq. 1). The darkened path represents the coalescence history of the two alleles (red and blue dots) that make up one locus in a diploid F1 hybrid. (b) Expected distribution of coalescence times for different values of Ne and t_s . Blue bars represent the components of coalescence time linked to coalescence in the ancestral population. Red bars represent the uniform increase in coalescence times brought by divergence between parental populations.

Because the proposed statistical procedure partitions observed heterozygosity between γ and θ , it is expected that estimates of both parameters will be positively correlated with the genetic diversity of samples. For instance, a sample with low heterozygosity can only yield low estimates of γ and θ . For this reason, we mostly relied on the ratio γ/θ , which is not directly related to sample heterozygosity. This ratio is expected to be close to zero in nonhybrids, and nonzero in F1 hybrids. Furthermore, a γ/θ ratio above one implies that the divergence time between parental populations is longer than $2Ne$ generations, which is the expected time for complete lineage sorting. Such a high value is very unlikely to be reached by nonhybrid individuals.

Simulated Test Loci Sets

To start evaluating our ability to detect F1 hybrids amongst diploid individuals, we simulated F1 hybrid, nonhybrid and first-generation backcross hybrid samples in the following manner. Individual diploid loci were simulated by using *ms v2014.03.04* (Hudson 2002) to sample pairs of alleles, together with the corresponding two alleles gene trees, under the demographic scenario described in figure 1a. To span across realistic values of both parameters of interest, values of θ and γ in simulations were set to be $\{10^{-4}, 10^{-3}, 10^{-2}\}$ and $\{0, 10^{-4}, 10^{-3}, 10^{-2}\}$, respectively. Once obtained, gene trees were converted to explicit nucleotide sequences pairs through the application of a HKY mutation model using *seq-gen v1.3* (Rambaut and Grassly 1997). The length of simulated sequences was set to be normally distributed with mean 1,000 bp and standard deviation 300 bp. At this point, simulated F1 hybrid and nonhybrid individuals were constructed by putting together independent collections of sequences pairs simulated under $\gamma > 0$ and $\gamma = 0$, respectively. First generation backcross individuals were constructed by putting together both types of sequences pairs in random proportions following a binomial distribution with $p = 0.5$, (i.e., as expected from a backcross with random meiosis and no linkage). Ten individuals of each type were constructed for each possible combination of parameter values, and for two possible loci set sizes (200 or 500 loci). Finally, simulated individuals (i.e., loci sets) were sequenced *in-silico* using *art_illumina v2.5.8* (Huang et al. 2012). We emulated standard PE150 sequencing on HiSeq 2500 with 10X coverage, using a standard normally distributed fragment size with mean 400 bp and standard deviation 20 bp.

UCE Datasets

We used ultraconserved elements (UCEs) in all applications to real data. UCEs are short (around 100 bp on average) independent genomic regions that are conserved without duplication across large phylogenetic groups (Faircloth et al. 2012). While these small regions themselves are too conserved to contain enough signal about recent divergence, their variable flanking regions are

mostly neutral and are thus expected to carry enough information to distinguish closely related species or lineages (i.e., as is done in standard UCE phylogenomics). UCEs are usually sequenced through hybridization capture protocols (Faircloth 2017; Miles Zhang et al. 2019), but subsets of UCEs that correspond to transcribed genomic regions can also be retrieved from transcriptomic data (Bossert et al. 2019; Miles Zhang et al. 2019). This last fact is convenient in the context of this study, because transcriptome sequencing data are available for known hybrid systems, featuring *a priori* identified F1 hybrids and nonhybrid individuals, and can be used to further test our procedure using real data. We retrieved transcriptome sequencing data published on *genbank* from two types of well-characterized F1 hybrids: 12 hybrid workers from the harvester ant *Messor barbarus* (Romiguier et al. 2017), and 18 *Equus caballus x asinus* hybrids (nine mules and nine hinny; Wang et al. 2019). Data from the same sources for seven haploid males and five nonhybrid queens of *M. barbarus*, as well as for five donkeys, were added for comparison. Genbank identifiers and metadata for *Messor* and *Equus* samples are available in supplementary tables S1 and S2, Supplementary Material online, respectively.

Sequencing data obtained through UCE-capture protocols has been published for a large number of nonmodel species, especially in Hymenoptera (Faircloth et al. 2012; Miles Zhang et al. 2019), thus allowing for a large-scale search for F1 hybrids in these groups. We retrieved from *genbank* UCE-capture sequencing data from diploid samples belonging to groups of Arthropods for which specific capture probe sets were available: Formicidae ("Insect Hymenoptera 2.5K version 2, Ant-Specific" probe set; Branstetter et al. 2017), nonFormicidae Hymenoptera ("Insect Hymenoptera 2.5K version 2, Principal" probe set; Branstetter et al. 2017), Hemiptera ("Insect Hemiptera 2.7K version 1" probe set; Branstetter et al. 2017 and Kieran et al. 2018), Coleoptera ("Insect Coleoptera 1.1K version 1" probe set; Faircloth 2017), Diptera ("Insect Diptera 2.7K version 1" probe set; Faircloth 2017), and Arachnida ("Arachnida 1.1K version 1" probe set; Faircloth 2017 and Starrett et al. 2016). To minimize the statistical weight of multiply sampled species, while maximizing statistical power at the group level, we kept only one sample per identified species (choosing samples with highest file size) and all samples lacking a complete identification (identified only to the genus level). Hymenoptera samples reported as males were considered as haploid and discarded. All remaining data files were downloaded from *genbank* using the *fasterq-dump* program from *SRA Toolkit v2.10.9*. Genbank identifiers and metadata for these samples are available in supplementary table S3, Supplementary Material online.

Parameters Estimation

To obtain estimates of γ and θ from simulated and real sequencing data, we systematically applied the following procedure. Raw read files were cleaned with *fastp v0.20.0*

(Chen et al. 2018) to remove adapters, reads shorter than 40 bp, and reads with less than 70% of bases with a phred score below 20. Cleaned reads were then assembled using *megahit v1.1.3* (Li et al. 2015) with k-mer size spanning from 31 to 101 by steps of 10. The *phyluce v1.6* (Faircloth 2016) tool suite was used to identify and isolate UCE loci from de-novo assemblies, by blasting contigs against UCE probe sets with the *phyluce_assembly_match_contigs_to_probes* function. In this step, assemblies obtained from test samples of *M. barbarus* and *Equus* were blasted against the “*Insect Hymenoptera 2.5K version 2, Ant-Specific*” (Branstetter et al. 2017) and the “*Tetrapods 5K version 1*” (Faircloth et al. 2012) UCE probe sets, respectively. Likewise, assemblies obtained from UCE-capture samples were blasted against the probe set associated with their phylogenetic group. As no probe set exists for simulated loci, these were blasted against custom probe sets constructed from their true sequence (i.e., as output by *seqgen*). Following this step, cleaned sequencing reads were realigned to isolated loci using *bwa v0.7.17* (Li and Durbin 2009) with default settings, and *angsd v0.921* (Korneliussen et al. 2014) was used to obtain allelic substitutions counts from read alignment files. Finally, we obtained estimates of θ and γ through bayesian estimations, using the R package *rstan v2.21.2* (Stan Development Team 2019, 2020) and uninformative priors spanning all realistic values for both parameters (i.e., uniform priors constrained between 0 and 0.2). The mean of the posterior distribution of each parameter was used as a point estimate, while credibility intervals were constructed from its 2.5% and 97.5% quantiles. R scripts and

the *stan* file necessary to run statistical estimations on a given set of observed allelic differences counts are available as [supplementary documents, Supplementary Material online](#) (<https://zenodo.org/record/5415947>).

Results

Simulations

Applying our estimation procedure on simulated data, we find that our method can be used to efficiently discriminate F1 hybrids from nonhybrids and first-generation backcross hybrids. Accurate divergence estimates can be obtained in simulated F1 hybrids using as little as 200 loci (fig. 2), provided that γ is in the same order of magnitude as the ancestral population mutation rate θ or higher (i.e., consistent with the model’s requirement of complete lineage sorting). Under the same condition, estimates of γ in nonhybrids and backcross hybrids are lower and do not exceed values one order of magnitude below estimated ancestral population size θ (which are themselves accurate; see [supplementary fig. S1, Supplementary Material online](#)). Across all simulations, 61.1% of simulated F1 hybrids yielded both estimates of γ higher than 10^{-3} and estimates of γ/θ higher 1, while no backcross hybrids or nonhybrids did, demonstrating the specificity of the method. Furthermore, we find that increasing the number of sequenced loci from 200 to 500 does not increase our ability to identify F1 hybrids (see fig. ??), which suggests than 200 loci is a good minimal requirement in applications to real data.

Simulations also revealed the statistical limits of our approach, which tends to overestimate the divergence

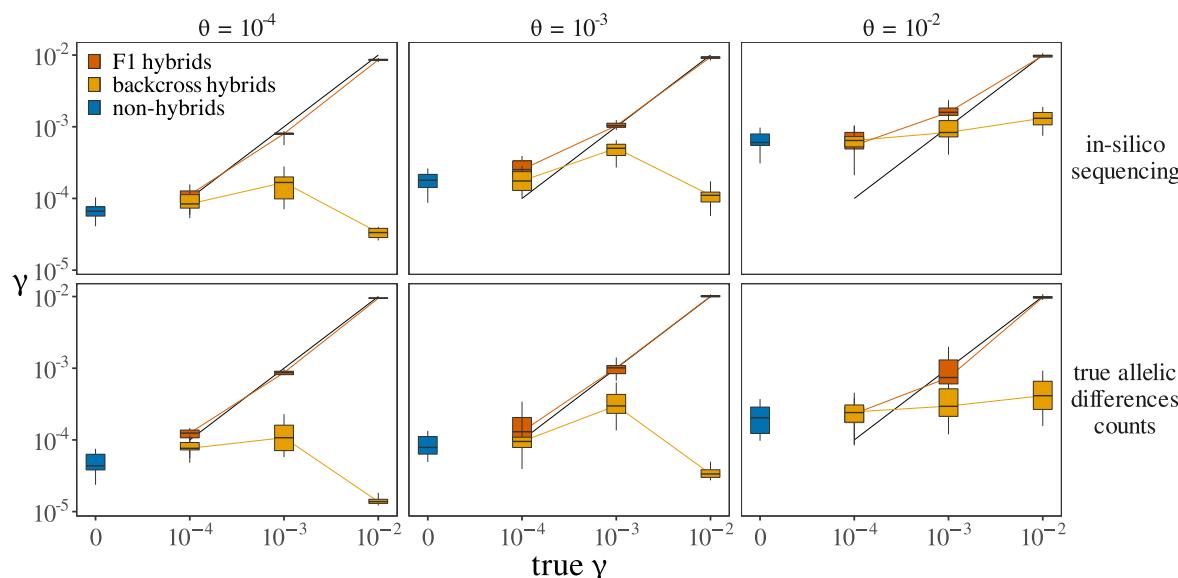


Fig. 2. Estimates of divergence in simulated individuals. Each box represents the distribution of estimated γ values across 10 simulated individuals. Every individual consists of a collection of 200 loci simulated under a given combination of true θ (given in headers) and γ (given in x-axis) values. In nonhybrid individuals γ is always zero. In backcross hybrids, the true value of γ is that given in the header, but only for a binomial proportion of loci (as described in the main text). The top row represents values obtained when estimating γ on loci sets obtained through the complete simulation procedure (including in-silico sequencing, read assembly and realignment, and substitutions counts estimation). The bottom row represent values obtained when estimating γ on sets of true counts data as output by ms (i.e., skipping subsequent simulation steps).

parameter γ whenever the true ancestral population mutation rate θ is high (fig. 2, top row). This translates into estimates of γ departing from zero in nonhybrids with high overall polymorphism. Interestingly, this overestimation can be shown to arise in part from error in genome assembly, reads alignments, and estimations of allelic differences counts. When estimating parameters using sets of true allelic differences counts as first output by ms (fig. 2, bottom row), divergence overestimation is less important in hybrids and nonhybrids. This suggests that in real data, a negative correlation will be expected between divergence estimates and overall sample quality.

Accurate Identification of F1 Hybrids in Two Known Hybrid Systems

To further quantify our ability to distinguish between nonhybrids and typical F1 hybrid individuals, we applied our estimation procedure to sequencing data from two types of well-characterized F1 hybrids, hybrid workers from the harvester ant *M. barbarus* (Romiguier et al. 2017), and *Equus caballus* × *asinus* hybrids (mules and hinnies) (Wang et al. 2019). Sequencing data from the same sources for males and nonhybrid queens of *M. barbarus*, as well as for donkeys, were added to the analysis for comparison. This analysis confirmed that F1 hybrids and nonhybrid individuals can be discriminated without ambiguity (fig. 3; parameters estimates are given in supplementary table S1 and S2, Supplementary Material online). Estimates of divergence (γ) in F1 hybrids always strongly departed from 0 and showed little variation across samples ($3.39 \times 10^{-3} \pm 2.05 \times 10^{-4}$ sd in *M. barbarus* workers; $1.47 \times 10^{-3} \pm 1.46 \times 10^{-4}$ sd in mules and hinnies). By

contrast, estimated values of γ in nonhybrid samples were always much closer to 0 in nonhybrid individuals ($2.34 \times 10^{-5} \pm 3.29 \times 10^{-5}$ sd in *M. barbarus* males and queens; $1.63 \times 10^{-5} \pm 3.31 \times 10^{-6}$ sd in donkeys). The ratio γ/θ reached the critical value of one in *M. barbarus* workers (1.067 ± 0.190 sd) while being two orders of magnitude lower in males and queens (0.012 ± 0.010 sd). This confirms that such a threshold value is reliable for discriminating true F1 hybrids. Ratios obtained in mules and hinnies are lower than 1 however (0.418 ± 0.061 sd), further suggesting that $\gamma/\theta > 1$ is a conservative requirement likely not to be reached by many true F1 hybrid. Interestingly, UCE-capture data for a single worker of *M. barbarus* (fig. 3a) led to slightly higher parameter estimates than transcriptomic data, but to a similar γ/θ ratio (1.131). This suggests that UCEs retrieved from transcriptomes of *M. barbarus* are less polymorphic on average, but contain the same information regarding relative divergence and hybrid status.

High Prevalence of F1 Hybrids in Hymenoptera and Formicidae

The application of our procedure to UCE-capture data, comprised of many samples of heterogeneous quality, led to the observation of the quality bias predicted from simulations. Specifically, we noted that older samples yielded slightly higher γ/θ estimates on average than recent ones, resulting in a significant correlation between the later ratio and specimen collection date ($\rho = -0.275$, p -value $< 2.2 \times 10^{-6}$). This bias is most likely due to lower sequence quality and increased data treatment error in old specimen, which leads to an

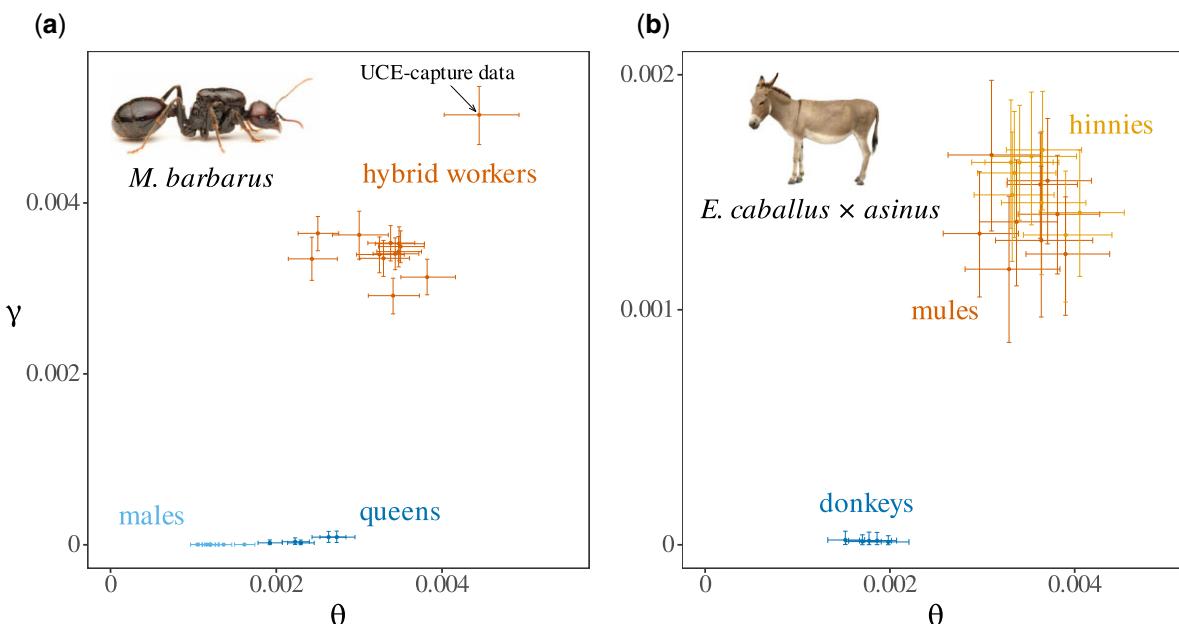


Fig. 3. Discrimination of F1 hybrids in transcriptomes of *M. barbarus* and *Equus*. Estimated values for the divergence parameter γ and the ancestral population mutation rate θ are represented for *M. barbarus* (a) and *Equus* (b). Colored points and lines represent point estimates and confidence intervals, respectively. Values obtained using UCE-capture data for a single worker of *M. barbarus* (genbank:SRR5437981) were added for comparison (arrow in panel a).

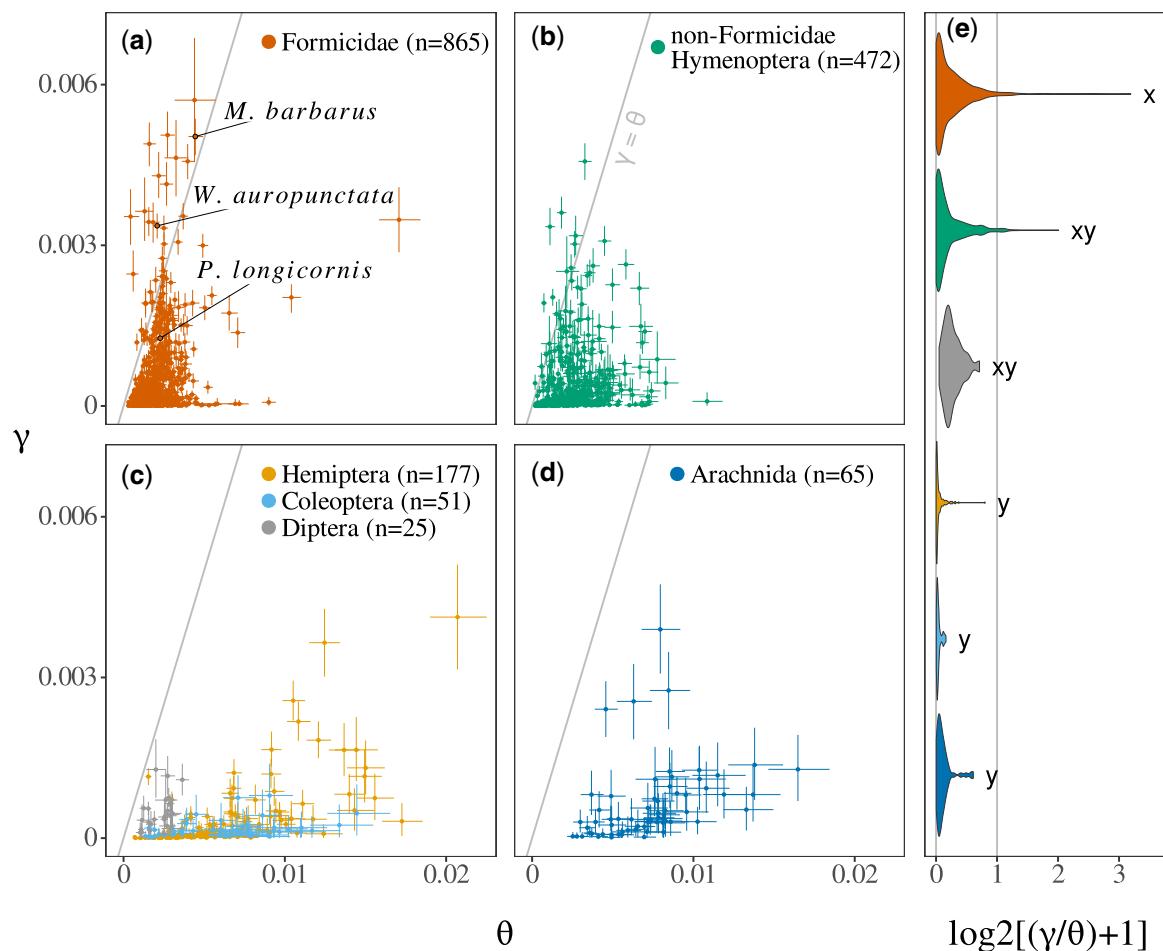


Fig. 4. Genomic scans for hybridization in six groups of arthropods. Estimates of the divergence parameter γ and the ancestral population mutation rate θ are represented for Formicidae (a), NonFormicidae Hymenoptera (b), other insects (c), and Arachnida (d). Colored points and lines represent bayesian point estimates and credibility intervals (see main text), respectively. (e) Distribution of the ratio γ/θ in each group. A one-shifted log 2-scale, under which the critical value of $\gamma/\theta = 1$ is unchanged, was used for visual convenience. Letters summarize the result of a post-hoc Tukey honest significance test, carried out using the *HSD.test* function of the R package *agricolae*. Groups with no letters in common have significantly different means (with $\alpha = 0.05$). All results were obtained using only dated and recent samples (see main text).

overestimation of γ as mentioned in the previous section. To take this effect into account, we excluded samples collected before 1980 and specimen with unknown collection date from subsequent analyses. This does not eliminate the mentioned correlation which remains significant ($\rho = -0.163$, p -value = 3.43×10^{-11}), but ensures that no old, highly degraded sample is wrongly interpreted as a F1 hybrid. This choice of a threshold date does not affect our subsequent statistical results (see [supplementary table S4, Supplementary Material online](#)). We also discarded samples for which less than 200 UCE loci could be retrieved to ensure sufficient statistical power. After application of these filters, we could obtain parameter estimates (fig. 4) for 850 Formicidae (223 represented genera), 472 other Hymenoptera (288 genera), 177 Hemiptera (121 genera), 51 Coleoptera (45 genera), 25 Diptera (5 genera), and 65 Arachnida (56 genera). All parameter estimates can be found in [supplementary table S3, Supplementary Material online](#).

Our results revealed important differences between phylogenetic groups regarding the prevalence of F1

hybrids. We found several candidate F1 hybrids ($\gamma/\theta > 1$) in Formicidae (29 candidates; fig. 4a) and other Hymenoptera (15 candidates; fig. 4b), while none were found in Hemiptera, Coleoptera, Diptera (fig. 4c), or Arachnida (fig. 4d). This result cannot be explained by the larger number of Hymenoptera available, as under the observed frequency of candidates in this group (0.033), the probability to observe no candidates in other groups would be 8.36×10^{-5} . Species names, divergence estimates and metadata for all candidate F1 hybrids can be found in [table 1](#). In Formicidae, two samples originating from species known to produce F1 hybrid workers (*M. barbarus* and *Wasmannia auropunctata*) were identified as candidate F1 hybrids, while a third (*Paratrechina longicornis*) was found to fall below the required value of $\gamma/\theta > 1$. Beyond individual candidates, Formicidae also displayed a significantly higher mean γ/θ ratio than nonHymenoptera insects, as evidence by a post-hoc Tukey honest significance test (fig. 4e). This suggests that, on average, successful interspecific mating is more frequent in ants than in other groups. Finally, candidates F1 hybrids displayed

Table 1. Candidate F1 Hybrids.

| | Family/Subfamily | Species | Collection | Origin | γ | γ/θ |
|---------------------------|------------------|-----------------------------------|------------|--------------|----------|-----------------|
| NonFormicidae Hymenoptera | Crabronidae | <i>Sphexius hogardii</i> | 2010 | unknown | 0.0033 | 3.0241 |
| | Cephidae | <i>Hartigia trimaculata</i> | 2013 | unknown | 0.0019 | 2.6289 |
| | Braconidae | <i>Pentatermus striatus</i> | 2016 | Thailand | 0.0004 | 2.3057 |
| | Crabronidae | <i>Microbembex cubana</i> | 2011 | unknown | 0.0036 | 1.9806 |
| | Sierolomorphidae | <i>Sierolomorpha sp.</i> | 2006 | unknown | 0.002 | 1.5881 |
| | Crabronidae | <i>Oxybelus analis</i> | 2011 | unknown | 0.0046 | 1.3963 |
| | Braconidae | <i>Macrostomion sumatranum</i> | 1999 | Japan | 0.0007 | 1.2737 |
| | Argidae | <i>Arge humeralis</i> | 2013 | unknown | 0.0032 | 1.1781 |
| | Braconidae | <i>Xenolobus sp.</i> | 2009 | Malawi | 0.0025 | 1.1694 |
| | Crabronidae | <i>Cerceris hatuey</i> | 2011 | unknown | 0.003 | 1.1464 |
| | Argidae | <i>Atomacera decepta</i> | 2013 | unknown | 0.0017 | 1.1371 |
| | Braconidae | <i>Cystomastax sp.</i> | 1989 | Costa Rica | 0.001 | 1.1018 |
| | Apidae | <i>Neolarva californica</i> | 2005 | Mexico | 0.0012 | 1.0878 |
| | Dryinidae | <i>Deinodryinus atriventris</i> | 2013 | unknown | 0.0026 | 1.0677 |
| | Braconidae | <i>Aleiodes coronopus</i> | 2003 | Thailand | 0.0017 | 1.0152 |
| Formicidae | Formicinae | <i>Paratrechina zanjonensis</i> | 2011 | Tanzania | 0.0035 | 8.1506 |
| | Myrmicinae | <i>Lachnomyrmex scrobiculatus</i> | 2008 | Guatemala | 0.0025 | 4.1156 |
| | Formicinae | <i>Agraulomyrmex sp.</i> | 2008 | Tanzania | 0.0049 | 3.0644 |
| | Myrmicinae | <i>Cyphomyrmex sp.</i> | 1992 | Brazil | 0.0036 | 2.7854 |
| | Myrmicinae | <i>Mycetagoicus triangularis</i> | 1992 | Brazil | 0.0034 | 2.1968 |
| | Formicinae | <i>Myrmecocystus creightoni</i> | 1997 | USA | 0.0043 | 1.9746 |
| | Formicinae | <i>Santschiella kohli</i> | 2000 | Gabon | 0.0034 | 1.8713 |
| | Dorylinae | <i>Aenictus hoelldobleri</i> | 2013 | China | 0.0051 | 1.855 |
| | Myrmicinae | <i>Wasmannia auropunctata</i> | 2001 | Cuba | 0.0034 | 1.6147 |
| | Formicinae | <i>Myrmecocystus cf. navajo</i> | 2003 | Mexico | 0.0041 | 1.555 |
| | Myrmicinae | <i>Ochetomyrmex sp.</i> | 2002 | Guyana | 0.0012 | 1.4361 |
| | Formicinae | <i>Brachymyrmex sp.</i> | 2009 | Brazil | 0.0046 | 1.4243 |
| | Dorylinae | <i>Simopone marleyi</i> | 1986 | South Africa | 0.0019 | 1.4099 |
| | Myrmicinae | <i>Tranopelta gilva</i> | 2006 | Costa Rica | 0.0019 | 1.3953 |
| | Dorylinae | <i>Sphinctomyrmex stali</i> | 2013 | Brazil | 0.0033 | 1.3282 |
| | Formicinae | <i>Polyrhachis hector</i> | 2010 | Indonesia | 0.0057 | 1.2988 |
| | Formicinae | <i>Teratomyrmex greavesi</i> | 2007 | Australia | 0.0021 | 1.2818 |
| | Formicinae | <i>Bajcaridris theryi</i> | 2010 | Morocco | 0.0014 | 1.2765 |
| | Dorylinae | <i>Ectiton mexicanum</i> | 2013 | Costa Rica | 0.0014 | 1.2643 |
| | Formicinae | <i>Polyrhachis mellita</i> | 2008 | Indonesia | 0.003 | 1.2046 |
| | Formicinae | <i>Myrmecocystus cf. mendax</i> | 2014 | USA | 0.0023 | 1.1768 |
| | Ponerinae | <i>Ponera coarctata</i> | 1990 | Italy | 0.0046 | 1.1492 |
| | Myrmicinae | <i>Kalathomyrmex emeryi</i> | 2012 | Brazil | 0.0028 | 1.1466 |
| | Myrmicinae | <i>Messor barbarus</i> | 2008 | Spain | 0.005 | 1.1311 |
| | Myrmicinae | <i>Cyphomyrmex costatus</i> | 1996 | Panama | 0.0019 | 1.0993 |
| | Myrmicinae | <i>Blepharidatta brasiliensis</i> | 2000 | Brazil | 0.0019 | 1.0743 |
| | Formicinae | <i>Polyrhachis taylori</i> | 2008 | Papua NG | 0.0013 | 1.0721 |
| | Formicinae | <i>Lepisiota sp.</i> | 2011 | South Africa | 0.0024 | 1.0187 |
| | Formicinae | <i>Acropyga stenotes</i> | 2002 | Guyana | 0.0025 | 1.0181 |

NOTE.—The table gives metadata and point parameter estimates for each candidate F1 hybrid (i.e., $\gamma/\theta > 1$) in our analysis.

higher average divergence γ in Formicidae than in other Hymenoptera ($T = 2.24$, p -value = 0.0324), suggesting that hybridization events in this group tend to occur between more divergent individuals. Note that these two last results are mostly unchanged under other reasonable choices of threshold dates (see table ??).

Samples for roughly two-thirds (223 represented genera) of the diversity of ants (about 300 genera) genera were available for this study. This allowed us to evaluate whether the distribution of hybridization within ants genera is random. Positioning candidate F1 hybrids on a phylogeny of ants genera (fig. 5) and the application of Abouheif's test (Abouheif 1999) revealed a significant positive phylogenetic correlation in mean γ/θ across genera ($C = 0.1758$; p -value = 0.007). This can be explained

by the absence of candidate F1 hybrids from widely sampled groups, such as the Crematogastrini tribe (171 species from 59 genera), and by their high prevalence in other groups, such as the Attini tribe (10 candidates representing 9.4% of the tribe's sampled species). Genera *Cyphomyrmex*, *Polyrhachis*, and *Myrmecocystus* also displayed several distinct candidate F1 hybrids.

Discussion

F1 Hybrid Detection from Single Genomes

In this article, we implement and showcase a fast and flexible statistical method for F1 hybrids detection. This method only relies on the distribution of heterozygosity across a

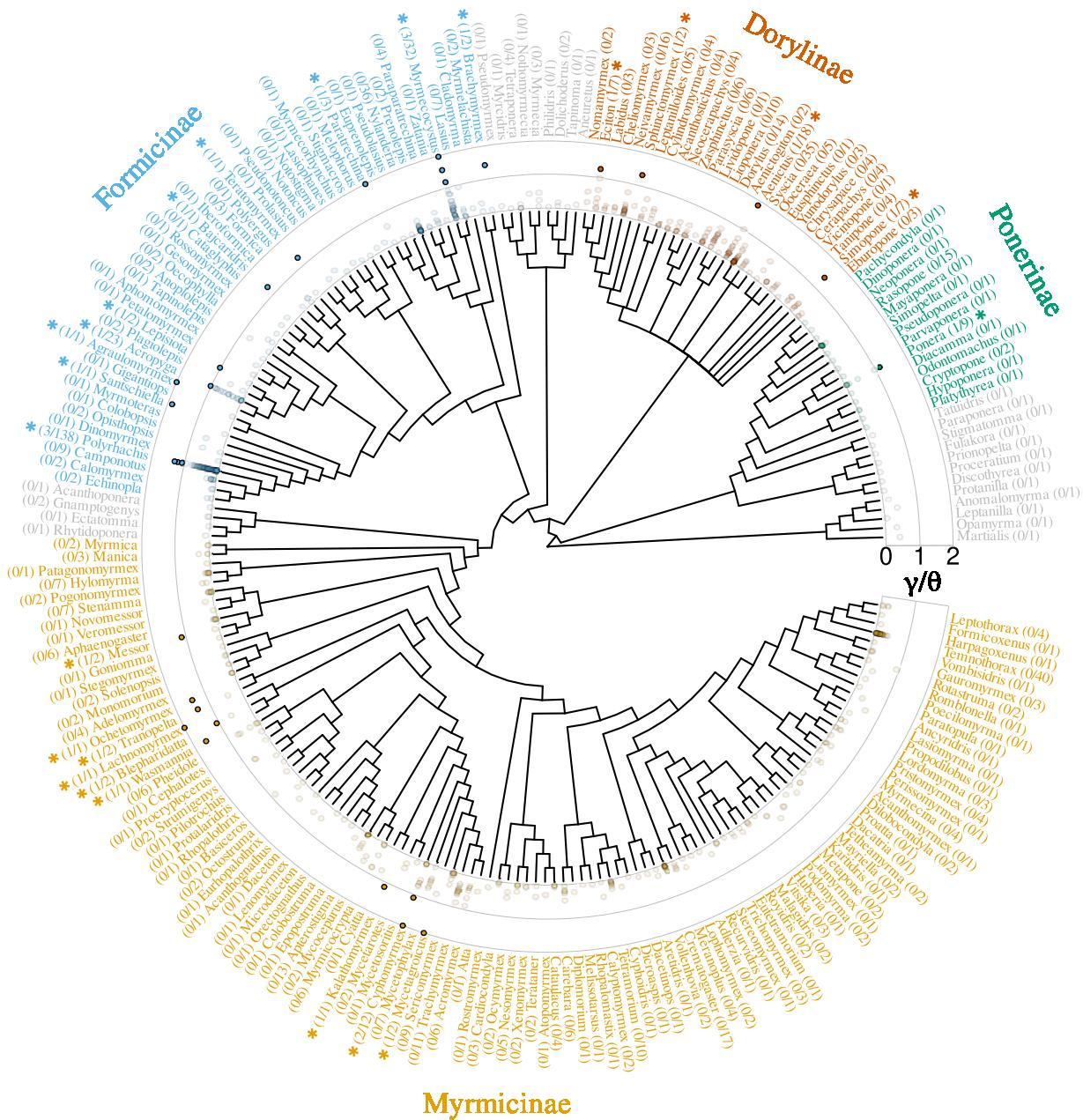


Fig. 5. Occurrence of F1 hybrids across genera of Formicidae. Estimates of the ratio γ/θ obtained in Formicidae are represented against the topology of genera in this group (retrieved from Antwiki). Genera counting at least one species with $\gamma/\theta > 1$ (i.e., probable F1 hybrid) are highlighted by a star. The number of such species per genera, as well as the total number of species per genera are given for each genus. γ/θ ratios higher than two were truncated to two for readability. Three genera with no candidate F1 hybrids (*Cryptopone*, *Pseudoatta*, and *Strongylognathus*), present in UCE capture data but not in the present topology, were not integrated in this representation or in statistical test for phylogenetic correlation.

set of diploid loci, and is thus theoretically applicable to any type of polymorphic loci set such as UCE loci, coding genes, or even RAD tags. Note however that chosen loci should ideally be 1-1 orthologs, in order to limit the risk of paralogy inflating observed substitutions counts and facilitate intra-group estimates comparisons. Besides its applicability to a large range of data types, the method is also flexible in that it does not rely on the use of parental genomes, unlike population-centered hybrid detection approaches ([Anderson and Thompson 2002](#); [Payseur and](#)

Rieseberg 2016; Schubert et al. 2017). It is thus especially suited for preliminary hybrid status assessment in single-species datasets composed of many nonmodel species (i.e., most phylogenomic datasets). In addition to applications in the study of hybridization prevalence across taxa (i.e., as done in this study), F1 hybrid detection could help preventing the use of error-inducing hybrids nuclear genomes in reconstructing species trees (McDade 1992).

Naturally, the presented method also has some shortcomings, which stem in the limited statistical power

provided by single-individual genomes. Perhaps the most important limitation of the method is that it is restricted to the discrimination of F1 hybrids, and cannot reliably be used to identify backcross hybrids. This restricts the use of the method to the study of present and recent hybridization and suggests that many hybrids can be missed, given the rarity of true F1 hybrids in natural populations. We have also shown that statistical error inherent to data treatment can inflate divergence estimates and lead to false identification of F1 hybrids. This is because our method relies on the assumption that divergence is characterized by a uniform increase in heterozygosity across loci. As sequencing, assembly and gene identification errors are likely to produce such an increase, their effect is mostly indistinguishable from true divergence using single genomes. This limits the application of our method to samples of good quality and limits its ability to identify F1 hybrids with low overall polymorphism. The sensitivity of the method is also hindered by any violation of the hypothesis of constant mutation rate in time and across loci. In fact, Yang (1997) has shown that variation in mutation rates generally reduces estimates of divergence by eroding any uniform component of heterozygosity. The limited sensitivity of the method might be problematic in other settings, but acts as a safeguard in our case by making F1 hybrids detection more conservative.

High Prevalence of F1 Hybrids in Hymenoptera and Particularly in Ants

F1 hybrids detection in 850 Formicidae, 472 nonFormicidae Hymenoptera, 177 Hemiptera, 51 Coleoptera, 25 Diptera, and 65 Arachnida revealed a heterogeneous distribution of F1 hybrids prevalence across these groups. We identified 29 and 15 candidate F1 hybrids in Formicidae and other Hymenoptera, respectively, and none in other groups, a result that cannot be explained by uneven group sampling. High hybridization rates in Hymenoptera have been predicted by other authors (Nonacs 2006; Feldhaar et al. 2008) under the rationale that haplodiploidy could mitigate the potential costs of out-breeding. More specifically, it was proposed that because haplodiploid females produce part of their descendants asexually, they should retain positive fitness even when engaging in nonviable interspecific mating, leading to a weaker long-term selection against such behavior. While our results are compatible with this hypothesis, similar analyses on haplodiploid groups other than Hymenoptera will be necessary to confirm that haplodiploidy is the only factor explaining this pattern. On a more general note, it is important to underline the fact that an absence of candidate F1 hybrids in nonHymenoptera does not mean that hybridization is absent in these groups. Instead, it suggests either that hybridization is generally less likely (i.e., rare enough to be undetected with our low sensitivity method), or that it more often leads to introgression and fewer F1 hybrids.

Within Hymenoptera, our analyses also revealed a significantly higher prevalence of F1 hybrids in Formicidae than in other Hymenoptera. High hybridization rates were previously described in several ant genera (e.g., in some North American *Solenopsis* or European *Temnothorax*, Feldhaar et al. 2008), and have been suspected to be frequent in ants in general on the basis of several arguments. Some authors have hypothesized that hybrid sterility has a minimal fitness cost in eusocial species because they produce a large majority of normally sterile individuals (i.e., workers), leading to weaker selection against hybridization (Nonacs 2006; Umphrey 2006). The same authors also proposed that eusocial queens could use interspecific mating as a “best of a bad situation” strategy allowing for the production of a workforce and the successful rearing of haploid sons in the absence of conspecific mates (e.g., in locally rare species). Such strategy, sometimes referred to as “sperm parasitism,” would be especially likely to arise if hybrid workers outperform regular ones, a hypothesis that received some empirical support in the *Pogonomyrmex* genus (James et al. 2002; Helms Cahan et al. 2010, but see Ross and Robertson 1990; Julian and Helms Cahan 2006; Feldhaar et al. 2008). Interestingly, the idea that eusociality facilitates or promotes hybridization is not clearly supported by the present analysis, as no candidate F1 hybrids were identified amongst 66 available nonFormicidae eusocial species (22 represented genera). While this might be because most of these species display relatively simple forms of eusociality as compared to ants (with 44 species belonging to either *Lasioglossum* or *Bombus*), it could also indicate that ants possess other traits relevant to frequent hybridization. Among characteristics unique to ants, the extreme functional simplification of workers (Peeters and Ito 2015) could have favored hybridization by making hybrid individuals less affected by inherent developmental defects (e.g., fluctuating asymmetry). Additionally, the typically low morphological and behavioral divergence observed between males of related ant species has been proposed to reduce pre-mating barriers to hybridization in this group (Feldhaar et al. 2008).

Phylogenetic and Ecological Characteristics of F1 Hybrids in Ants

Beyond the higher prevalence of candidate F1 hybrids in ants, our analysis reveals that their phylogenetic distribution in the group follows a nonrandom pattern, hinting towards a potential connection with variation in ecological and life-history characteristics of species. One peculiar characteristic of some ants that is especially relevant to our findings is their display of hybridization-dependent reproductive systems. In these systems, strong genetic caste determination enforces that all workers are F1 hybrids developing from eggs fertilized by allo-specific males (i.e., social hybridogenesis, as in *Messor*, *Pogonomyrmex*, *Solenopsis*, or *Cataglyphis*; Helms Cahan et al. 2002; Helms Cahan and Vinson 2003; Anderson et al. 2006;

Romiguier et al. 2017; Lacy et al. 2019; Kuhn et al. 2020) or by males from a divergent lineage of the same species (i.e., as in *W. auropunctata*, *Vollenhovia emeyri*, or *P. longicornis*; Fournier et al. 2005; Ohkawara et al. 2006; Pearcy et al. 2011), while queens are produced through regular intra-lineage mating or thelytokous parthenogenesis. In genera where it has been described, strong genetic caste determination has typically evolved independently multiple times (Anderson et al. 2006; Romiguier et al. 2017; Kuhn et al. 2020), indicating that phylogenetic correlation in this trait is expected. Furthermore, out of the three available species known to display such system, two clearly stand out as F1 hybrids (*M. barbarus* and *W. auropunctata*), indicating that our method is in some cases able to detect the divergence signal present in individual genomes of their workers. While this result was expected in *M. barbarus*, where the divergence between hybridizing lineages is known to be high (Romiguier et al. 2017), it was more surprising in *W. auropunctata*. In this species, the divergence between male and female lineages, which are thought to originate from the same ancestral population (Fournier et al. 2005), is expected to be much lower (i.e., more similar to what is observed for *P. longicornis*). This might suggest that the isolation between males and females of this species is more ancient than previously thought, or that divergence has quickly accumulated. The extent to which our method can reliably detect reproductive systems such as that of *W. auropunctata* and *P. longicornis* is still largely unknown. A better understanding will require more in-depth population genetic analyses and independent estimates of divergence between lineages. Besides remaining uncertainties, the detection of *M. barbarus* and *W. auropunctata* as F1 hybrids suggests that other candidates identified in this work might belong to species with similar reproductive systems, which would help explain why we detected a larger proportion of F1 hybrids in ants. This possibility echoes the prediction of some authors that the prevalence of strong genetic caste determination in Formicidae might have been largely underestimated (Anderson et al. 2008).

If some detected candidates correspond to unknown cases of strong genetic caste determination, our results might help shed new light on the conditions that drive the evolution of such systems. For instance, it has been hypothesized that genetic caste determination evolves more frequently in taxa with a highly specialized diet (such as granivory), as a reduced dietary spectrum would impede the use of differential larval feeding as a mean to drive caste determination (Romiguier et al. 2017). Interestingly, we found significantly higher γ/θ ratios in genera listed as strictly herbivorous (fungus-growing, granivorous, or specialized aphid-rearing diets; Blanchard and Moreau 2016) than in omnivorous or carnivorous genera (one-sided Welsch t -test; $t = 3.3292$, $df = 154.75$, p -value = 0.00054). This may suggest that highly specialized diets do favor the evolution of genetic caste determination. This remains highly speculative however without an extended study on more genera and clear

confirmation that γ/θ variations are mainly due to unusual reproductive systems across ants. While the exact proportion of detected F1 hybrids that are due to such reproductive systems is unknown at this stage, species with γ/θ ratios superior to known cases (*M. barbarus*, *W. auropunctata*, *P. longicornis*, see fig. 4) would be good first candidates for future studies.

Besides unusual reproductive systems, high hybridization rates in Dorylinae and in Attini could be linked to the unusually high polyandry observed in these group (Keller and Reeve 1994; Strassmann 2001). Queens that mate multiply are less likely to mate only with interspecific males (Umphrey 2006), and are therefore expected to display lower pre-mating barriers to hybridization. Such effect of polyandry is especially likely when both types of males are easily accessible, as in species with massive mating flights that are synchronized with other sympatric species. Such pattern is more frequent in species inhabiting temperate and arid climates, where mating flights are often triggered by heavy rainfall (Dunn et al. 2007). In favor of such connection, we find that the previously unsuspected xerophile genus *Myrmecocystus* counts several candidate F1 hybrids. As a final remark, we note that some ant groups display a high proportion of candidate F1 hybrids, while presenting no obvious life-history or ecological features likely to produce such pattern. This is especially true of the paraphyletic group of attines composed of *Ochetomyrmex*, *Tranopelta*, *Lachnomyrmex*, *Blepharidatta*, and *Wasmannia*. This suggests the existence of other unknown factors in species predisposition to hybridization, and new biological models for the study of such factors.

Conclusion

Hybridization is a widespread and fundamental phenomenon that carries implications for many central processes of biological evolution, including speciation and adaptation. Here we present the first large-scale comparative study of F1 hybrids prevalence in Arthropods, analyzing genomic data for more than 1,500 nonmodel species obtained from public repositories. We report high rates of recent hybridization in Hymenoptera, and especially in ants, confirming previous predictions found in the literature. We also find the prevalence of F1 hybrids to be heterogeneously distributed within ants, with probable links with ecological and life-history features. These results were produced through the implementation of a scalable F1 hybrids detection method, which is applicable to virtually any modern sequencing data. Further applications of this method should help better assessing the frequency of hybridization across the tree of life, and understanding its determinants.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution online*.

Acknowledgments

We thank Hugo Darras and Bernhard Seifert for useful comments. We thank the European Research Council (ERC-2020-StG-948688 RoyalMess project). The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

Author Contributions

A.W. and J.R. conceived the study. A.W. and N.G. developed statistical methods. L.B. and J.R. preformed preliminary analyses. A.W. performed the final analysis and wrote the first draft of the manuscript under the guidance of J.R. and N.G.. All authors contributed to the final version.

Data Availability

Supplementary tables containing all results produced in this work, as well as scripts and files necessary to apply our statistical procedure, are available here: <https://zenodo.org/record/5415947>.

References

- Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, et al. 2013. Hybridization and speciation. *J Evol Biol.* **26**(2):229–246.
- Abouheif E. 1999. A method for testing the assumption of phylogenetic independence in comparative data. *Evol Ecol Res.* **1**(8):895–909.
- Anderson E. 1953. Introgressive hybridization. *Biol Rev.* **28**(3):280–307.
- Anderson EC, Thompson EA. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics.* **160**(3):1217–1229.
- Anderson KE, Gadau J, Mott BM, Johnson RA, Altamirano A, Strehl C, Fewell JH. 2006. Distribution and evolution of genetic caste determination in *Pogonomyrmex* seed-harvester ants. *Ecology.* **87**(9):2171–2184.
- Anderson KE, Linksvayer TA, Smith CR. 2008. The causes and consequences of genetic caste determination in ants (Hymenoptera: Formicidae). *Myrmecol News.* **11**:119–132.
- Barton NH. 2001. The role of hybridization in evolution. *Mol Ecol.* **10**(3):551–568.
- Barton NH, Hewitt GM. 1985. Analysis of hybrid zones. *Annu Rev Ecol Syst.* **16**:113–148.
- Blanchard BD, Moreau CS. 2016. Defensive traits exhibit an evolutionary trade-off and drive diversification in ants. *Evolution (NY).* **71**(2):315–328.
- Bordenstein SR, O'hara FP, Werren JH. 2001. Wolbachia-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature.* **409**(6821):707–710.
- Bossert S, Murray EA, Almeida EA, Brady SG, Blaimer BB, Danforth BN. 2019. Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae. *Mol Phylogenet Evol.* **130**(July 2018):121–131.
- Branstetter MG, Longino JT, Ward PS, Faircloth BC. 2017. Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. *Methods Ecol Evol.* **8**(6):768–776.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* **34**(17):i884–i890.
- Cordonnier M, Escarguel G, Dumet A, Kaufmann B. 2020. Multiple mating in the context of interspecific hybridization between two Tetramorium ant species. *Heredity (Edinb).* **124**(5):675–684.
- Dobzhansky T. 1940. Speciation as a stage in evolutionary divergence. *Am Nat.* **74**(753):312–321.
- Dunn RR, Parker CR, Geraghty M, Sanders NJ. 2007. Reproductive phenologies in a diverse temperate ant fauna. *Ecol Entomol.* **32**(2):135–142.
- Edmands S. 2002. Does parental divergence predict reproductive compatibility?. *Trends Ecol Evol.* **17**(11):520–527.
- Faircloth BC. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics.* **32**(5):786–788.
- Faircloth BC. 2017. Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods Ecol Evol.* **8**(9):1103–1112.
- Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst Biol.* **61**(5):717–726.
- Feldhaar H, Foitzik S, Heinze J. 2008. Lifelong commitment to the wrong partner: hybridization in ants. *Philos Trans R Soc B Biol Sci.* **363**(1505):2891–2899.
- Fishman L, Stathos A, Beardsley PM, Williams CF, Hill JP. 2013. Chromosomal rearrangements and the genetics of reproductive barriers in *mimulus* (monkey flowers). *Evolution (NY).* **67**(9):2547–2560.
- Fournier D, Estoup A, Orivel J, Foucaud J, Jourdan H, Le Breton J, Keller L. 2005. Clonal reproduction by males and females in the little fire ant. *Nature.* **435**(7046):1230–1234.
- Hamilton JA, Miller JM. 2016. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conserv Biol.* **30**(1):33–41.
- Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of species boundaries. *J Hered.* **105**(S1):795–809.
- Helms Cahan S, Daly A, Schwander T, Woods H. 2010. Genetic cast determination does not impose growth rate costs in *Pogonomyrmex* harvester ants. *Funct Ecol.* **24**:301–309.
- Helms Cahan S, Parker JD, Rissing SW, Johnson RA, Polony TS, Weiser MD, Smith DR. 2002. Extreme genetic differences between queens and workers in hybridizing *Pogonomyrmex* harvester ants. *Proc R Soc B Biol Sci.* **269**(1503):1871–1877.
- Helms Cahan S, Vinson SB. 2003. Reproductive division of labor between hybrid and nonhybrid offspring in a fire ant hybrid zone. *Evolution (NY).* **57**(7):1562–1570.
- Huang W, Li L, Myers JR, Marth GT. 2012. ART: a next-generation sequencing read simulator. *Bioinformatics.* **28**(4):593–594.
- Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics.* **18**(2):337–338.
- James SS, Pereira RM, Vail KM, Ownley BH. 2002. Survival of imported fire Ant (Hymenoptera: Formicidae) species subjected to freezing and near-freezing temperatures. *Environ Entomol.* **31**(1):127–133.
- Julian GE, Helms Cahan S. 2006. Behavioral differences between *Pogonomyrmex rugosus* and dependent lineages (H1/H2) harvester ants. *Ecology.* **87**(9):2207–2214.
- Keller L, Reeve HK. 1994. Genetic variability, queen number, and polyandry in social Hymenoptera. *Evolution (NY).* **48**(3):694–704.
- Kieran TJ, Gordon ER, Forthman M, Hoey-Chamberlain R, Kimball RT, Faircloth BC, Weirauch C, Glenn TC. 2018. Insight from an ultraconserved element bait set designed for hemipteran phylogenetics integrated with genomic resources. *Mol Phylogenet Evol.* **130**:297–303.
- Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of Next Generation Sequencing Data. *BMC Bioinform.* **15**(1):1–13.
- Kuhn A, Darras H, Paknia O, Aron S. 2020. Repeated evolution of queen parthenogenesis and social hybridogenesis in *Cataglyphis* desert ants. *Mol Ecol.* **29**(May 2019):549–564.

- Lacy KD, Shoemaker D, Ross KG. 2019. Joint evolution of asexuality and queen number in an ant. *Curr Biol*. **29**(8):1394–1400.e4.
- Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. **31**(10): 1674–1676.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**(14):1754–1760.
- Mallet J. 2005. Hybridization as an invasion of the genome. *Trends Ecol Evol*. **20**(5):229–237.
- Mayr E. 1942. *Systematics and the origin of species*. New York: Columbia University Press.
- Mayr E. 1963. *Animal species and evolution*. Cambridge (MA): Harvard University Press.
- McDade LA. 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. *Evolution (NY)*. **46**(5): 1329–1346.
- Miles Zhang Y, Williams JL, Lucky A. 2019. Understanding UCEs: a comprehensive primer on using ultraconserved elements for arthropod phylogenomics. *Insect Syst Divers*. **3**(5):1–12.
- Nonacs P. 2006. Interspecific hybridization in ants: at the intersection of ecology, evolution, and behavior. *Ecology*. **87**(9): 2143–2147.
- Ohkawara K, Nakayama M, Satoh A, Trindl A, Heinze J. 2006. Clonal reproduction and genetic caste differences in a queen-polymorphic ant, *Vollenhovia emeryi*. *Biol Lett*. **2**(3):359–363.
- Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation. *Mol Ecol*. **25**(11):2337–2360.
- Pearcy M, Goodisman MAD, Keller L. 2011. Sib mating without inbreeding in the longhorn crazy ant. *Proc R Soc B: Biol Sci*. **278** (1718):2677–2681.
- Peeters C, Ito F. 2015. Wingless and dwarf workers underlie the ecological success of ants (Hymenoptera: Formicidae). *Myrmecol News*. **21**:117–130.
- Pfennig KS. 2007. Facultative mate choice drives adaptive hybridization. *Science (80-)*. **318**(5852):965–967.
- Prentis PJ, Wilson JR, Dormontt EE, Richardson DM, Lowe AJ. 2008. Adaptive evolution in invasive species. *Trends Plant Sci*. **13**(6): 288–294.
- Price TD, Bouvier MM. 2002. The evolution of F1 postzygotic incompatibilities in birds. *Evolution (NY)*. **56**(10):2083–2089.
- Rambaut A, Grassly NC. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics*. **13**(3):235–238.
- Randler C. 2002. Avian hybridization, mixed pairing and female choice. *Anim Behav*. **63**(1):103–119.
- Roberts HF. 1919. Darwin's contribution to the knowledge of hybridization. *Am Nat*. **53**(629):535–554.
- Romiguier J, Fournier A, Yek SH, Keller L. 2017. Convergent evolution of social hybridogenesis in *Messor* harvester ants. *Mol Ecol*. **26**: 1108–1117.
- Ross KG, Robertson JL. 1990. Developmental stability, heterozygosity, and fitness in two introduced fire ants (*Solenopsis invicta* and *S. richteri*) and their hybrid. *Heredity (Edinb)*. **64**:93–103.
- Schubert M, Mashkour M, Gaunitz C, Fages A, Seguin-Orlando A, Sheikhi S, Alfarhan AH, Alquraishi SA, Al-Rasheid KA, Chuang R, et al. 2017. Zonkey: a simple, accurate and sensitive pipeline to genetically identify equine F1-hybrids in archaeological assemblages. *J Archaeol Sci*. **78**:147–157.
- Smadja CM, Butlin RK. 2011. A framework for comparing processes of speciation in the presence of gene flow. *Mol Ecol*. **20**(24): 5123–5140.
- Stan Development Team. 2019. Stan Modeling Language Users Guide and Reference Manual, 2.17.
- Stan Development Team. 2020. RStan: the R interface to Stan. R package version 2.21.2.
- Starrett J, Derkarabetian S, Hedin M, Bryson RW, McCormack JE, Faircloth BC. 2016. High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. *Mol Ecol Resour*. **17**(4):812–823.
- Strassmann J. 2001. The rarity of multiple mating by females in the social Hymenoptera. *Insectes Soc*. **48**:1–13.
- Takahata N, Satta Y, Klein J. 1995. Divergence time and population size in the lineage leading to modern humans. *Theor Popul Biol*. **48**:198–221.
- Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat Ecol Evol*. **3**(2):170–177.
- Umphrey GJ. 2006. Sperm parasitism in ants: selection for interspecific mating and hybridization. *Ecology*. **87**(9):2148–2159.
- Wakeley J. 2008. *Coalescent theory: an introduction*. Greenwood Village (CO): Roberts and Company Publishers. p. 230.
- Wang Y, Gao S, Zhao Y, Chen WH, Shao JJ, Wang NN, Li M, Zhou GX, Wang L, Shen WJ, et al. 2019. Allele-specific expression and alternative splicing in horse×donkey and cattle×yak hybrids. *Zool Res*. **40**(4):293–304.
- Yang Z. 1997. On the estimation of ancestral population sizes of modern humans. *Genet Res*. **69**(2):111–116.

Ant phylogenomics reveals a natural selection hotspot preceding the origin of complex eusociality

Highlights

- We sequenced 65 ant genomes representing the 17 ant subfamilies
- Leptanillomorphs are the sister clade of all other extant ant species
- A natural selection hotspot occurred during the emergence of formicoids
- Genomic foundations of complex eusociality may have evolved early in ant evolution

Authors

Jonathan Romiguier,
Marek L. Borowiec, Arthur Weyna, ...,
Brian L. Fisher, Philip S. Ward,
Laurent Keller

Correspondence

jonathan.romiguier@umontpellier.fr

In brief

Romiguier et al. sequence 65 genomes to produce a phylogenetic tree of all ant subfamilies, with the subterranean leptanillomorph clade as the sister group of all other ants. A natural selection hotspot is detected during the emergence of formicoid clade, which is an ant clade grouping the most complex forms of sociality.



Report

Ant phylogenomics reveals a natural selection hotspot preceding the origin of complex eusociality

Jonathan Romiguier,^{1,2,8,9,*} Marek L. Borowiec,³ Arthur Weyna,¹ Quentin Helleu,² Etienne Loire,⁴ Christine La Mendola,² Christian Rabeling,⁵ Brian L. Fisher,⁶ Philip S. Ward,⁷ and Laurent Keller²

¹Institut des Sciences de l'Evolution de Montpellier (ISEM), CNRS, IRD, EPHE, Université de Montpellier, Montpellier, France

²Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland

³Department of Entomology, Plant Pathology and Nematology, University of Idaho, Moscow, ID, USA

⁴ASTRE, Cirad, INRAE, University of Montpellier, Montpellier, France

⁵School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

⁶Department of Entomology, California Academy of Sciences, San Francisco, CA 94118, USA

⁷Department of Entomology and Nematology, University of California, Davis, Davis, CA 95616, USA

⁸Twitter: @SelfishMeme

⁹Lead contact

*Correspondence: jonathan.romiguier@umontpellier.fr

<https://doi.org/10.1016/j.cub.2022.05.001>

SUMMARY

The evolution of eusociality has allowed ants to become one of the most conspicuous and ecologically dominant groups of organisms in the world. A large majority of the current ~14,000 ant species belong to the formicoids,¹ a clade of nine subfamilies that exhibit the most extreme forms of reproductive division of labor, large colony size,² worker polymorphism,³ and extended queen longevity.⁴ The eight remaining non-formicoid subfamilies are less well studied, with few genomes having been sequenced so far and unclear phylogenetic relationships.⁵ By sequencing 65 genomes, we provide a robust phylogeny of the 17 ant subfamilies, retrieving high support to the controversial leptanillomorph clade (Leptanillinae and Martialinae) as the sister group to all other extant ants. Moreover, our genomic analyses revealed that the emergence of the formicoids was accompanied by an elevated number of positive selection events. Importantly, the top three gene functions under selection are linked to key features of complex eusociality, with histone acetylation being implicated in caste differentiation, gene silencing by RNA in worker sterility, and autophagy in longevity. These results show that the key pathways associated with eusociality have been under strong selection during the Cretaceous, suggesting that the molecular foundations of complex eusociality may have evolved rapidly in less than 20 Ma.

RESULTS AND DISCUSSION

A reference tree of ant subfamilies

To build a comprehensive phylogenetic tree including representatives of all extant ant subfamilies, we conducted two main types of analyses on the 4,300,911 amino acids from 4,151 single-copy protein-coding genes that we generated from 83 species (including 9 hymenopteran outgroups and 65 newly sequenced genomes; [STAR Methods](#)). First, we performed supermatrix approaches, where all the genes were concatenated for estimating a single species tree. Second, we performed supertree approaches, where the species trees were estimated from all gene trees. The two best resulting trees from each approach are summarized in [Figure 1](#). Each node of the tree is supported with maximal support by the supermatrix analyses, and nearly all nodes (78/82) are congruently supported by both the supermatrix and supertree approaches. Importantly, the deepest and most important nodes of the ant phylogeny, including every relationship among the 17 ant subfamilies are

maximally supported by both the supermatrix and supertree approaches.

Our results confirm that the so-called poneroid subfamilies (Ponerinae, Paraponerinae, Agroecomyrmecinae, Proceratinae, Apomyrminae, and Amblyoponinae) are monophyletic,^{6–9} rather than paraphyletic.^{10–12} Both the supermatrix and supertree provide maximum support for the poneroid monophyly. The relationships among poneroid subfamilies are also all congruently supported by both approaches ([Figure 1](#)), including for the Paraponerinae (one living species) and Agroecomyrmecinae (two living species), which are inherently difficult to relate with other subfamilies as their deep phylogenetic divergence result in long branches. An analysis of ultraconserved elements (UCE) markers for an increased dataset of 166 taxa ([STAR Methods](#)) retrieved the same subfamily relationships ([Figure S1](#)), further supporting the view that our phylogeny ([Figure 1](#)) is robust.

These analyses are important because the rooting of the ant phylogeny has been a controversial issue since the discovery of *Martialis heureka*, an extremely rare and morphologically



FORMICOIDS
PONEROIDS

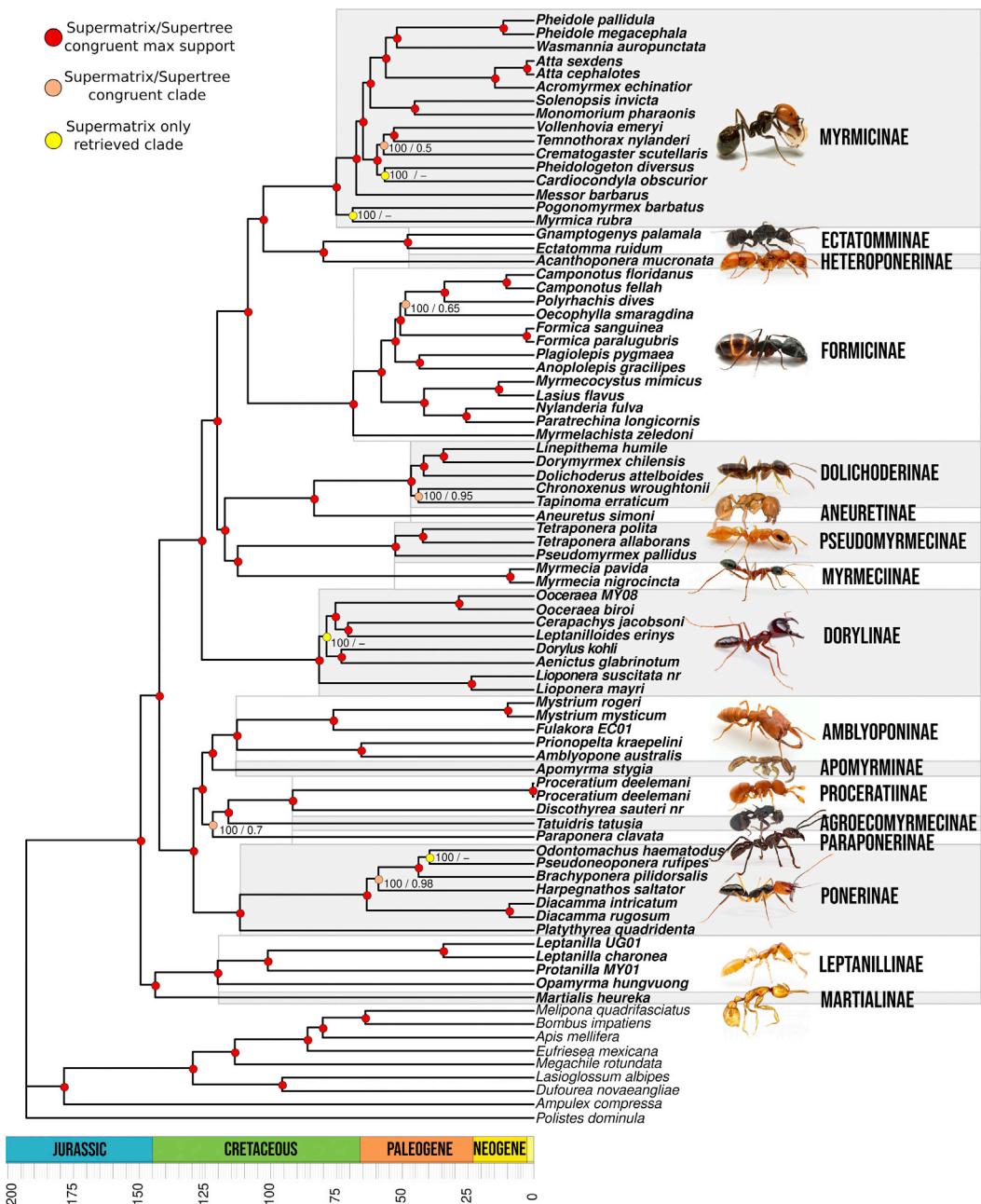


Figure 1. Phylogeny and timeline of ant evolution based on whole-genome data

The topology is inferred according to the main supermatrix analysis (1,692,052 amino acid sites after cleaning; maximum likelihood; IQ-TREE PMSF C20 profile mixture model; Figure S1D); ultrafast bootstrap support is displayed first when node support is not maximal. Node support of the main supertree analyses (from gene trees of the 1,552 alignments of more than 500 amino acid sites; gene trees inferred with model search in IQ-TREE, and species tree with ASTRAL; Figure S1E) is displayed second when not maximal (dash when node shows incongruence with the supermatrix analysis). Time divergence has been estimated using chronos with 12 calibration nodes (Table S2). Ant images from Alex Wild; used with permission. See also Figure S1 and Tables S1 and S2.

divergent ant species that was initially inferred as the sister group of all other ants.¹⁰ Some studies suggested that Leptanillinae is the subfamily sister to all other ants,^{11,13} whereas a recent study suggested that Leptanillinae and Martialinae may form a monophyletic group.⁹ Our analyses support this last hypothesis, with Leptanillinae + Martialinae forming a clade (hereafter referred to as the leptanillomorph clade) that is the sister group to all other

ants (Figure 1). This conclusion was also supported by two further analyses controlling for outgroup composition, which has been suggested to affect the rooting of the ant tree.⁹ First, we built an alternative supermatrix of 2,343 genes (983,951 amino acids) containing many outgroups, with 115 non-ant aculeate species borrowed from published transcriptomes.¹⁴ These analyses revealed that the rooting of the ant phylogeny and subfamily relationships

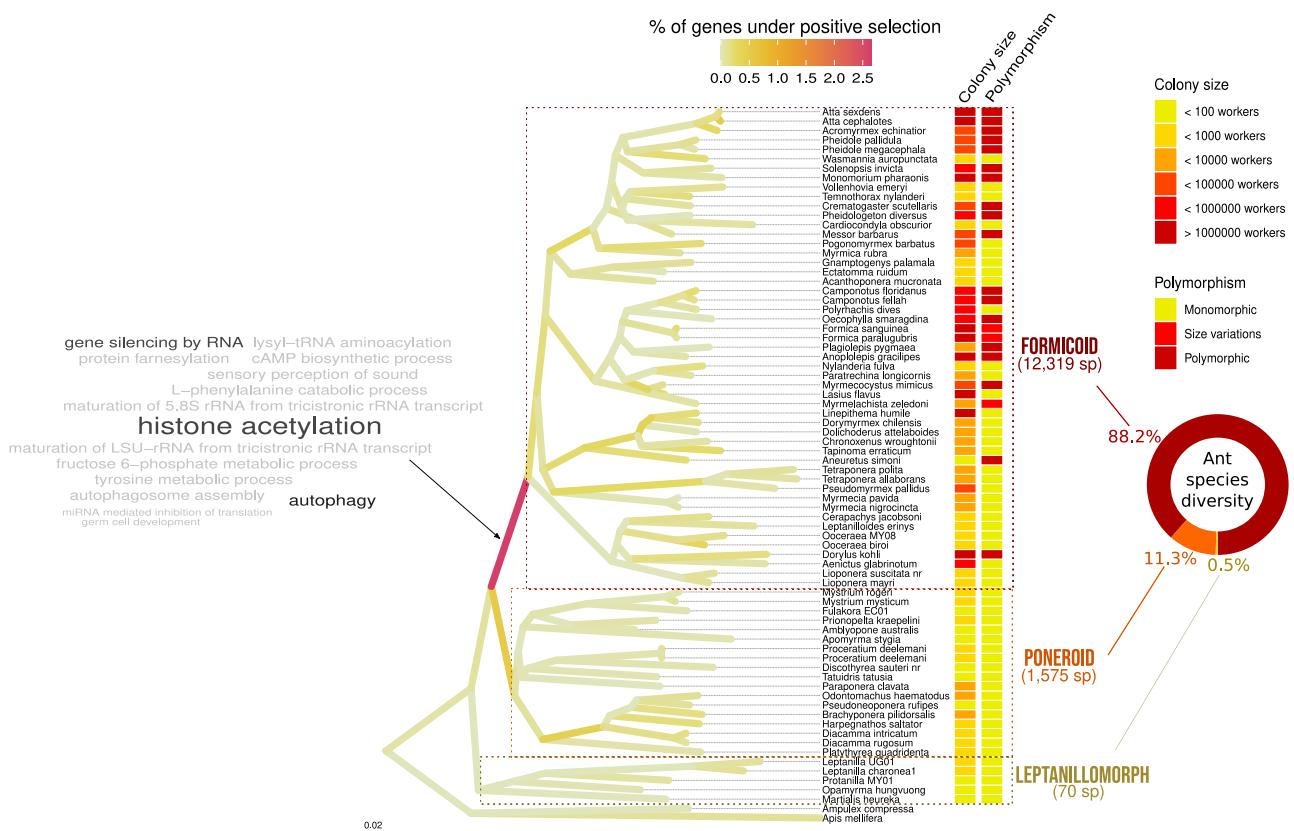


Figure 2. Branch subtending socially diverse formicoid ants shows increased rates of positive selection

Colors indicate the percentage of genes significantly under positive selection (aBSREL analysis) on each branch. Colony size and polymorphism data represent the maximum observed value in the genus, and they have been extracted from the literature.²³ Significantly enriched functional categories (biological process) under positive selection are represented as a word cloud for the formicoid branch. The size of the font is proportional to the p value (Fisher's exact test; larger font indicating the most significant ones, from 0.011 to 0.047). Darker colors indicate the three functions, with the highest numbers of significant annotated genes under positive selection. See also Data S1.

are not affected by the inclusion of all these outgroups (Figure S1B). Second, using random combinations of outgroups (see STAR Methods for details), we always found the same rooting, with strong support for the leptanillomorph clade being sister to all other ants (bootstrap values: 100 for 109 trees and 99 for the remaining 6 trees). Because all species of the leptanillomorph clade are pale, blind, and have a similar hypogean ecology, this suggests that some early ants may have escaped extinction by retreating to these stable subterranean habitats before other lineages diversified by developing novel morphological and behavioral adaptations.¹⁰ According to our divergence date estimates, the common ancestor of the leptanillomorphs lived around the Jurassic-Cretaceous boundary (~145 Ma) shortly after the common ancestor of all extant ants (~150 Ma). This suggests the possibility that subterranean lifestyles existed in the ancestors of extant ants or, more likely, that a hypogean lifestyle originated at an early stage in the history of leptanillomorphs. This result contrasts with the fossil evidence because the earliest-known fossilized crown ants were not specialized to subterranean habitats, and they come from Burmese amber deposits that are ~99 Ma old.^{15,16} Set against our divergence date estimates, this indicates a gap in the ant fossil record of ~50 Ma, further emphasizing an

existing discrepancy between fossils and molecular data when it comes to the question of ant origins.¹⁷ This is reminiscent of the debate on the origin of placental mammals, which are estimated to be in the middle Cretaceous by molecular data, whereas there is no fossil record before the K-T crisis.¹⁸ It has been suggested that the lack of fossils may stem from the occurrence of only a few lineages of placental mammals and perhaps small population sizes during the Cretaceous.¹⁹ Similarly, it is possible that the abundance of crown ants was low at first and only sufficiently increased with the rise of angiosperm⁶ to be represented in the fossil record. Alternatively, there may be methodological biases leading to overestimation of divergence dates^{20,21} or incorrect phylogenetic placement of early ant fossils.¹⁶

The pervasive positive selection is associated with the origin of the socially diverse formicoid clade

To investigate the molecular changes associated with the evolution of complex eusociality, we conducted positive selection analyses on the 4,151 ortholog genes of the 75 ant genomes (STAR Methods). The percentage of genes under positive selection varied greatly among the 38 branches ranging from 0% to 2.6% in a single branch (Figure 2). Strikingly, the branch leading to the

formicoid clade stood out as a clear outlier with a 30-fold higher rate of positive selection compared with the average of other tree branches. There were 110 positively selected genes on the branch subtending the formicoid clade, whereas the average number of genes with positive selections was only 3.1 in other branches (maximum value, 20 genes). This finding is particularly remarkable, given that the genes considered in our analysis are highly conserved universal orthologs across Hymenoptera.²² This indicates that extensive molecular changes in well-conserved core genes occurred along the branch giving rise to the formicoids. By contrast, there was no evidence of a further burst of positive selection later in the evolutionary history of the formicoids, including in the multiple branches leading to the most complex eusocial species (Figure 2). This suggests that most of the genetic innovations that are specific to complex eusociality in formicoids occurred in less than ~20 Ma during the early Cretaceous (Figures 1 and 2).

Functional enrichment analyses for the formicoid branch revealed that histone acetylation was the most significantly overrepresented function among the 110 positively selected genes. Histone acetylation is well known for controlling transcriptional activity,²⁴ reprogramming the foraging behavior of the major worker caste into the minor worker caste,²⁵ colony activity rhythms,²⁶ and the longevity/fecundity trade-off in workers.²⁷ Histone acetylation is also involved in the caste determination of honeybees through the effect of royal jelly,²⁸ and it has been identified as a key caste-specific enhancer of transcription regulating the differential larval development of queens and workers.²⁹ Interestingly, our analyses revealed positive selection on *histone acetyltransferase* (Data S1), a gene previously linked to functions potentially relevant to eusociality, such as the regulation of worker polymorphism.³⁰ The second most significant function was autophagy. Autophagy has repeatedly been shown to be essential for queen lifespan extension^{31,32} and the caste-specific programmed cell death responsible for the divergent ovary development in queen and worker honeybees.³³ Finally, the third most significant function was gene silencing by RNA (Figure 2). From our results, we retrieved the gene *Tudor-SN*, which is a candidate for controlling worker sterility in honeybees.³⁴ Altogether, these results reveal that the common ancestor of the formicoid ants underwent important genomic changes relative to the regulation of gene expression (e.g., histone acetylation and gene silencing RNA) and soma maintenance (e.g., autophagy).

These changes may have been important in allowing the evolution of extreme division of labor in formicoid clade, which is the ant clade comprising the vast majority of species exhibiting extreme forms of complex eusociality (e.g., maximum colony size of 3 million polymorphic workers in formicoid *Dorylus* species compared with a maximum of 50,000 monomorphic workers in some poneroid species of the genus *Leptogenys*²³). However, given that the nine formicoid subfamilies also display some species with less complex levels of eusociality, this implies that although the genomic changes that occurred during the early Cretaceous may have favored the emergence of extreme division of labor and more overtly complex forms of eusociality, they did not necessarily lead to such changes in social organization. Knowing which selective pressures triggered these dramatic molecular changes remains an open and intriguing question.

Conclusions

By providing genome-wide data for all ant subfamilies, this study infers the leptanillomorph clade as the sister clade to all other extant ants and clarifies controversial subfamily relationships that will be important for further comparative studies in ants. The comparative genome analysis also reveals important changes in key molecular pathways implicated in the differential gene expression of queens and workers. This burst of molecular innovations, which occurred over ~20 Ma in the early Cretaceous, possibly played an important role in facilitating the evolution of complex eusociality, including the large colony sizes, extensive caste polymorphism, and extreme fecundity/longevity of queens that characterize multiple lineages of formicoids.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Sequencing and genome assembly
 - Phylogeny
 - Divergence dating analyses
 - Positive selection analysis
 - Gene predictions and gene family analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2022.05.001>.

ACKNOWLEDGMENTS

We thank the FEBS (Federation of European Biochemical Societies) for the long-term fellowship granted to J.R., the funding from the French ANR (t-ERC grant RoyalMess) and ERC (RoyalMess, grant 948688) and the ERC and Swiss NSF to L.K., a grant from the US National Science Foundation (CAREER DEB-1943626) to C.R., and the three reviewers for their very useful comments. We thank the following for helping us to collect samples: C.S. Moreau, F.A. Esteves, S. Van Noort, N. Gunawardene, C. Poteaux, T. Bruscht, J.M. Gómez Durán, M. Molet, E. Vargo, Bui Tuan Viet, M.D. Goodisman, T. Delsinne, and J. Orivel.

AUTHOR CONTRIBUTIONS

J.R. and L.K. conceived the study; J.R., M.L.B., C.R., B.L.F., P.S.W., and L.K. coordinated the sample collection efforts; C.L.M. performed the DNA extractions; J.R., M.L.B., Q.H., A.W., and E.L. performed the analyses; and J.R., M.L.B., C.R., B.L.F., P.S.W., and L.K. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 17, 2021

Revised: March 9, 2022

Accepted: May 2, 2022

Published: May 26, 2022

REFERENCES

1. Bolton, B. (2021). An online catalog of the ants of the world. <https://antcat.org>.
2. Dornhaus, A., Powell, S., and Bengston, S. (2012). Group size and its effects on collective organization. *Annu. Rev. Entomol.* 57, 123–141.
3. Oster, G.F., and Wilson, E.O. (1978). *Caste and Ecology in the Social Insects* (Princeton University Press).
4. Keller, L., and Genoud, M. (1997). Extraordinary lifespans in ants: a test of evolutionary theories of ageing. *Nature* 389, 958–960.
5. Ward, P.S. (2014). The phylogeny and evolution of ants. *Annu. Rev. Ecol. Evol. Syst.* 45, 23–43.
6. Moreau, C.S., Bell, C.D., Vila, R., Archibald, S.B., and Pierce, N.E. (2006). Phylogeny of the ants: diversification in the age of angiosperms. *Science* 312, 101–104.
7. Ward, P.S., and Fisher, B.L. (2016). Tales of dracula ants: the evolutionary history of the ant subfamily Amblyoponinae (Hymenoptera: Formicidae). *Syst. Entomol.* 41, 683–693.
8. Branstetter, M.G., Longino, J.T., Ward, P.S., and Faircloth, B.C. (2017). Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. *Methods Ecol. Evol.* 8, 768–776.
9. Borowiec, M.L., Rabeling, C., Brady, S.G., Fisher, B.L., Schultz, T.R., and Ward, P.S. (2019). Compositional heterogeneity and outgroup choice influence the internal phylogeny of the ants. *Mol. Phylogenet. Evol.* 134, 111–121.
10. Rabeling, C., Brown, J.M., and Verhaagh, M. (2008). Newly discovered sister lineage sheds light on early ant evolution. *Proc. Natl. Acad. Sci. USA* 105, 14913–14917.
11. Kück, P., Hita Garcia, F., Misof, B., and Meusemann, K. (2011). Improved phylogenetic analyses corroborate a plausible position of *Martialis heureka* in the ant tree of life. *PLOS One* 6, e21031.
12. Brady, S.G., Schultz, T.R., Fisher, B.L., and Ward, P.S. (2006). Evaluating alternative hypotheses for the early evolution and diversification of ants. *Proc. Natl. Acad. Sci. USA* 103, 18172–18177.
13. Moreau, C.S., and Bell, C.D. (2013). Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. *Evolution* 67, 2240–2257.
14. Peters, R.S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., Lanfear, R., et al. (2017). Evolutionary history of the Hymenoptera. *Curr. Biol.* 27, 1013–1018.
15. Barden, P., and Grimaldi, D.A. (2016). Adaptive radiation in socially advanced stem-group ants from the Cretaceous. *Curr. Biol.* 26, 515–521.
16. Boudinot, B.E., Khouri, Z., Richter, A., Griebenow, Z.H., van de Kamp, T., Perrichot, V., and Barden, P. (2022). Evolution and systematics of the Aculeata and kin (Hymenoptera), with emphasis on the ants (Formicoidea: †@#@idae fam. nov., Formicidae). Preprint at bioRxiv. <https://doi.org/10.1101/2022.02.20.480183>.
17. Borowiec, M.L., Moreau, C.S., and Rabeling, C. (2020). Ants: phylogeny and classification. In *Encyclopedia of Social Insects*, C.K. Starr, ed. (Springer International Publishing), pp. 1–18.
18. dos Reis, M., Inoue, J., Hasegawa, M., Asher, R.J., Donoghue, P.C.J., and Yang, Z. (2012). Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. Biol. Sci.* 279, 3491–3500.
19. Cooper, A., and Fortey, R. (1998). Evolutionary explosions and the phylogenetic fuse. *Trends Ecol. Evol.* 13, 151–156.
20. Springer, M.S., Foley, N.M., Brady, P.L., Gatesy, J., and Murphy, W.J. (2019). Evolutionary models for the diversification of placental mammals across the KPg boundary. *Front. Genet.* 10, 1241.
21. Bromham, L., Duchêne, S., Hua, X., Ritchie, A.M., Duchêne, D.A., and Ho, S.Y.W. (2018). Bayesian molecular dating: opening up the black box. *Biol. Rev. Camb. Philos. Soc.* 93, 1165–1191.
22. Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* 1962, 227–245.
23. Blanchard, B.D., and Moreau, C.S. (2017). Defensive traits exhibit an evolutionary trade-off and drive diversification in ants. *Evolution* 71, 315–328.
24. Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* 12, 599–606.
25. Simola, D.F., Graham, R.J., Brady, C.M., Enzmann, B.L., Desplan, C., Ray, A., Zwiebel, L.J., Bonasio, R., Reinberg, D., Liebig, J., et al. (2016). Epigenetic (re)programming of caste-specific behavior in the ant *Camponotus floridanus*. *Science* 351, aac6633.
26. Libbrecht, R., Nadrau, D., and Foitzik, S. (2020). A role of histone acetylation in the regulation of circadian rhythm in ants. *iScience* 23, 100846.
27. Choppin, M., Feldmeyer, B., and Foitzik, S. (2021). Histone acetylation regulates the expression of genes involved in worker reproduction and lifespan in the ant *Temnothorax rugatulus*. *BMC Genomics* 22, 871.
28. Spannhoff, A., Kim, Y.K., Raynal, N.J.-M., Gharibyan, V., Su, M.B., Zhou, Y.Y., Li, J., Castellano, S., Sbardella, G., Issa, J.P., et al. (2011). Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. *EMBO Rep.* 12, 238–243.
29. Wojciechowski, M., Lowe, R., Maleszka, J., Conn, D., Maleszka, R., and Hurd, P.J. (2018). Phenotypically distinct female castes in honey bees are defined by alternative chromatin states during larval development. *Genome Res.* 28, 1532–1542.
30. Alvarado, S., Rajakumar, R., Abouheif, E., and Szyf, M. (2015). Epigenetic variation in the Egfr gene generates quantitative variation in a complex trait in ants. *Nat. Commun.* 6, 6513.
31. Madeo, F., Zimmermann, A., Maiuri, M.C., and Kroemer, G. (2015). Essential role for autophagy in life span extension. *J. Clin. Invest.* 125, 85–93.
32. Nakamura, S., and Yoshimori, T. (2018). Autophagy and longevity. *Mol. Cells* 41, 65–72.
33. Barchuk, A.R., Cristina, A.S., Kucharski, R., Costa, L.F., Simões, Z.L., and Maleszka, R. (2007). Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Dev. Biol.* 7, 70.
34. Hartfelder, K., Tiberio, G.J., Lago, D.C., Dallacqua, R.P., and Bitondi, M.M.G. (2018). The ovary and its genes—developmental processes underlying the establishment and function of a highly divergent reproductive system in the female castes of the honey bee, *Apis mellifera*. *Apidologie* 49, 49–70.
35. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
36. Jackman, S.D., Vandervalk, B.P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S.A., Jahesh, G., Khan, H., Coombe, L., Warren, R.L., and Birol, I. (2017). ABYSS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* 27, 768–777.
37. Chikhi, R., and Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. *Bioinformatics* 30, 31–37.
38. Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: interrogation of genome assemblies. *F1000Res* 6, 1287.
39. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., and Pyshkin, A.V. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477.
40. Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simão, F.A., Ioannidis, P., Seppey, M., Loetscher, A., and Kriventseva, E.V. (2017). OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* 45, D744–D749.

41. Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548.
42. Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33, 511–518.
43. Borowiec, M.L. (2019). Spruceup: fast and flexible identification, visualization, and removal of outliers from large multiple sequence alignments. *J. Open Source Softw.* 4, 1635.
44. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAI: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973.
45. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274.
46. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19, 153.
47. Ranwez, V., Harispe, S., Delsuc, F., and Douzery, E.J.P. (2011). MACSE: multiple alignment of coding sequences accounting for frameshifts and stop codons. *PLOS One* 6, e22594.
48. Di Franco, A., Poujol, R., Baurain, D., and Philippe, H. (2019). Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol. Biol.* 19, 21.
49. Smith, M.D., Wertheim, J.O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S.L. (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* 32, 1342–1353.
50. Alexa, A., and Rahnenführer, J. (2009). Gene set enrichment analysis with topGO. *Bioconductor Improv.* 27, 1–26.
51. Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12, 491.
52. Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezhuk, Y., et al. (2013). BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* 41, W29–W33.
53. Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* 5, 4–10.
54. Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6, 31.
55. Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5, 59.
56. Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33, W465–W467.
57. Schattner, P., Brooks, A.N., and Lowe, T.M. (2005). The tRNAscan-SE, snoScan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, W686–W689.
58. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238.
59. Mendes, F.K., Vanderpool, D., Fulton, B., and Hahn, M.W. (2020). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518.
60. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314.
61. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760.
62. Yu, G. (2020). Using ggtree to visualize data on tree-like structures. *Curr. Protoc. Bioinformatics* 69, e96.
63. Aljanabi, S.M., and Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.* 25, 4692–4693.
64. Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522.
65. Naser-Khdour, S., Minh, B.Q., Zhang, W., Stone, E.A., and Lanfear, R. (2019). The prevalence and impact of model violations in phylogenetic analysis. *Genome Biol. Evol.* 11, 3341–3352.
66. Wang, H.C., Minh, B.Q., Susko, E., and Roger, A.J. (2018). Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* 67, 216–235.
67. Tihelka, E., Cai, C., Giacomelli, M., Lozano-Fernandez, J., Rota-Stabelli, O., Huang, D., Engel, M.S., Donoghue, P.C.J., and Pisani, D. (2021). The evolution of insect biodiversity. *Curr. Biol.* 31, R1299–R1311.
68. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528.
69. Gatesy, J., and Springer, M.S. (2014). Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Mol. Phylogenet. Evol.* 80, 231–266.
70. Faircloth, B.C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32, 786–788.
71. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.
72. dos Reis, M., and Yang, Z. (2011). Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol. Biol. Evol.* 28, 2161–2172.
73. Economo, E.P., Narula, N., Friedman, N.R., Weiser, M.D., and Guénard, B. (2018). Macroecology and macroevolution of the latitudinal diversity gradient in ants. *Nat. Commun.* 9, 1778.
74. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904.
75. Yang, Z., and Dos Reis, M. (2011). Statistical properties of the branch-site test of positive selection. *Mol. Biol. Evol.* 28, 1217–1228.
76. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300.
77. Galtier, N., and Duret, L. (2007). Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends Genet.* 23, 273–277.
78. Romiguier, J., and Roux, C. (2017). Analytical biases associated with GC-content in molecular evolution. *Front. Genet.* 8, 16.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|------------------------------|--|---|
| Biological samples | | |
| Ant tissue samples | This study | See Table S1 |
| Critical commercial assays | | |
| QIAamp DNA Micro Kit | QIAGEN | Cat# 56304 |
| Deposited data | | |
| Genome raw reads | This study | See Table S1 and ENA: PRJEB48742; |
| Genome assemblies | This study | Zenodo: https://doi.org/10.5281/zenodo.5705739 |
| Alignments | This study | Zenodo: https://doi.org/10.5281/zenodo.5705739 |
| Phylogenetic trees | This study | Zenodo: https://doi.org/10.5281/zenodo.5705739 |
| Analyses raw output | This study | Zenodo: https://doi.org/10.5281/zenodo.5705739 |
| Software and algorithms | | |
| Trimmomatic v 0.36 | Bolger et al. ³⁵ | https://github.com/usadelab/Trimmomatic |
| AbySS 2.0.2 | Jackman et al. ³⁶ | https://github.com/bcgsc/abyss |
| KmerGenie v1.7016 | Chikhi and Medvedev ³⁷ | http://kmergenie.bx.psu.edu/ |
| Blobtools v1.1 | Laetsch and Blaxter ³⁸ | https://github.com/DRL/blobtools |
| SPAdes 3.9.0 | Bankevich et al. ³⁹ | https://github.com/ablab/spades |
| OrthoDB v9 | Zdobnov et al. ⁴⁰ | https://www.orthodb.org/ |
| BUSCO v 3.02 | Waterhouse et al. ⁴¹ | https://busco.ezlab.org/ |
| MAFFT v 7.310 | Katoh et al. ⁴² | https://github.com/GSLBiotech/mafft |
| Spruceup | Borowiec ⁴³ | https://github.com/marekborowiec/spruceup |
| TrimAI v 1.2 | Capella-Gutiérrez et al. ⁴⁴ | http://trimal.cgenomics.org/ |
| IQ-TREE v 2.0.5 | Nguyen et al. ⁴⁵ | http://www.iqtree.org/ |
| ASTRAL 5.7.4 | Zhang et al. ⁴⁶ | https://github.com/smirarab/ASTRAL |
| macse v1.2 | Ranwez et al. ⁴⁷ | https://bioweb.supagro.inra.fr/macse/ |
| hmmcleaner v1.8 | Di Franco et al. ⁴⁸ | https://doi.org/10.5281/zenodo.5705739 |
| aBSREL v 2.2 (HyPhy package) | Smith et al. ⁴⁹ | https://stevenweaver.github.io/hyphy-site/ |
| topGO | Alexa and Rahnenführer ⁵⁰ | https://bioconductor.org/packages/release/bioc/html/topGO.html |
| MAKER v 2.31.8 | Holt and Yandell ⁵¹ | https://www.yandell-lab.org/software/maker.html |
| ncbi-blast v 2..2.28 | Boratyn et al. ⁵² | https://blast.ncbi.nlm.nih.gov/Blast.cgi |
| RepeatMasker v 4.0.5 | Chen ⁵³ | https://www.repeatmasker.org/ |
| exonrate v 2.2.0 | Slater and Birney ⁵⁴ | https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonrate |
| snap v 2013.11.29 | Korf ⁵⁵ | https://github.com/KorfLab/SNAP |
| augustus v 3.2.2 | Stanke and Morgenstern ⁵⁶ | http://augustus.gobics.de/ |
| tRNAscan-SE | Schattner et al. ⁵⁷ | http://lowelab.ucsc.edu/tRNAscan-SE/ |
| snoScan 0.9 | Schattner et al. ⁵⁷ | http://lowelab.ucsc.edu/snoScan/ |
| orthofinder v 2.2.1 | Emms and Kelly ⁵⁸ | https://github.com/davidemms/OrthoFinder |
| CAFE v5 | Mendes et al. ⁵⁹ | https://github.com/hahnlab/CAFE5 |
| eggNOG v5 | Huerta-Cepas ⁶⁰ | http://eggnog5.embl.de |
| Bwa v0.7.15 | Li and Durbin ⁶¹ | http://bio-bwa.sourceforge.net/ |
| ggtree | Yu ⁶² | https://bioconductor.org/packages/release/bioc/html/ggtree.html |

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources be directed to and will be fulfilled by the lead contact, Jonathan Romiguier (jonathan.romiguier@umontpellier.fr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw reads of sequenced genomes data have been deposited at ENA (European Nucleotide Archive). Accession numbers are listed in the [key resources table](#). Genome assemblies, data, raw results and command lines for reproducibility are available in the following Zenodo repository and are publicly available as of the date of publication Zenodo: <https://doi.org/10.5281/zenodo.5705739>. DOIs are listed in the [key resources table](#).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We sampled 65 species (64 ants, 1 jewel wasp *Ampulex compressa*) across all ant subfamilies. We collected specimen from various sources and collectors (details and full overview of samples available in [Table S1](#)).

METHOD DETAILS

Sequencing and genome assembly

DNA extractions have been performed based on a high salt method⁶³ on samples conserved in ethanol or in -80°C freezers.

We sequenced the genomes of 65 samples (64 ants and the jewel wasp *Ampulex compressa*) using Illumina Hiseq technology (paired-end, 150 bp reads). Reads were cleaned using *Trimmomatic*³⁵ and first assembled using *AbySS* 2.0.2³⁶ with kmer size set on 61 or the optimal value as estimated by KmerGenie v1.7016.³⁷ We removed potential contaminations using the blobtools pipeline,³⁸ with the exclusion of contigs that blasted on any non-arthropod phylum against the Genebank database (*nt*). Reads mapping on these contaminant contigs were filtered and remaining reads were used for a second genome assembly using *SPADEs* 3.9.0.³⁹ For each species, we selected the best assembly between *SPADEs* and *AbySS* based on the number of complete and single copy ortholog genes using *BUSCO*²² with a 4,415 Hymenoptera ortholog dataset from OrthoDB v9⁴⁰ (see [Table S1](#) for assembler, *BUSCO* scores and N50 of each species).

Phylogeny

We built our phylogenomic dataset by complementing the 4,415 ortholog genes of our 65 species with 17 (9 ants and 8 hymenopteran outgroups) supplemental reference genomes from OrthoDB v9.⁴⁰ We aligned the amino-acid sequences using *mafft*⁴² and cleaned the resulting alignments with *Spruceup*⁴³ and *trimal*⁴⁴ with the “automated1” option.

We concatenated the alignments in a supermatrix that we analysed with two different substitution models with IQ-TREE v 2.0.5⁴⁵ and 1,000 ultrafast bootstraps.⁶⁴ First, we performed a gene partitioned analysis (all partitions share the same set of branch lengths but have their own evolution rate) with LG+F+G4 models after having removed partitions that failed at symmetry tests testing stationarity and homogeneity assumptions⁶⁵ (resulting tree is presented in [Figure S1C](#) and has been used as the *guide tree* for the next analysis). Second, because the most controversial ant subfamily relationships are expected to be affected by long-branch-attraction artefacts (Martialinae and Paraponerinae are monotypic subfamilies with long branches), we used the posterior mean site frequency model (PMSF model, LG+C20+F+G) which has been designed to correct for such artefacts by modeling site heterogeneity.^{66,67} The resulting tree is presented in [Figures 1](#) and [S1D](#) and is referred in the results as the main supermatrix analysis. We also used this tree for estimating divergence times via penalised maximum likelihood approaches⁶⁸ and a set of 12 node calibrations (details of calibrations and references in [Table S2](#)).

A coalescent-based species tree analysis was performed by first producing gene trees using IQ-TREE⁴⁵ with the substitution model selected for each alignment by the built-in ModelFinder option MFP+MERGE. Because coalescent-based species tree approaches are sensitive to inaccurate gene trees,⁶⁹ we only kept gene trees from long alignments of more than 500 amino-acids (n=1,366) and used ASTRAL 5.7.4⁴⁶ for producing the supertree ([Figure S1E](#)).

In case outgroup composition affected our results,⁹ we built an alternative dataset containing 115 outgroups of Aculeate species by matching our OrthoDB IDs with the IDs provided by the alignments of an Hymenoptera phylogenomic dataset,¹⁴ resulting in a 2,343 gene supermatrix (983,951 sites) after an automated *trimal* cleaning. We first performed the same PMSF analysis as described above (see tree in [Figure S1B](#)). Second, we produced a reduced supermatrix by keeping only alignments with at least 90% of species

and sites with 90% of non-ambiguous characters, resulting in a supermatrix with 271,959 sites. We then removed from 0 to 115 random outgroups from this supermatrix, producing 116 new supermatrices for testing the effect of random outgroup removal. The same PMSF tree inference as described above was then performed, resulting in 116 phylogenetic trees (available in the Zenodo repository Zenodo: <https://doi.org/10.5281/zenodo.5705739>).

We built an alternative dataset of ultra-conserved elements loci (UCE) from our genomes and merged the data with the phylogenetic dataset of Branstetter et al.⁸ We used phyluce⁷⁰ to extract UCE loci from our genome assemblies. 2,510 loci were retrieved but we only kept the 1,855 that were common with the Branstetter et al.⁸ dataset. We aligned the data using mafft,⁴² cleaned the alignment using trimal⁴⁴ with the *-automated1* option and removed alignments that contained fewer than 75% of the total number of taxa ($n = 166$). We concatenated the 1,230 remaining alignments in a supermatrix with 426,015 sites and inferred a tree using IQ-TREE with a locus partitioned analysis (all partitions share the same set of branch lengths but have their own evolution rate) with GTR+I+G4 models and 1000 ultrafast bootstraps (Figure S1A).

Divergence dating analyses

First, we used a simple approach exploiting the largest supermatrix by using the topology and branch lengths of Figure S1D (supermatrix of 1,692,052 amino acids, PMSF model LG+C20+F+G) to estimate divergence times via penalised maximum likelihood approaches⁶⁸ and a set of 12 node calibrations (details of calibrations and references in Table S2). To confirm the retrieved estimations (presented in Figure 1), we analysed a reduced dataset using a Bayesian approach, as implemented in MCMCTree, a part of the PAML package, v4.10.⁷¹ MCMCTree utilizes rapid approximate likelihood computation,⁷² which makes it suitable for divergence dating of genome-scale data sets.¹⁸ Due to computational constraints, we used an alignment with loci containing a minimum of 95% of our 83 taxa, totalling 182,809 amino acid sites. We fixed the topology to be the same as our analysis of the full alignment. We constrained our root node with a soft bound maximum age of 236 Ma, corresponding to the lower bound of the 95% highest posterior density (HPD) interval for that split in Hymenoptera tree estimations.¹⁴ We also set soft bounds on the root of the Formicidae to be 103 Ma and 169 Ma, corresponding to the upper 95% bound of HPD in Borowiec et al.⁹ and lower bound in Economo et al.,⁷³ the most divergent of recent estimates for the crown age of the family.¹⁷ We also used minimum node age constraints based on fossils presented in Table S2. We ran each analysis unpartitioned, under the LG model for 5 million generations. We examined each run's statistics in Tracer⁷⁴ and confirmed convergence and sufficient effective sample sizes (>>200) for all parameters. Retrieved estimations were close to those retrieved with penalised likelihood on the whole dataset (Figure 1) and are available with all output files in the zenodo repository (Zenodo: <https://doi.org/10.5281/zenodo.5705739>).

Positive selection analysis

We performed a positive selection detection analysis by using 4,415 nucleotide alignments. Nucleotide alignments have been produced and refined from amino acid alignments using the command *reportGapsAA2NT* and *refineAlignment* from macse v1.2.⁴⁷ We cleaned the alignments of potential errors further by using *hmmcleaner v1.8*, a tool that has been reported as especially effective for reducing false positives for detection of positive selection.⁴⁸ We used the value of 5 for the threshold parameter then removed every species with fewer than 20% nucleotides remaining after the cleaning. To ensure that gap-rich regions did not bias our analyses, we applied two different supplemental cleaning treatments by keeping only codons shared with more than 50 and 75% of the species of the alignment. All of the following analyses were performed with the three cleaning strategies (hmmcleaner only; hmmcleaner+50% complete codons; hmmcleaner+75% complete codons) and retrieved consistent results regarding the relatively high percentage of genes under positive selection in the Formicoid branch compared to other branches (38.13, 36.26 and 30.58-fold increases, respectively, see Data S1). Only the results of the “hmmcleaner+more than 75% complete codons” treatment are presented in the main text.

We used the ABSREL method (adaptive Branch-Site Random Effects Likelihood) from the HyPhy package,⁴⁹ an improved implementation of the branch-site model typically used to test whether positive selection has occurred on some branches via the estimation of dN/dS (non-synonymous substitution rate over synonymous substitution rate).⁷⁵ All branches were tested for positive selection for each gene, with p values corrected for multiple testing on multiple branches (using the built-in correction in aBSREL). An additional correction was conducted for multiple testing on multiple genes.⁷⁶ For each internal branch in Figure 2, we only considered alignments containing at least one species for each of its three connected clades, ensuring that the positive selection test reflects this exact part of the evolutionary history of the gene. For each gene, we reported the gene ontology function of the *Apis mellifera* ortholog gene as available in OrthoDB.⁴⁰ Gene ontology enrichment analyses have been performed using topGO⁵⁰ with Fisher exact tests and the default weight01 algorithm.

GC-content variations are known to potentially bias detection of positive selection methods via the process of biased gene conversion.^{77,78} Particularly for our results, strong GC-content variations among Formicoid and Poneroid species could lead to an overestimation of positive selection in the branch leading to Formicoids. To ensure that it is not the case, we measured the average GC-content of all species and retrieved no significant difference between Formicoids and Poneroids when considering all genes (45.43% vs 46.14% in Formicoids and Poneroids, p value = 0.14 from Welch two sample test) or only considering the 110 genes retrieved as positively selected from the main analysis (48.10% vs 48.92% in Formicoids and Poneroids, p value = 0.15 from Welch two sample test).

Gene predictions and gene family analyses

We performed gene predictions for our 65 genomes by using MAKER2 v 2.31.8,⁵¹ a pipeline for genome annotation using ncbi-blast v 2.2.28,⁵² RepeatMasker v 4.0.5,⁵³ exonerate v 2.2.0,⁵⁴ snap v 2013.11.29,⁵⁵ augustus v 3.2.2,⁵⁶ tRNAscan-SE v1.3.1 and snoScan

0.9.⁵⁷ We filtered genes with fewer than 1000 nucleotides and individual protein sets were blasted against each other as well as against 28 additional Hymenoptera protein sets (detailed list available in output files available on Zenodo: <https://doi.org/10.5281/zenodo.5705739>) using orthofinder v 2.2.1.⁵⁸ The resulting gene count data file was then used for a gene family evolution analysis with CAFE v5⁵⁹ with base model default setting values. After trying several filtering methods, we removed gene families with difference in gene number larger than 50 to prevent “-inf” likelihood scores.

We assigned GO terms for our 65 protein sets using eggNOG v5⁶⁰ and used it to assign GO terms to gene families analysed with CAFE. To identify gene functions over-represented in gene families that underwent significant expansion/contraction, GO term enrichment analyses were performed using topGO,⁵⁰ with Fisher exact tests and the default weight01 algorithm. Gene functions significantly over-represented and potentially related to eusociality include *autophagy*, determination of adult lifespan, oogenesis, detection of chemical stimulus involved in sensory perception of smell, maintenance of chromatin silencing, olfactory receptor activity, histone deacetylase binding, histone kinase activity (see the topGOresults tables in the Zenodo repository for the whole list). However, these results should be taken with caution because our analyses showed that the genomes available in public databases tended to exhibit greater increases/decreases of gene families than our 65 genomes. This is probably due to the fact that we had to use short-read sequencing technologies to be able to analyse low amounts of DNA and/or degraded DNA for some species that are rare and very difficult to collect. We therefore chose not to present these results in the main text but instead made them available in a Zenodo repository (Zenodo: <https://doi.org/10.5281/zenodo.5705739>) where we provide MAKER2 control file and options, resulting protein fasta files, orthofinder main output files, CAFE5 input and output files, eggNOG annotations and topGO analysis input/output.