Relaxation of purifying selection suggests low effective population size in eusocial Hymenoptera and solitary pollinating bees

Arthur Weyna[1], Jonathan Romiguier[1]

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

Corresponding author: jonathan.romiguier@umontpellier.fr

ABSTRACT

~~Eusociality and parastitism are specific life-history strategies particularly common in Hymenoptera.~~With one of the highest number of parasitic, eusocial and pollinator species among all insect orders, Hymenoptera features a great diversity of lifestyles. At the population genetics level, ~~low effective population size and reduced efficiency of purifying selection have been suggested as an evolutionary consequence of both social life and parasitism.~~such life-history strategies are expected to decrease effective population size and the efficiency of purifying selection. In this study, we tested this hypothesis~~these hypotheses~~ by estimating the relative rate of non-synonymous substitution in 169 species to investigate the variation in natural selection efficiency ~~and effective population size~~ throughout the hymenopteran~~Hymenoptera~~ tree of life. We found no effect of parasitism, but we show that relaxed selection is associated with eusociality, suggesting that the division of reproductive labour decreases effective population size in ants, bees and wasps. Unexpectedly, the effect of eusociality is marginal compared to a striking and widespread relaxation of selection in both social and non social bees, which indicates that these keystone pollinator species generally feature low effective population sizes. T~~Albeit strong, t~~his widespread pattern ~~in pollinating bees is difficult to clearly explain, but might suggest links between drops of effective population size and evolution of social life, as bees are the Hymenoptera group with the most appearances of eusociality in the tree of life.~~suggests specific constraints in pollinating bees potentially linked to limited resource availability and high parental investment. The particularly high load of deleterious mutations we report in the genome of these crucial ecosystem engineer species also raises new concerns about their ongoing population decline.

INTRODUCTION

The intensity of ~~the~~ genetic drift experienced by a population depends on its effective population size *Ne* (Wright 1931). Deleterious mutations reach fixation with a higher probability in small populations, that~~which~~ undergo more drift, than in large populations where~~in which~~ purifying selection is more efficient. *Ne* is usually defined for any observed population as the theoretical census size an ideal Wright-Fisher population should have to show the level of drift displayed by the observed population (Wang et al.~~, Santiago, and Caballero~~ 2016). While different definitions of ***Ne*** exist depending on the field, it generally correlates negatively with any process breaking the assumption of panmixia, which underlies the Wright-Fisher model (i.e., population structure, homogamy, inbreeding…). Building on this knowledge, it has been proposed that basic traits influencing the reproductive output and mating choices of organisms, such as life-history traits, should correlate with their genome-wide deleterious substitutions rates. Several examples confirming these predictions have been uncovered in the last two decades: species generation time, longevity or body mass were found to be~~is~~ positively correlated with the genome-wide dN/dS (i.e., ratio of the non-synonymous substitution rate to the synonymous substitution rate) (Nikolaev et al. 2007; Romiguier~~, Lourenco,~~ et al. 2014; Popadin et al. 2007; Figuet et al. 2016; Botero-Castro et al. 2017; Rolland et al. 2020). However, most known examples are clustered within a few v~~V~~ertebrate taxa: mainly mammals, birds or reptiles. To date, only few examples of such patterns have been found in invertebrates, which casts doubt on the existence of a general relationship between life history strategies and the efficiency of natural selection in Metazoa. Various reasons might explain the difficulty to demonstrate such relationships in invertebrates: There is relatively fewer~~less~~ genomic data available than in mammals or birds, and gathering life-history data in a large number of non-model invertebrates can be difficult as they have generally received less attention than ~~Vertebrates~~vertebrates. Effective population size comparisons among invertebrate~~s~~ clades can also be particularly difficult, as the existence of reproductive systems such as haplo-diploidy, where the male is haploid, affect ***Ne*** estimations (Wang et al. 2016).

Among invertebrates, all Hymenoptera conveniently share the same haplo-diploid system, while displaying a particularly wide diversity of life-history strategies. Notably, they exhibit extreme lifestyles that can be predicted to strongly influence their reproductive output, and thus their long-term *Ne*. First, many species within this clade are parasites of plants (phytophageous) or parasitoids of other a~~A~~rthropods (parasitoids) (Mayhew 2016), which could shape their demography, as population structure and size of the host can restrict~~influence~~ that of the parasite or, especially, of the parasitoid (Mazé-Guilmo et al. 2016). Second, the h~~H~~ymenopteran order contains a large number of pollinators, such as bees, that are involved in keystone insect-plant mutualisms and strictly depend on a limited floral resource ~~likely to be~~generally scattered in time and space. Because of the limited availability of resources required for their brood, pollinating bees have to invest a lot of time and energy in terms of foraging, nesting and food processing for their offsprings (Zayed et al. 2005; Zayed and Packer 2007), which should limit their reproductive output and population size. Supporting this hypothesis, high parental investment has been previously identified as a general proxy for low *Ne* in animals (Romiguier et al. 2014a). Third,~~Finally,~~ eusociality, which is a rare lifestyle in Animals, is relatively common in the Hymenoptera order with at least 9 independent~~independant~~ ~~apparitions~~appearances (Hughes et al. 2008). Eusociality is characterized by the differentiation of reproductive and non-reproductive castes of individuals within a colony, which cooperate for resource acquisition and brood care (Crespi and Yanega 1995). Because reproduction is typically monopolized by few long-lived reproductive individuals (Keller and Genoud 1997), a decrease in~~of~~ long-term ***Ne***~~,~~ and~~ of~~ the efficiency of natural selection, is often thought~~believed~~ to be a general consequence of eusociality (Bromham and Leys 2005; Romiguier~~, Lourenco,~~ et al. 2014; Settepani et al.~~, Bechsgaard, and Bilde~~ 2016). Maintenance of high relatedness within low-*Ne* inbred groups has also been proposed~~raised~~ as a prerequisite to the evolution of eusociality because it favors altruistic behaviors through kin-selection **(Hamilton 19**64~~72~~**; Husseneder et al. 1999; Hughes et al. 2008; Tabadkani et al. 2012).** Ancestral population bottlenecks could thus be a typical feature of taxa in which eusociality frequently evolves, which is the case for several independent clades in Hymenoptera. Alternatively, it has been hypothesized that important population bottlenecks may be rare in Hymenoptera, because the associated loss of genetic diversity on single locus sex determination would lead to the costly production of sterile diploid males (Asplen et al. 2009; Rabeling and Kronauer 2013).~~.~~

~~Some early studies have tried to detect *Ne* decreases in eusocial lineages by comparing their heterozygosity (Owen 1985; Berkelhamer 1983; Reeve, Reeve, and Pfennig 1985) with that of solitary lineages. These studies did not, however, allow a consensus to be reached, maybe in part because of the limiting size of the genetic data set available at the time. A more recent study showed that the genome-wide genetic diversity of four eusocial lineages was closer to that of~~

~~mammals and birds (Romiguier et al. 2014) than to that of four solitary insects, suggesting a reduced *Ne* in eusocial lineages. This Ne decrease is expected to come at a particularly high cost in haplo-diploid Hymenoptera, as most species are expected rely on a single locus sex determination system (Asplen et al. 2009), which can trigger the production of diploid sterile males when genetic diversity is reduced (Rabeling and Kronauer 2013). However, the results of this study are to be taken with care, as the number of compared taxa is low, and as none of the solitary reference taxa are Hymenoptera, while almost all compared eusocial taxa belong to this group. Indeed, Ne could be inherently reduced in Hymenoptera, irrespective of eusociality, due to their haploid males (Wang, Santiago, and Caballero 2016). A study accommodating this problem by comparing genetic data for eusocial and solitary sister species in Hymenoptera and other taxa did not reveal any significant effect of eusociality on molecular evolution (Bromham and Leys 2005). This represents the strongest argument to date against a systematic reduction of *Ne* in eusocial lineages, as pair comparisons allow correction for the effects of ploidy and phylogeny, and as implicated pairs are numerous (25) and distributed over most known eusocial lineages. However, the small size of the genetic dataset (1 to 3 genes) in this study might have made it statistically difficult to detect any effect.~~So far, few studies have investigated how *Ne* varies among Hymenoptera, and all were restricted to the effect of eusociality alone (Owen 1985; Berkelhamer 1983; Reeve et al. 1985; Bromham and Leys 2005; Romiguier et al. 2014; Imrit et al. 2020). Only recent studies with genome-wide datasets have detected associations between eusociality and decreases in *Ne* (Romiguier et al. 2014; Imrit et al. 2020), but these studies are typically restricted to few taxa compared to studies that rejected any significant effect (Bromham and Leys 2005). Disregarding the joint effect of other potential *Ne* determinants (e.g., body-size, parasitism, pollen-feeding, haplo-diploidy) may bias results and explain the discrepancy among studies with low vs high number of species comparisons.

Here, we tried to better assess the respective effects of potential *Ne* determinants in Hymenoptera. We used a phylogenomic dataset of 3256 genes in 169 species of Hymenoptera ~~species~~ (Peters et al. 2017), including 10 eusocial species distributed among 4 independent origins of eusociality (Formicidae : 3 species; Polistinae/Vespinae wasps : 3 species; Stenogastrinae wasps : 1 species; Corbiculate bees : 3 species), ~~with the goal of disentangling the links between effective population size, deleterious mutation load, and causes and consequences of eusociality. This dataset also contains~~ 112 parasitic species and 32 solitary pollinating bees. We estimated~~ the efficiency of purifying selection by estimating~~ mean genomic dN/dS and~~. We~~ compared these estimations ~~between eusocial and solitary species, parasitic and free-living species~~between solitary and eusocial taxa, as well as between free and parasitic taxa (see figure S3),. We also ~~, and~~ correlated these~~m~~ to body size, life-history descriptor variables of parasitoids (see table S1 for details) and geographical range descriptors (see table S5 for details). We further~~also~~ confirmed ~~the links between increased dN/dS and drops in the efficiency of purifying selection in eusocial species and solitary bees~~that detected increases in dN/dS do correspond to relaxed purifying selection (and thus to drops in *Ne*), via specialized analyses that differentiate~~ing~~ positive selection from relaxed purifying selection. Unexpectedly, we found that, instead of large species, parasites or eusocial taxa, pollinating bees display by far the lowest long-term *Ne* among Hymenoptera.

RESULTS

dN/dS distribution across the Hymenoptera phylogeny

We estimated dN/dS in 3241 gene alignments of 169 species of Hymenoptera~~ species~~ using the mapNH program (Romiguier et al. 2012;~~,~~ https://github.com/BioPP/testnh) from the *testnh* program suite (Dutheil and Boussau 2008; Guéguen and Duret 2017~~8~~). We used the tree obtained by Peters et al. (2017) and its topology through all analyses to correct for phylogenetic inertia. As eusocial Hymenoptera are known to have high recombination rates (Wilfert et al. 2007; Sirviö et al. 2011; Wallberg et al. 2015;~~,~~ Jones et al. 2019), which in turn are known to inflate dN/dS when associated to biased gene conversion in ~~Vertebrates~~vertebrates (Duret and Galtier 2009; Lartillot 2012~~3~~; Galtier et al. 2018), we estimated dN/dS considering GC-conservative substitutions only. Estimated rates should, therefore, be impervious to the effects of biased gene conversion (Galtier et al. 2018). Average corrected genomic dN/dS values are displayed along the h~~H~~ymenopteran tree on figure 1 (see the d~~D~~istribution of uncorrected dN/dS values~~ are displayed~~ in figure S1). The largest and smallest mean ratios were inferred for *Eucera nigrescens* (0.1901) and Cimbex rubida (0.0684), respectively. As expected for conserved coding regions, the distribution of genomic dN/dS ratios is close to 0 (overall average of 0.0947±0.003sd), indicative of the large prevalence of purifying selection. We observed above average dN/dS ratios in 3 of the~~all~~ 4 available eusocial clades: Formicidae (0.1068 ± 0.0093sd, 3 species), Polistinae/Vespinae wasps (0.1033 ± 0.0088sd, 3 species), and~~Stenogastrinae (0.0951, 1 species only),~~ the *Apis/Bombus/Tetragonula* clade (0.1086 ± 0.0352sd). This last clade of bees does not clearly stand out however, as most bees in the dataset (Anthophila, species characterized by pollen feeding of larvae: Apidae, Megachilidae, Halictidae, Colettidae, Andrenidae, and Melittidae) show high dN/dS ratios (0.1190 ± 0.0302sd, 41 species) with no dependence on their social organization. Finally, only two purely solitary taxa displayed comparable dN/dS ratios: Siricoidea (0.1025 ± 0.0251sd, 3 species) and Cynipoidea (0.1005 ± 0.0175sd, 5 species).

We further used simple linear modeling to try and relate variation in dN/dS ratios to life history traits and geographical range descriptors. Phylogenetic independent contrasts were used to transform the data and account for phylogenetic relationships (Felsenstein 1985). We also used terminal branch length as a covariable in all models. This is because short terminal branches are known to bias dN/dS estimations upward as they yield more inaccurate estimations of this parameter, whose~~ich~~ real value is often close to its zero boundary at a genomic scale. There is strong association between dN/dS ratios and branch length in this study (table 1). Variation in dN/dS estimation accuracy can also stem from variation in the number ~~-~~of genes available for each species. For example, four of the 10 available eusocial Hymenoptera (Apis mellifera and the three available ants), are species with published and annotated genomes (Consortium and The Honeybee Genome Sequencing Consortium 2006; Bonasio et al. 2010; Nygaard et al. 2011), and were used by Peters et al. (2017) as reference species for the identification of 1-1 orthologous genes, along with only one solitary reference species, Nasonia vitripennis (Werren et al. 2010). This translated into a relatively better power for gene prediction by Peters et al. (2017) in eusocial species, and thus into a significant~~–~~ (T = 3.0567, df = 9.3549, p-value = 0.01305) over-representation of these eusocial species in alignments (mean number of alignments available per species: 2732.40 ± 88.09sd)~~ as~~ compared to solitary species (2276.7 ± 90.74sd). To control for potential biases originating from varying precision in estimations, we replicated all the analyses of this study using a balanced subsampled dataset containing 134 alignments, each of them containing data for the same 88 species (most represented half of the species, referred later as the 88-species dataset). Unless specified otherwise, presented results were obtained with the full dataset. Average corrected genomic dN/dS estimated using the 88-species subsampled dataset are displayed along the Hymenoptera tree i~~o~~n figure S2.

No effect of body size, parasitism and geographical range on relative protein evolution rates

Unlike findings in birds and mammals (Figuet et al. 2016; Botero-Castro et al. 2017), we found no significant effect of body size on dN/dS ratio in Hymenoptera (table 1). When testing for a difference in dN/dS ratios between parasitic (parasitoid or parasites) and free-living Hymenoptera (see table S3), we found a significant effect (df= 167, F= 46.327, p-val = 1.715e-10, R^2 = 0.217~~2~~), but which completely disappeared~~s~~ when taking phylogeny into account (df= 166, F= 1.211~~1~~, p-val = 0.272~~7~~, R^2 = 0.007~~243~~). We thus interpret this as being a confounding effect of sampling disequilibrium, as groups with elevated ratios completely lack parasites (with the exception of the cuckoo bumblebee *Bombus rupestris and Sphecodes albilabris*), and discarded this grouping from our models. We further tried to~~and~~ test for an association between dN/dS ratios of reproductive strategy and diet specialization within parasitoids using life-history and host range descriptors found in the lit~~t~~erature (Traynor and Mayhew 2005a, Traynor and Mayhew~~[b]~~ 2005b; Jervis et al.~~, Ferns, and Heimpel~~ 2003; Mayhew 2016), and summarized in table S1. However these descriptors were very seldom available for the species contained in the present phylogenomic dataset, forcing us to use genus-level averaging for both traits and dN/dS ratios. We detected no significant associations between average dN/dS ratios and life-history in parasitoids at the genus-level. We also tested for an association between dN/dS ratios and four proxies of species geographical range obtained using occurrence data available on the GBIF database. dN/dS ratios showed no significant correlation with mean latitude of occurrences, maximal distance between occurrences, or two additional estimators of species range (table S5).

Anthophila bees and eusocial taxa reveal relaxed~~display relaxation of~~ selection at the genomic scale

High dN/dS ratios in Anthophila bees is by far the strongest pattern observed in our results. Treating~~ appartenance to~~ Anthophila as a covariable allows to significantly explain (df = 167; F = 175.84; p-val < 2.2.10-16) more than half the observed variation (R2 = 0.512~~9~~). Despite Anthophila being only one monophyletic group, this effect is still present when accounting for phylogeny (table 1), and when accounting for sampling effort variation by using the 88-species subsampled dataset (table S3). This effect is strong enough to completely mask the effect of eusociality when using the full dataset. Indeed, the social status of a terminal branch significantly explains dN/dS variations in the dataset only if removing all Anthophila samples from the analysis. This is because eusocial corbiculate bees do not show any increase in dN/dS values when compared to other Anthophila. The increase of dN/dS in ants and eusocial wasps, remains significant when accounting for sampling effort variation by using the 88-species subsampled dataset (table S3).

To ensure that previous results stem from a relaxation of selection and not from strong positive selection, we applied the Hyphy RELAX procedure (Pond et al. 2005; Wertheim et al. 2015) on each available alignment separately. This procedure allows to formally test for selection relaxation by modelling the distribution of dN/dS ratios along the branches a phylogeny and by comparing the distribution fitted on a focal group of branches (eusocial taxa and Anthophila, alternatively) to the distribution fitted for the rest of the tree. Out of 3236 realized tests, 1743 (53.9%) detected relaxed selection on eusocial branches (including eusocial bees) and 184 (5.7%) detected intensified selection. Genes under relaxation of selection thus represent 90% of the genes for which a difference of selection efficiency between eusocial branches and focal branches could be detected. Results of a gene ontology enrichment analysis conducted with genes under intensified selection in eusocial species as focal genes are presented in table S4, but revealed no clear pattern. Using a conservative bonferroni correction for multiple testing in this procedure still leads to the detection of selection relaxation in 751 genes and of selection intensification in 28 genes. These results also hold if the more balanced 88-species subsampled dataset is used, as out of 134 alignments, 68 genes supported a relaxation of selection and 16 genes supported an intensification of selection. Moreover, the detected effect of eusociality does not seem to be driven by any over-representation of bees within eusocial species. The average number of eusocial bee sequences available for genes with relaxed selection (2.427 ±~~+-~~ 0.018sd) is not different than within genes without relaxed selection (2.463 ±~~+-~~ 0.024sd) (F=2.11; pval=0.146). These verifications are needed as bees experience an even stronger relaxation of selection. If this was apparent from simple modelling of genomic dN/dS ratios, it is made even more obvious by the application of the RELAX procedure with Anthophila branches as focal branches. Out of 3239 realized tests, 2000 (61.74%) detected relaxed selection on eusocial branches, while 294 detected an intensification of selection (9.07%). Using a conservative bonferroni correction for multiple testing in this procedure still leads to the detection of selection relaxation in 1210 genes and of selection intensification in 66 genes.

DISCUSSION

Molecular consequences of body-size, parasitism and eusociality

Contrary to observed patterns in vertebrates (Nikolaev et al. 2007; Romiguier et al. 2014; Popadin et al. 2007; Figuet et al. 2016; Botero-Castro et al. 2017; Rolland et al. 2020), we did not detect any significant effect of body size on dN/dS in Hymenoptera. This suggests that the general association between body size and *Ne* observed in vertebrates is not universal in Metazoa, particularly in Hymenoptera. Parasitism is also not significantly associated to *Ne* decreases, which is surprising given the theoretical constrains imposed by the host (Papkou et al. 2016). This surprising negative result might be partly explained by the fact that Hymenoptera parasitoids lifecycle mostly requires a single insect host, a resource that might be not as limiting as parasites of vertebrates with complex lifecyles (Strobel et al. 2019).

We observed a significantly higher accumulation of non-synonymous substitutions in eusocial genomes, although the effect is relatively modest compared to the global pattern of increased dN/dS in pollinating bees (Anthophila). This increase ~~was associated with the detection of relaxed selection on eusocial branches for more than half of the available alignments. This result supports the hypothesis of a relaxation of selection associated with eusociality through demographic effects, as any other selective effects would be expected to have more heterogenous and localised effects.~~

~~Because eusocial Hymenoptera are known to display very high recombination rates (Wallberg, Glémin, and Webster 2015; Sirviö et al. 2011; Wilfert, Gadau, and Schmid-Hempel 2007; Jones et al. 2019), these results could have been~~can not be imputed to biased gene conversion, because ~~which is known to increase dN/dS by promoting the fixation of any G/C alleles (including deleterious alleles). Biases in dN/dS estimation due to gBGC can be excluded however, as~~ our results are obtained using dN/dS ratios accounting only for GC-conservative substitutions. This result can not be imputed to positive selection either, as RELAX analyses detected relaxed selection on eusocial branches for more than half of the available alignments.~~This is confirmed by RELAX analyses which would be expected to have detected gBGC effects as directional/adaptive selection, because both signals are very similar (Rousselle et al. 2019).~~ This result supports the hypothesis of a relaxation of selection associated with eusociality through demographic effects.

~~We rather suggest that the high dN/dS ratios we report are primarily due to *Ne* reduction and purifying selection relaxation in eusocial lineages.~~ Long-lived reproductive female with delayed sexual maturity, as well as a biased sex-ratio and monopolization of the reproductive labo~~u~~r by few individuals, are typical features of eusocial species, which are bound to reduce effective population size. The hypothesis of a life-history effect matches well with the observation of a higher *dN/dS* in the highly eusocial formicoids ants *Acromyrmex echinatior* and *Camponotus floridanus* than in *Harpegnatos saltator*, which possesses a less complex social organization (Hölldobler and Wilson 1990). ~~Ants, which display a variety of complexity levels in their social organisation, could represent an ideal model for a more quantitative approach (Bourke, 1999), allowing to test for an effect of variation in eusocial characteristics of species on selection efficiency.~~

These results are however to be taken with care, as the number of eusocial species in the dataset is low, and as no significant increase in dN/dS due to eusociality has been detected within bees. In this study, a choice was made not to increment the original dataset with additionnal eusocial species, because this addition would have introduced heterogeneity in sample treatment, and translated into new bias in the estimation of dN/dS ratios. It might be necessary to replicate our analyses using a separate, tailor-made and more exhaustive dataset in terms of eusocial species number in order to confirm the effect of eusociality on demography. Another exciting prospect will be to study *Ne* variation within eusocial groups. Ants, which display a variety of complexity levels in their social organisation, could represent an ideal model for a more quantitative approach (Bourke 1999), allowing to test for an effect of variation in eusocial characteristics of species on selection efficiency.

Ecological and molecular predisposition to eusociality in bees

High genomic dN/dS ratios in all social and solitary bees unexepectedly appears as the major pattern of our results.

~~Interestingly, Anthophila (bees) is the taxa with the highest number of independent origins of eusociality in the tree of life (Hughes et al. 2008). This suggests that high dN/dS might be not only a consequence of group-living, but might also be associated to pre-requisites facilitating evolution toward eusociality. Several ecological characteristics have been suggested as facilitating the emergence of eusociality in the litterature, such as communal behaviors in nesting, the ability to defend such communal nests, parental care, progressive provisioning of food to the larvae or co-occurrence of nesting/feeding sites (Wilson 2008). Some of these ecological characteristics can be linked to *Ne*, such as communal behaviors in nesting and parental care, that theoretically increase within-group relatedness through inbreeding, thus reducing *Ne* and increasing dN/dS (Hamilton 1972; Hussedener et al. 1999, Tabadkani et al., 2012, but see Nowak et al. 2010). Few genomic evidences supporting such a link have been observed so far. By showing a striking increase in dN/dS ratio in all Anthophila bees - the taxa concentrating more than half of the origins of eusociality in the tree of life - our results are the first genomic insight supporting the idea that low-Ne might have preceded and/or favoured evolution towards eusociality. As suggested previously in the litterature, the evolution towards eusociality might have been favoured by the emergence of small groups of inbred individuals, despite the cost associated to genetic diversity loss at the sex determination single locus (Rabeling and Kronauer 2013).~~Besides many independent transitions toward eusociality, Anthophila are also characterized by their pollen-collecting behaviors which might explain our results. This dependence to large amounts of pollen to feed ~~their~~ larvae is indeed believed to be a potential constraint on *Ne*, particularly in specialist species (Zayed and Packer 2007). Pollen is a resource which is scattered in space and time and require a large energetic investment to come by and exploit (through progressive provisioning), thus constraining the very fecundity of females, which invest a lot of time and energy in their descent. Parental investment has already been highlighted as the major determinant of genetic diversity and long-term ***Ne*** in animal species (Romiguier et al. 2014). We suggest that high parental investment in pollinating bees might be a major factor limiting their ***Ne***. This could in turn provide an explanation for the absence of differences between dN/dS ratios in social and solitary pollen-collecting species. Group-living might indeed represent a way to enhance the productivity of pollen collecting and metabolizing, thus compensating the decrease of *Ne* linked to eusociality in Anthophila. ~~Measuring parental investment through propagule size leaving parents (as in Romiguier et al. 2014b) of every Hymenoptera species would be a way to confirm these hypotheses, but such precise life-history data are for now too scarce in this taxa. Additionally, it could be argued that constraints on *Ne* imposed by pollen scarcity might be less important in tropical species, as pollen is less limitant in these regions. Similarly to Romiguier et al (2014b), we found no associations between geographical distribution and Ne. Mean latitude and dN/dS ratios were not significantly associated in Anthophila, but this might be due to the over-representation of species from temperate regions in our dataset.~~

Contrary to the general pattern in animals, pollinating bees appears as an exception and display higher species richness at high latitudes compared to tropics (Orr et al. 2020). This suggests that the diversification and origin of a pollinating bee lifestyle stems to environments with strong seasonality and important long-term climatic oscillations, which might have led to frequent bottlenecks in their population history. One previous study in Teleost fishes has shown that species of temperate regions display lower *Ne* than species of tropical regions (Rolland et al. 2020), while another found no such link across Metazoans (Romiguier et al. 2014b). Similarly to the latter, we found no associations between mean latitude (or other range descriptors) and dN/dS ratios in Anthophila (see table S5). However this result might be simply due to the massive over-representation of species from temperate regions in our dataset, and more thorough studies focusing on more tropical species will be necessary to draw any conclusions. In any case, specialized feeding on flowers appears here as a specialization to ecosystems with relatively low carrying capacity (Orr et al. 2020) requiring high parental investment for a scarce resource. Pollinating bees thus represent an ideal model to study the links between long-term demographics and seasonal variation in resource availability in temperate or arid environments. Anthophila could be also used to formally test whether the degree of specialization of a species towards one or a few plant species constrains *Ne* at the genomic scale in the general case (Zayed et al. 2005). Finally, bees might represent an opportunity to gain novel insights about the links between long-term demographics and characteristics more specific to Hymenoptera, such as nest parasite load (Wcislo 1987).

Interestingly, besides pollen-collecting, Anthophila (bees) is also the taxa with the highest number of independent origins of eusociality in the tree of life (Hughes et al. 2008). This suggests that low *Ne* is not only a consequence of group-living, but might also facilitate evolution toward eusociality. Supporting this hypothesis, low *Ne* due to intense inbreeding has been associated to communal behaviors in nesting and high parental care, two classical pre-requesites to an eusocial lifestyle (Hamilton 1964; Hussedener et al. 1999; Tabadkani et al. 2012; Wilson et al. 2008). Inbreeding also tends to increase within-group relatedness, which theoretically increases the benefit of kin selection, potentially favouring the emergence of eusociality (Hamilton 1964; Hussedener et al. 1999; Tabadkani et al. 2012; Kay et al. 2020; but see Nowak et al. 2010). Few genomic evidences supporting such a link have been observed so far. By showing a striking increase in dN/dS ratio in all Anthophila bees - the taxa concentrating more than half of the origins of eusociality in the tree of life - our results are the first genomic insight supporting the idea that low-*Ne* might have preceded and/or favoured evolution towards eusociality. As suggested previously in the literature, the evolution towards eusociality might have been favoured by the emergence of small groups of inbred individuals, despite the cost associated to genetic diversity loss at the sex determination single locus (Rabeling and Kronauer 2013).

Besides their implication regarding the evolution of eusociality, our results have important consequences for the conservation field. P~~Indeed, p~~ollination has been found to rely heavily on wild and domesticated bees, which ensure the majority of animal-mediated pollination of wild and domesticated plants in most ecosystems (Winfree 2010). Our finding of particularly high deleterious substitution rates within this group raises the additional concern that bee species might be especially sensitive to any further population decline, which are already known as particularly alarming (~~(Powney et al. 2019); (~~Murray et al.~~, Kuhlmann, and Potts~~ 2009; Arbetman et al. 2017; Powney et al. 2019).

Conclusion

This study brings supplemental~~new~~ genomic evidence supporting~~ the existence of~~ an association between eusociality and reduced~~-~~ effective population size. We thus bring further support to the hypothesis that the extreme life-history traits of eusocial species constrain their molecular evolution. More i~~I~~nterestingly, ~~the observation of a~~ the surprisingly massive and widespread reduction of selection efficiency in both eusocial and solitary bees suggests unexpectedly high constraints of a pollinator lifestyle, potentially linked to limiting resource and high parental investment. This also brings genomic support to the hypothesis that some ecological characteristics associated with low ***Ne*** might have facilitated evolution towards eusociality. Altogether, this study suggests that, ~~as in~~conversely to v~~V~~ertebrates, purifying selection efficiency in invertebrates is more constrained by lifestyle and~~the~~ ecology than siple body size~~and life-history strategies of species~~.

Acknowledgment

We thanks Nicolas Galtier for advices during the writing of the manuscript and Laurent Keller for useful discussions.

Data Availability

The original dataset of Peters *et al. (2017),* with alignments and trees, is available at *http://dx.doi.org/10.17632/trbj94zm2n.2*. Detailed tables containing data used for this paper as well as obtained results are available at Zenodo.org : https://zenodo.org/record/3999857#.X0UsBBk6-it.

Conflict of interest disclosure

The authors of this article declare that they have no financial conflict of interest with the content of this article.

MATERIALS AND METHODS

Genetic data

Data was downloaded from the authors' online repository (http://dx.doi.org/10.17632/trbj94zm2n.2). It originally contained nucleotide and amino-acids multi-sample alignments for 3256 protein coding genes predicted to be 1-1 orthologs in 174 species (see Peters et al. 2017 for details about the production of these alignments), 5 of which are outgroups to the Hymenoptera (2 Coleoptera, 1 Megaloptera, 1 Neuroptera and 1 Raphidioptera), and 10 of which are eusocial species. The latter belong to 5 independent eusocial clades: corbiculate bees (*Tetragonula carbonaria*, *Bombus rupestris* and *Apis mellifera*), ants (*Acromyrmex echinatior*, *Camponotus floridanus* and *Harpegnathos saltator*), Polistinae/Vespinae wasps (*Vespa crabro*, *Vespula germanica* and *Polistes dominula*), Stenogastrinae wasps (*Parischnogaster nigricans*). (Cardinal and Danforth 2011). The data also contained the trees inferred using this data by the original authors. We used the dated chronogram inferred by the authors using amino-acid data throughout this study. This tree corresponds to their main results and is contained in the file dated_tree_aa_inde_2_used_in_Fig1.tre available on the authors' online repository.

Data cleaning

Each amino-acid alignment was first checked for potential false homology using HmmCleaner (Di Franco et al. 2019; Philippe et al. 2017) with default settings. The resulting maskings were then reported on corresponding nucleotide sequences using the reportMaskAA2NT program from the MASCE program suite (Ranwez et al. 2011). At this point, we discarded individual sequences containing less than 50% of informative site within one alignment.

dN/dS ratios estimation

Cleaned alignments were then used, along with the tree topology inferred by Peters et al. (2017) and the mapNH binary (Romiguier et al 2012, https://github.com/BioPP/testnh), to estimate synonymous and non-synonymous substitution rates along the branches of the Hymenoptera tree. MapNH allows a fast estimation of those rates by using tree-wide parameters obtained *a priori* by fitting a homogeneous model (YN98) to the data with the help of paml (Yang 1997), to parsimoniously map observed substitutions to the supplied topology. Estimated substitution counts for specific branches, obtained separately for each alignments, can then be summed to obtain genome-wide substitution rates. We used this method to obtain dN/dS ratios of terminal branches, susceptible to carry information about the long-term drift regime of extant lineages. 15 alignments did not contain enough data to allow correct convergence of the homogeneous model needed by mapNH.

Controlling for biased gene conversion

We produced a corrected dN/dS using only GC conservative substitutions to estimate dN/dS. This was achieved using a custom version of mapNH developed in our lab (Rousselle et al. 2019) which categorizes mapped substitutions into GC-conservative (GC->GC or AT->AT) and GC-modifying (AT->GC or GC->AT) substitutions, and uses only the former to compute dN/dS ratios. Ratios obtained this way show more sampling variance, as they are obtained from smaller substitution counts. This translates in higher genomic dN/dS, as this parameter is usually close to its zero bound in exons. These rates are however supposedly impervious to gBGC.

Controlling for sampling bias

Four Hymenoptera (*Apis mellifera* and the three ants), which represent nearly half the eusocial species considered, are species with published genomes. This translates into a better power for gene prediction and thus, into an over-representation of these species in the dataset. Imprecisions in dN/dS ratios estimations are in turn known to yield higher values, because the real value of this ratio in functional sequences is often close to its zero boundary. We thus applied an additional sub-sampling procedure, designed to correct for any potential bias in our estimations that could stem from variation in the quantity of information available for each species. We applied every analysis mentioned before to a reduced but complete dataset containing data only for the most represented half of the species (88 species), and only alignments containing information for each of these species (135 alignments).

Linear modelling of dN/dS ratios

Estimated rates, corrected rates and rates obtained from the 88-species subsampled dataset were then modelled through simple linear models using the R software environment, using adult size, social status (eusocial or solitary) and membership to Anthophila as covariables. We also used this statistical setting to evaluate the effect of branch length. Short branches are known to bias dN/dS estimations upward because they yield more inaccurate and thus generally higher estimations of this parameter. The phylogenetic setting was taken into account by using phylogenetic independent contrast (Felsenstein 1981) for each variable. This was done using the pic() function in the R package ape. To try and further uncover the potential links between dN/dS ratios and life-history within Hymenoptera, we also attempted to correlate dN/dS ratios with major descriptors of parasitic type within parasitoid Hymenoptera. These descriptors were gathered from databases designed to describe the reproductive strategy of parasitoids (Traynor & Mayhew 2005a, Traynor & Mayhew 2005b, Jervis & Ferns 2011; Mayhew 2016) and are summarized in table S1. We conducted the analysis at the genus level using genus-averaged dN/dS ratios and descriptors. This was necessary because the species-level concordance between databases was too low (only 6 species in common between the genomic database and the parasitoid life-history database). We used Pearson’s linear correlation coefficient for continuous descriptors and Kruskal-Wallis tests for discrete descriptors.

Finally, we tested the correlation of dN/dS ratios with four proxies of species range. For each species (and for all known synonyms) in the sample, we queried all available occurrence points from the GBIF database, using the R package rgbif. Occurrence data was then used to calculate for classical proxies of species range. The mean latitude was calculated as a simple unweighted mean between occurrences. The maximum distance between two occurrences was calculated taking all occurrences into account, even when the species occurred on more than one continent. The circular area around occurrence was calculated by casting 100km-radius circles around each occurence, and estimating the total land surface contained in at least one circle. The convex hull area around occurrence was calculated by estimating the total land surface contained in the smallest convex hull containing all occurrences. When a species occurred on more than one continent, a separate convex hull was used per continent.

RELAX analyses

We used the RELAX procedure (Wertheim et al. 2015) from the HyPhy program suite (Pond, Frost, and Muse 2005) to test for the presence of a systematic relaxation of selection on branches belonging to eusocial groups (thereafter called “eusocial branches”), that is all branches descending from the ancestral node of one of the eusocial clade present in the dataset. Hyphy allows, for a specific sequence alignment, to model the distribution of dN/dS ratios along the branches of a tree. The RELAX procedure consists first in defining focal and background branches, associated with one focal and one background distribution of dN/dS ratios. It then consists in comparing a model where the two mentioned distribution are identical (null model, no differences between branch sets) to a model where the focal distribution is a power transform of the background distribution ($\boldsymbol{\omega_f=\omega_b^k}$). Relaxation of selection is inferred when the second model appears superior based on a log-ratio test (differences between branch sets), and when the focal distribution is narrower than the background distribution (k parameter estimated to be less than 1). Indeed, strong selection is thought to produce both low (close to 0) and high (greater than 1) dN/dS ratios, while neutrality should produce rates close to 1. This test thus correctly takes into account the fundamental two-sided nature of dN/dS ratios. 20 (0.61%) of the original alignments did not contain enough data to allow models necessary to the HyPhy RELAX procedure to be fitted with eusocial branches as background branches, and 17 (0.52%) of the original alignments didn’t allow the procedure with Anthophila branches as background branches.

REFERENCES

Arbetman, Marina P., Gabriela Gleiser, Carolina L. Morales, Paul Williams, and Marcelo A. Aizen. 2017. “Global Decline of Bumblebees Is Phylogenetically Structured and Inversely Related to Species Range Size and Pathogen Incidence.” *Proceedings of the Royal Society B: Biological Sciences* 284 (1859). https://doi.org/10.1098/rspb.2017.0204.

Asplen, M. K., J. B. Whitfield, J. G. DE Boer, and G. E. Heimpel. 2009. “Ancestral State Reconstruction Analysis of Hymenopteran Sex Determination Mechanisms.” *Journal of Evolutionary Biology* 22 (8): 1762–69.

Berkelhamer, Rudi C. 1983. “INTRASPECIFIC GENETIC VARIATION AND HAPLODIPLOIDY, EUSOCIALITY, AND POLYGYNY IN THE HYMENOPTERA.” *Evolution; International Journal of Organic Evolution* 37 (3): 540–45.

Bonasio, Roberto, Guojie Zhang, Chaoyang Ye, Navdeep S. Mutti, Xiaodong Fang, Nan Qin, Greg Donahue, et al. 2010. “Genomic Comparison of the Ants Camponotus Floridanus and Harpegnathos Saltator.” *Science* 329 (5995): 1068–71.

Botero-Castro, Fidel, Emeric Figuet, Marie-Ka Tilak, Benoit Nabholz, and Nicolas Galtier. 2017. “Avian Genomes Revisited: Hidden Genes Uncovered and the Rates versus Traits Paradox in Birds.” *Molecular Biology and Evolution* 34 (12): 3123–31.

Bromham, Lindell, and Remko Leys. 2005. “Sociality and the Rate of Molecular Evolution.” *Molecular Biology and Evolution* 22 (6): 1393–1402.

Cardinal, Sophie, and Bryan N. Danforth. 2011. “The Antiquity and Evolutionary History of Social Behavior in Bees.” *PLoS ONE*. https://doi.org/10.1371/journal.pone.0021086.

Consortium, The Honeybee Genome Sequencing, and The Honeybee Genome Sequencing Consortium. 2006. “Insights into Social Insects from the Genome of the Honeybee Apis Mellifera.” *Nature*. https://doi.org/10.1038/nature05260.

Crespi, Bernard J., and Douglas Yanega. 1995. “The Definition of Eusociality.” *Behavioral Ecology*. https://doi.org/10.1093/beheco/6.1.109.

Di Franco, Arnaud, Raphaël Poujol, Denis Baurain, and Hervé Philippe. 2019. “Evaluating the Usefulness of Alignment Filtering Methods to Reduce the Impact of Errors on Evolutionary Inferences.” *BMC Evolutionary Biology* 19 (1): 21.

Duret, Laurent, and Nicolas Galtier. 2009. “Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes.” *Annual Review of Genomics and Human Genetics* 10: 285–311.

Dutheil, Julien, and Bastien Boussau. 2008. “Non-Homogeneous Models of Sequence Evolution in the Bio+

+ Suite of Libraries and Programs." *BMC Evolutionary Biology* 8 (September): 255.

Felsenstein, Joseph. 1985. "Phylogenies and the Comparative Method." *The American Naturalist*. https://doi.org/10.1086/284325.

Figuet, Emeric, Benoît Nabholz, Manon Bonneau, Eduard Mas Carrio, Krystyna Nadachowska-Brzyska, Hans Ellegren, and Nicolas Galtier. 2016. "Life History Traits, Protein Evolution, and the Nearly Neutral Theory in Amniotes." *Molecular Biology and Evolution* 33 (6): 1517–27.

Galtier, Nicolas, Camille Roux, Marjolaine Rousselle, Jonathan Romiguier, Emeric Figuet, Sylvain Glémin, Nicolas Bierne, and Laurent Duret. 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 (5): 1092–1103.

Guéguen, Laurent, and Laurent Duret. 2018. "Unbiased Estimate of Synonymous and Nonsynonymous Substitution Rates with Nonstationary Base Composition." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msx308.

Hamilton, W. D. 1972. "Altruism and Related Phenomena, Mainly in Social Insects." *Annual Review of Ecology and Systematics*. https://doi.org/10.1146/annurev.es.03.110172.001205.

Hölldobler, Bert, and Edward O. Wilson. 1990. "The Ants." https://doi.org/10.1007/978-3-662-10306-7.

Hughes, W. O. 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*. https://doi.org/10.1126/science.1156108.

Husseneder, Husseneder, J. T. Epplen, Brandl, Epplen, and Kaib. 1999. "WITHIN-COLONY RELATEDNESS IN A TERMITE SPECIES: GENETIC ROADS TO EUSOCIALITY?" *Behaviour*. https://doi.org/10.1163/156853999501702.

Jervis, M. A., P. N. Ferns, and G. E. Heimpel. 2003. "Body Size and the Timing of Egg Production in Parasitoid Wasps: A Comparative Analysis." *Functional Ecology*. https://doi.org/10.1046/j.1365-2435.2003.00742.x.

Jones, Julia C., Andreas Wallberg, Matthew J. Christmas, Karen M. Kapheim, and Matthew T. Webster. 2019. "Extreme Differences in Recombination Rate between the Genomes of a Solitary and a Social Bee." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msz130.

Keller, Laurent, and Michel Genoud. 1997. "Extraordinary Lifespans in Ants: A Test of Evolutionary Theories of Ageing." *Nature*. https://doi.org/10.1038/40130.

Lartillot, Nicolas. 2013. "Phylogenetic Patterns of GC-Biased Gene Conversion in Placental Mammals and the Evolutionary Dynamics of Recombination Landscapes." *Molecular Biology and Evolution* 30 (3): 489–502.

Lynch, Michael. 2007. "The Frailty of Adaptive Hypotheses for the Origins of Organismal Complexity." *Proceedings of the National Academy of Sciences of the United States of America* 104 Suppl 1 (May): 8597–8604.

Lynch, Michael, and John S. Conery. 2003. "The Origins of Genome Complexity." *Science*. https://doi.org/10.1126/science.1089370.

Mayhew, Peter J. 2016. "Comparing Parasitoid Life Histories." *Entomologia Experimentalis et Applicata*. https://doi.org/10.1111/eea.12411.

Mazé-Guilmo, Elise, Simon Blanchet, Karen D. McCoy, and Géraldine Loot. 2016. "Host Dispersal as the Driver of Parasite Genetic Structure: A Paradigm Lost?" *Ecology Letters* 19 (3): 336–47.

Murray, Tomás E., Michael Kuhlmann, and Simon G. Potts. 2009. "Conservation Ecology of Bees: Populations, Species and Communities." *Apidologie*. https://doi.org/10.1051/apido/2009015.

Nikolaev, Sergey I., Juan I. Montoya-Burgos, Konstantin Popadin, Leila Parand, Elliott H. Margulies, National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program, and Stylianos E. Antonarakis. 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 (51): 20443–48.

Nygaard, Sanne, Guojie Zhang, Morten Schiøtt, Cai Li, Yannick Wurm, Haofu Hu, Jiajian Zhou, et al. 2011. "The Genome of the Leaf-Cutting Ant Acromyrmex Echinatior Suggests Key Adaptations to Advanced Social Life and Fungus Farming." *Genome Research* 21 (8): 1339–48.

Owen, Robin E. 1985. "DIFFICULTIES WITH THE INTERPRETATION OF PATTERNS OF GENETIC VARIATION IN THE EUSOCIAL HYMENOPTERA." *Evolution*. https://doi.org/10.1111/j.1558-5646.1985.tb04093.x.

Peters, Ralph S., Lars Krogmann, Christoph Mayer, Alexander Donath, Simon Gunkel, Karen Meusemann, Alexey Kozlov, et al. 2017. "Evolutionary History of the Hymenoptera." *Current Biology: CB* 27 (7): 1013–18.

Philippe, Hervé, Damien M. de Vienne, Vincent Ranwez, Béatrice Roure, Denis Baurain, and Frédéric Delsuc. 2017. “Pitfalls in Supermatrix Phylogenomics.” *European Journal of Taxonomy*. https://doi.org/10.5852/ejt.2017.283.

Pond, Sergei L. Kosakovsky, Simon D. W. Frost, and Spencer V. Muse. 2005. “HyPhy: Hypothesis Testing Using Phylogenies.” *Bioinformatics* 21 (5): 676–79.

Popadin, Konstantin, Leonard V. Polishchuk, Leila Mamirova, Dmitry Knorre, and Konstantin Gunbin. 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 (33): 13390–95.

Powney, Gary D., Claire Carvell, Mike Edwards, Roger K. A. Morris, Helen E. Roy, Ben A. Woodcock, and Nick J. B. Isaac. 2019. “Widespread Losses of Pollinating Insects in Britain.” *Nature Communications* 10 (1): 1–6.

Rabeling, Christian, and Daniel J. C. Kronauer. 2013. “Thelytokous Parthenogenesis in Eusocial Hymenoptera.” *Annual Review of Entomology* 58: 273–92.

Rahnenfuhrer, Jorg, and Adrian Alexa. 2019. *topGO: Enrichment Analysis for Gene Ontology* (version 2.38.1.).

Ranwez, Vincent, Sébastien Harispe, Frédéric Delsuc, and Emmanuel J. P. Douzery. 2011. “MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons.” *PloS One* 6 (9): e22594.

Reeve, Hudson K., Jan Shellman Reeve, and David W. Pfennig. 1985. “EUSOCIALITY AND GENETIC VARIABILITY: A RE-EVALUATION.” *Evolution; International Journal of Organic Evolution* 39 (1): 200–201.

Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, et al. 2014. “Comparative Population Genomics in Animals Uncovers the Determinants of Genetic Diversity.” *Nature* 515 (7526): 261–63.

Romiguier, J., J. Lourenco, P. Gayral, N. Faivre, L. A. Weinert, S. Ravel, M. Ballenghien, et al. 2014. “Population Genomics of Eusocial Insects: The Costs of a Vertebrate-like Effective Population Size.” *Journal of Evolutionary Biology*. https://doi.org/10.1111/jeb.12331.

Rousselle, Marjolaine, Alexandre Laverré, Emeric Figuet, Benoit Nabholz, and Nicolas Galtier. 2019. “Influence of Recombination and GC-Biased Gene Conversion on the Adaptive and Nonadaptive Substitution Rate in Mammals versus Birds.” *Molecular Biology and Evolution* 36 (3): 458–71.

Settepani, Virginia, Jesper Bechsgaard, and Trine Bilde. 2016. “Phylogenetic Analysis Suggests That Sociality Is Associated with Reduced Effectiveness of Selection.” *Ecology and Evolution* 6 (2): 469–77.

Sirviö, Anu, J. Spencer Johnston, Tom Wenseleers, and Pekka Pamilo. 2011. “A High Recombination Rate in Eusocial Hymenoptera: Evidence from the Common Wasp Vespula Vulgaris.” *BMC Genetics* 12 (November): 95.

Tabadkani, Seyed Mohammad, Jamasb Nozari, and Mathieu Lihoreau. 2012. “Inbreeding and the Evolution of Sociality in Arthropods.” *Die Naturwissenschaften* 99 (10): 779–88.

Traynor, Ruth E., and Peter J. Mayhew. 2005a. “A Comparative Study of Body Size and Clutch Size across the Parasitoid Hymenoptera.” *Oikos*. https://doi.org/10.1111/j.0030-1299.2005.13666.x.

———. 2005b. “Host Range in Solitary versus Gregarious Parasitoids: A Laboratory Experiment.” *Entomologia Experimentalis et Applicata*. https://doi.org/10.1111/j.1570-7458.2005.00331.x.

Wallberg, Andreas, Sylvain Glémin, and Matthew T. Webster. 2015. “Extreme Recombination Frequencies Shape Genome Variation and Evolution in the Honeybee, Apis Mellifera.” *PLOS Genetics*. https://doi.org/10.1371/journal.pgen.1005189.

Wang, J., E. Santiago, and A. Caballero. 2016. “Prediction and Estimation of Effective Population Size.” *Heredity*. https://doi.org/10.1038/hdy.2016.43.

Werren, John H., Stephen Richards, Christopher A. Desjardins, Oliver Niehuis, Jürgen Gadau, John K. Colbourne, Nasonia Genome Working Group, et al. 2010. “Functional and Evolutionary Insights from the Genomes of Three Parasitoid Nasonia Species.” *Science* 327 (5963): 343–48.

Wertheim, Joel O., Ben Murrell, Martin D. Smith, Sergei L. Kosakovsky Pond, and Konrad Scheffler. 2015. “RELAX: Detecting Relaxed Selection in a Phylogenetic Framework.” *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msu400.

Wilfert, L., J. Gadau, and P. Schmid-Hempel. 2007. “Variation in Genomic Recombination Rates among Animal Taxa and the Case of Social Insects.” *Heredity* 98 (4): 189–97.

Wilson, Edward O. 2008. “One Giant Leap: How Insects Achieved Altruism and Colonial Life.” *BioScience*. https://doi.org/10.1641/b580106.

Winfree, Rachael. 2010. “The Conservation and Restoration of Wild Bees.” *Annals of the New York Academy of Sciences* 1195 (May): 169–97.

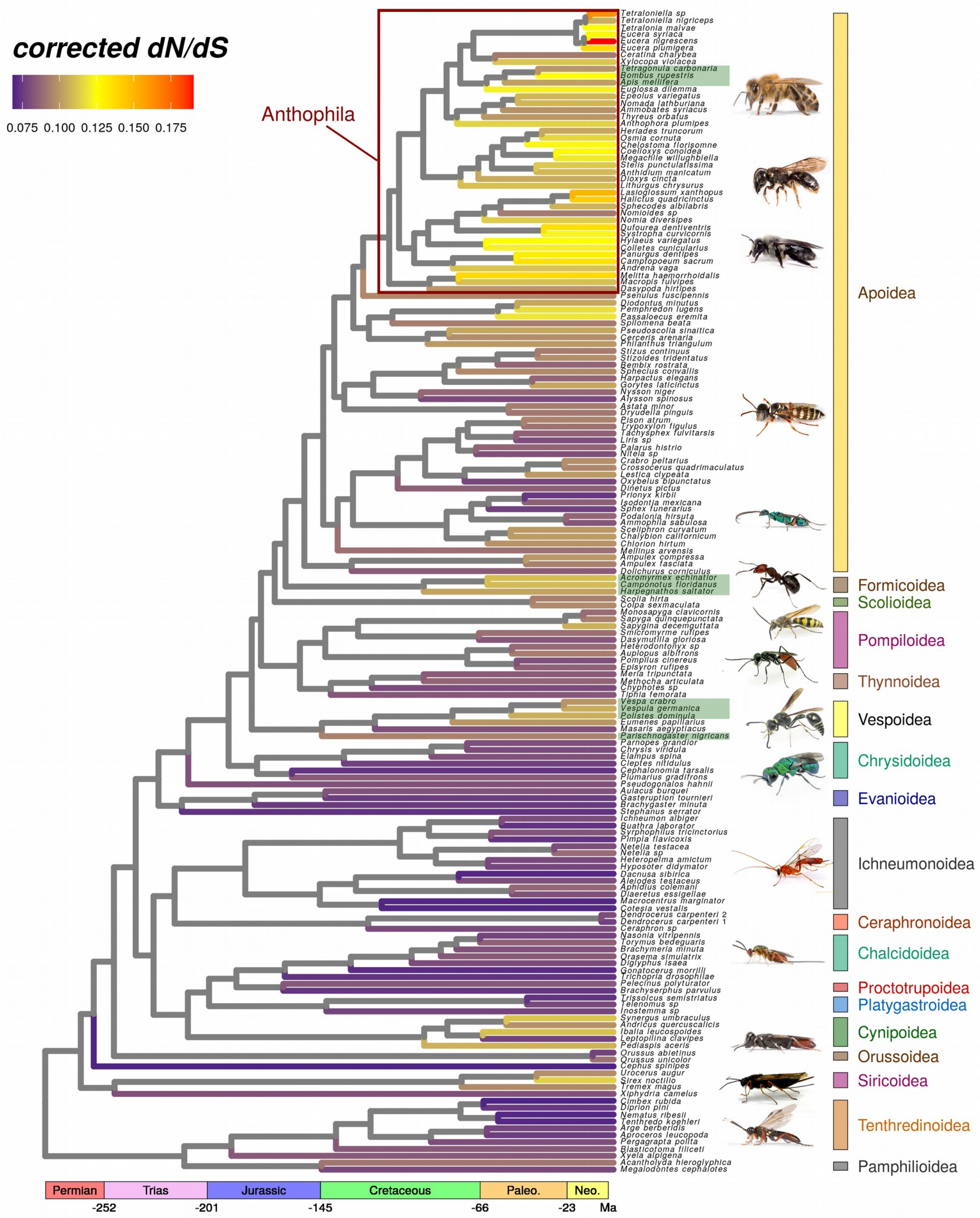
Wright, S. 1931. “Evolution in Mendelian Populations.” *Genetics* 16 (2): 97–159.

Yang, Ziheng. 1997. “PAML: A Program Package for Phylogenetic Analysis by Maximum Likelihood.” *Bioinformatics*. https://doi.org/10.1093/bioinformatics/13.5.555.

Zayed, A., and L. Packer. 2007. “The Population Genetics of a Solitary Oligolectic Sweat Bee, Lasioglossum (Sphecodogastra) Oenotherae (Hymenoptera: Halictidae).” *Heredity* 99 (4): 397–405.

	All samples residual df = 127; R^2=0.082			**Non-Anthophila samples** residual df = 97; R^2=0.117		
covariables	R^2	F	p-value	R^2	F	p-value
branch length	0.0347	4.8040	**0.0302**	0.0751	8.2422	**0.0051**
adult size	0.0085	1.1793	0.2795	0.0048	0.5363	0.4657
Anthophila	0.0381	5.2677	**0.0233**			
Eusociality	0.0008	0.1232	0.7261	0.0372	4.0915	**0.0458**

Table 1: Linear modelling of corrected dN/dS ratios. Corrected dN/dS are obtained using 3241 genes and GC-conservative substitutions only. Displayed results are obtained when simultaneously using all covariables inside a multiple linear model. Phylogenetic independent contrasts are used for all variables so as to account for phylogenetic autocorrelation.



645 **Figure 1: Corrected genomic dN/dS ratios for 169 Hymenoptera species.** dN/dS ratios estimated on terminal branches using 3241 genes are represented on the chronogram inferred by Peters *et al.* (2017). Green rectangles around labels indicate eusocial taxa.

SUPPLEMENTARY

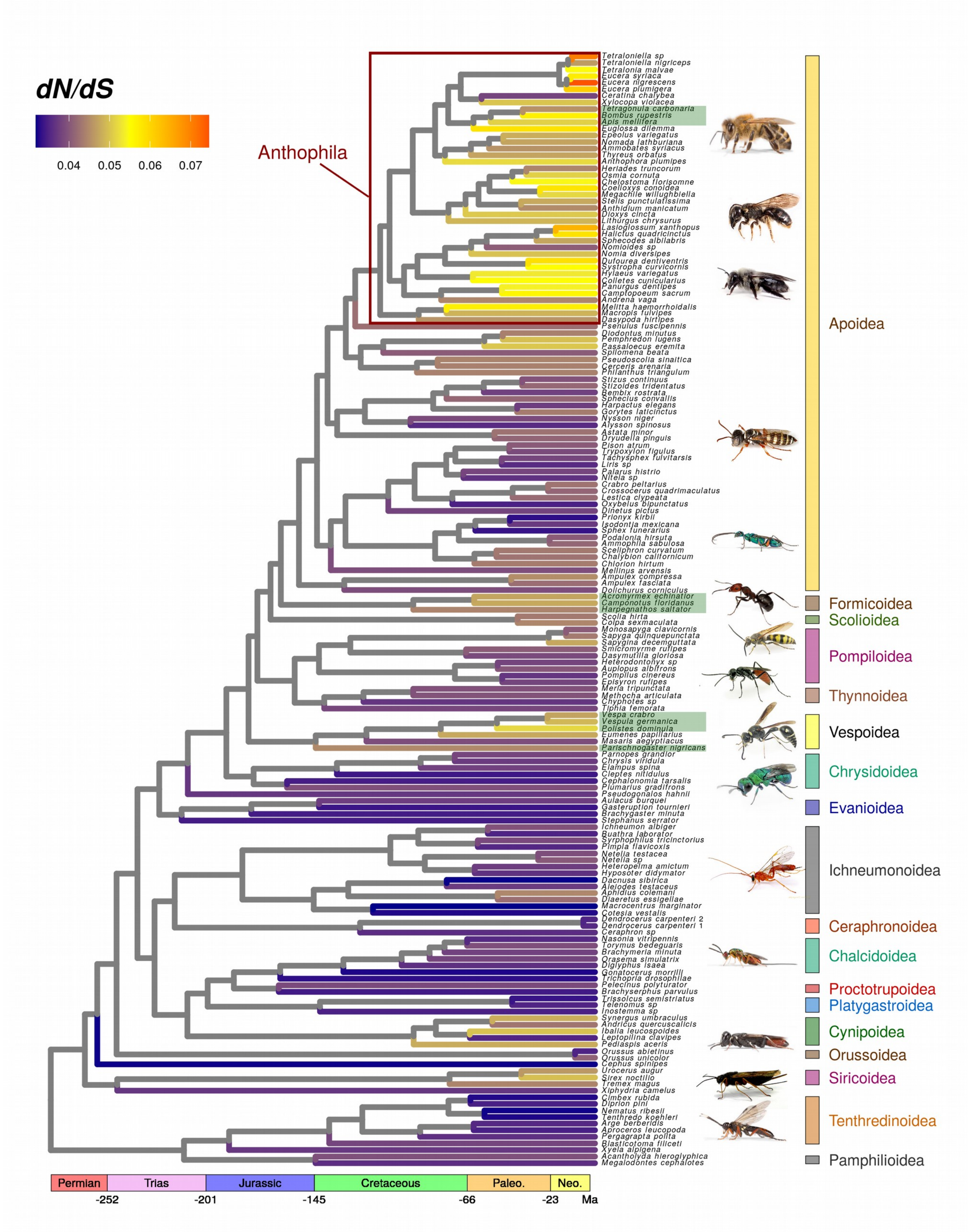


Figure S1: Uncorrected genomic dN/dS ratios for 169 Hymenoptera species. dN/dS ratios estimated on terminal branches using 3241 genes and GC conservative substitutions are represented on the chronogram inferred by Peters *et al.* (2017). Green rectangles around labels indicate eusocial taxa.

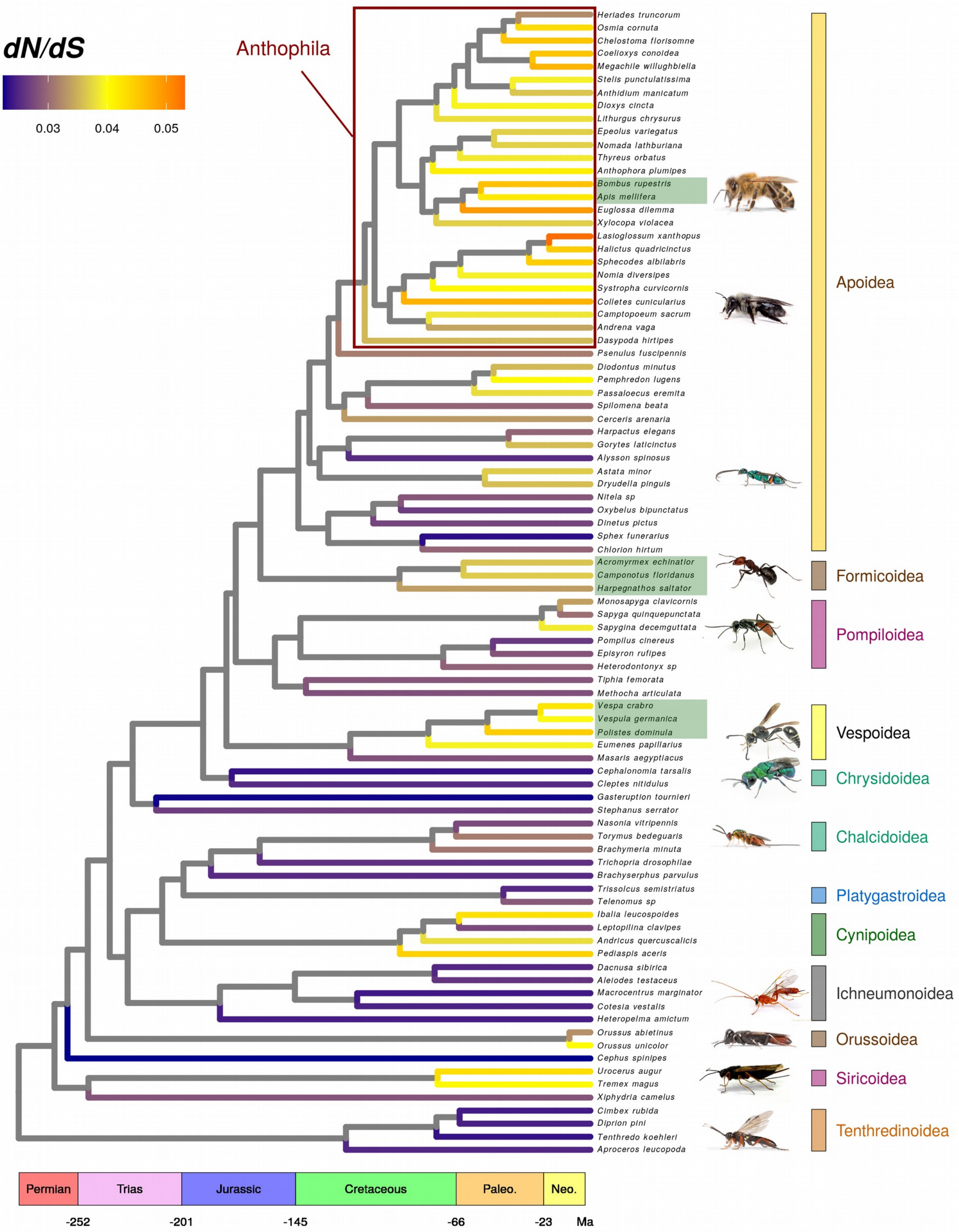


Figure S2: Genomic dN/dS ratios for 88 Hymenoptera species. dN/dS ratios estimated on terminal branches using 134 genes with data for each of the displayed species are represented on the chronogram inferred by Peters *et al.* (2017). Green rectangles around labels indicate eusocial taxa.

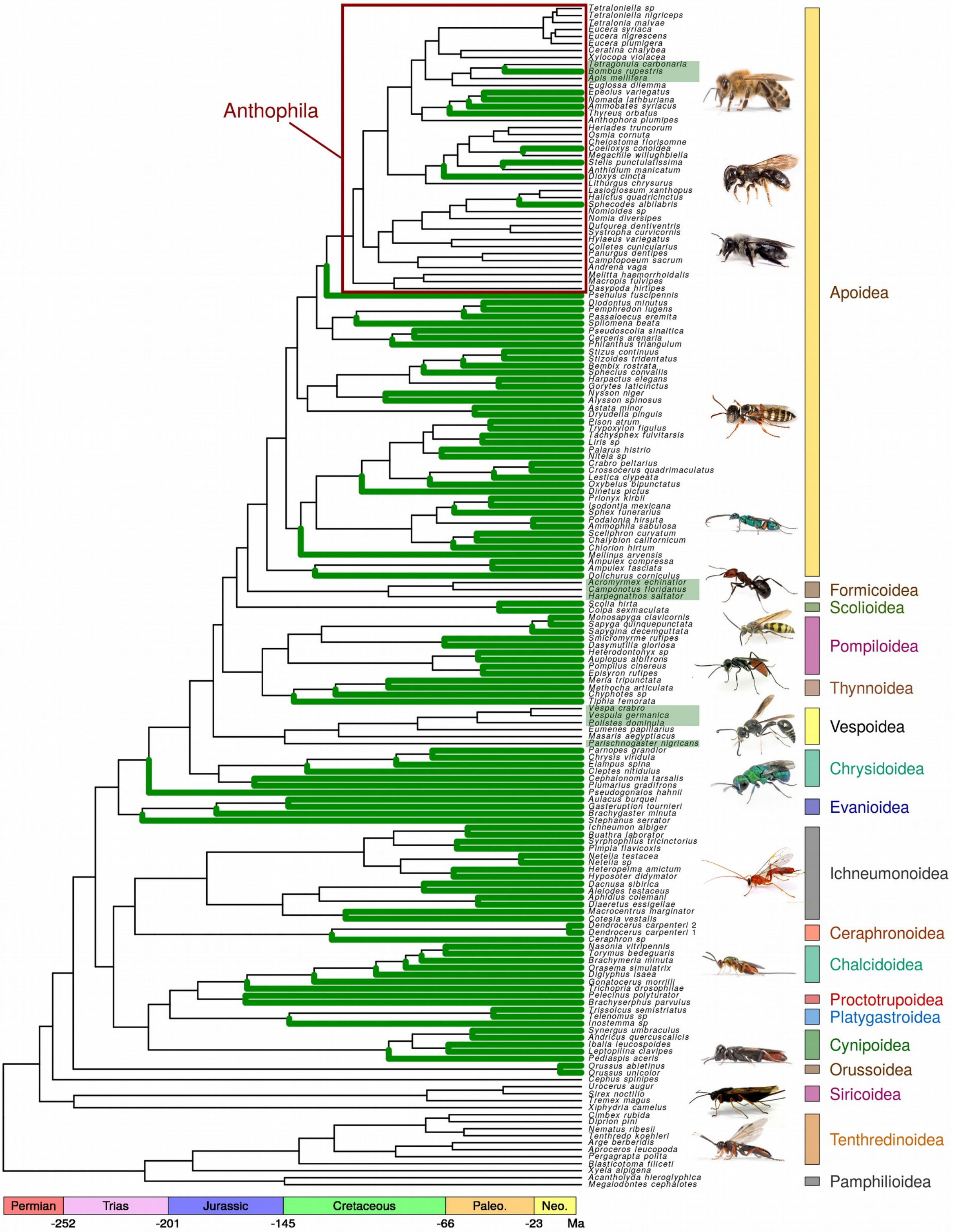


Figure S3: Parasitic species in the dataset. Parasitics species (parasitoids, parasites of plants and social parasites) are indicated by green terminal branches. Green rectangles around labels indicate eusocial taxa.

Variable		**Description**	**Number of genus with information**	**correlation**	**p-val**	**source**
continuous	Number of mature eggs at emergence	Maximal observed number during dissections of emerged females. This is a proxy for lifetime reproductive success for pro-ovigenic parasitoid species-	11	-0.048	0.889	Jervis & Ferns (2011) Mayhew (2016)
	Maximal number of mature eggs	Maximal observed number during dissections. This is a proxy for lifetime reproductive success for synovigenic parasitoid species-	10	0.146	0.688	Jervis & Ferns (2011) Mayhew (2016)
	Ovigenic index	Ratio of the last two variables	10	0.114	0.711	Jervis & Ferns (2011) Mayhew (2016)
	Oviposition rate	Average number of eggs laid by one female in one day	10	0.176	0.627	Jervis & Ferns (2011)
	Egg length	Propagule size	3	-0.5	0.667	Traynor & Mayhew (2005a)
	Clutch size	Average number of eggs in one clutch. This is 1 for non-gregarious species.	5	-0.707	0.182	Traynor & Mayhew (2005a)
	Fecundity	Average number of eggs laid by one female until death (continuous)	14	0.415	0.14	Jervis & Ferns (2011) Traynor & Mayhew (2005a)
	Developpement time : egg -> adult	Proxy of developmental speed	11	-0.2	0.55	Jervis & Ferns (2011)
	Time as an egg	Proxy of developmental speed	4	0.2	0.8	Traynor & Mayhew (2005a)
	Time as a pupa	Proxy of developmental speed	5	-0.3	0.624	Jervis & Ferns (2011) Traynor & Mayhew (2005a)
	Time as an adult	~~Proxy of developmental speed~~Proxy of longevity	14	0.226	0.436	Jervis & Ferns (2011) Traynor & Mayhew (2005a)
	Host body length	Loose proxy of res~~s~~ource availability~~abundance~~	8	0.119	0.779	Traynor & Mayhew (2005b)
	Number of potential host species	Proxy of the level of specialization	10	-0.234	0.515	Traynor & Mayhew (2005b)
discrete	Egg hydropisy	High or low nutritional res~~s~~ources in eggs	11	0.375	0.54	Jervis & Ferns (2011)
	Gregarious behavior	One or more larvae on one individual host	12	1.444	0.229	Jervis & Ferns (2011)
	Idiobiont/Koïnobiont	Parasitoid types	13	1.371	0.242	Jervis & Ferns (2011)
	Ecto/Endoparasitism	Parasitoid types	18	0.022	0.882	Jervis & Ferns (2011) Traynor & Mayhew (2005b)
	Hyperparasitism	~~Host is free or a parasite itself~~Host is another parasitoid or is free-living	12	0.185	0.667	Jervis & Ferns (2011)
	Host consumption behavior	Is the adult able to feed on the host too ?~~Is Adult able to feed on host too ?~~	12	0.26	0.61	Jervis & Ferns (2011)
	Host developpement stage at infection	~~On which developmental stage is the host when eggs are laid.~~The developmental stage of the host when the parasitoid lays her eggs.	10	3.836	0.28	Traynor & Mayhew (2005b)
	Host hiding behavior	How well is the host concealed when paraisitized ~~How well is the host dissimulated~~	7	2.4	0.121	Traynor & Mayhew (2005b)

Table S1: Life-history and specialisation descriptors for parasitoids. Tested variables and their description are displayed along with the value of the statistic obtained for each correlation test with corrected dN/dS ratios. Correlation tests are Spearman tests for continuous variables and Kruskal-

Wallis tests for discrete variables.

	All samples residual df = 127; R^2=0.113			**Non-Anthophila samples** residual df = 97; R^2=0.159		
covariables	R^2	F	p-value	R^2	F	p-value
branch length	0.0389	5.5767	**0.0197**	0.0813	9.3867	**0.0028**
adult size	0.0269	3.8587	0.0516	0.0275	3.1808	0.0776
Anthophila	0.0459	6.5762	**0.0115**			
Eusociality	0.0011	0.1598	0.6900	0.0501	5.7797	**0.0181**

Table S2: Linear modelling of uncorrected dN/dS ratios. Displayed results are obtained when simultaneously using all covariables inside a multiple linear model. Phylogenetic independent contrasts are used for all variables so as to account for phylogenetic autocorrelation.

	All samples residual df = 69; R^2=0.1494			**Non-Anthophila samples** residual df = 47; R^2=0.191		
covariables	R^2	F	p-value	R^2	F	p-value
branch length	0.0281	2.2860	0.1351	0.0677	3.9381	0.0530
adult size	0.0283	2.2963	0.1342	0.0524	3.0525	0.0871
Anthophila	0.0691	5.6131	**0.0206**			
Eusociality	0.0237	1.9294	0.1692	0.0714	4.1567	**0.0471**

Table S3: Linear modelling of uncorrected dN/dS ratios in the 88-species subsampled dataset. Corrected dN/dS are obtained using GC-conservative substitutions only. Displayed results are obtained when simultaneously using all covariables inside a multiple linear model. Phylogenetic independent contrasts where used for all variables so as to account for phylogenetic autocorrelation.

domain	GO ID	Term	p-val
biological process	GO:0043623	cellular protein complex assembly	0.00011
	GO:0016043	cellular component organization	0.00011
	GO:0043604	amide biosynthetic process	0.00012
molecular function	GO:0003723	RNA binding	0.00018
	GO:0008092	cytoskeletal protein binding	0.00030
	GO:0003735	structural constituent of ribosome	0.00083
	GO:0005488	binding	0.00109
	GO:0051020	GTPase binding	0.00113
	GO:0005085	guanyl-nucleotide exchange factor activity	0.00350
	GO:0017069	snRNA binding	0.00376
	GO:0019899	enzyme binding	0.00491
	GO:0005198	structural molecule activity	0.00500
	GO:0030246	carbohydrate binding	0.01089
	GO:0019904	protein domain specific binding	0.01286
	GO:0008536	Ran GTPase binding	0.02269
	GO:0003924	GTPase activity	0.02838
	GO:0017016	Ras GTPase binding	0.03243
	GO:0031267	small GTPase binding	0.03243

670 **Table S4: Go terms enriched in genes supporting an intensification of selection in eusocial Hymenoptera.** P-values are those of a Fisher hypergeometric test used for significance in the GO enrichment analysis, as implemented in the **R** package ***topGO*** *(Rahnenfuhrer and Alexa 2019)*

Variables	Description	All samples			Anthophila only		
		df	**correlation**	**p-val**	**df**	**correlation**	**p-val**
Mean Latitude	Mean latitude among occurences	132	-0.0317	0.7154	30	0.1912	0.2944
Maximal geographic distance	Maximal distance between occurence pairs	126	-0.1108	0.2131	28	-0.2650	0.1569
Circular area	Total area contained in 100km-radius circles around ocurrences	132	0.0819	0.3466	30	-0.2038	0.2631
Convex hull area	Total area contained in the smallest convex hull containing all occurences	114	-0.0138	0.8823	26	-0.2023	0.3019

Table S5: Correlation between corrected dN/dS values and proxies of the geographical range of species. The presented correlation coefficients are Pearson's product moment correlation coefficient. P-values are those of the associated correlation test.