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

Arthur Weyna¹, Jonathan Romiguier¹

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

Corresponding author: jonathan.romiguier@umontpellier.fr

ABSTRACT

Eusociality and parasitism are specific life-history strategies particularly common in a rare life history strategy that evolved repeatedly in Hymenoptera. At the population genetics level, inbreeding and low effective population size and reduced efficiency of purifying selection have been suggested as an evolutionary consequence of both social life and parasitism. In this study, we tested 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 Hymenoptera tree of life. We found no effect of parasitism, but show that relaxed selection is strongly associated with eusociality. This suggests that the division of reproductive labour decreases effective population size in ants, bees and wasps. Unexpectedly, the effect of eusociality is marginal compared to the effect of parasitism. Interestingly, we also report 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. Albeit strong, this 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. We suggest that a particularly high inbreeding rate in bees might increase the benefits of kin selection, which would explain why most independent origins of eusociality in the tree of life occurred in this taxon. 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 $N_e$ (Wright 1931). Deleterious mutations reach fixation with a higher probability in small populations, which undergo more drift, than in large populations in which purifying selection is more efficient. $N_e$ 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, Santiago, and Caballero et al. 2016). While different definitions of $N_e$ 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 is positively correlated with the genome-wide dN/dS (ratio of the non-synonymous substitution rate to the synonymous substitution rate) (Nikolaev et al. 2007; Romiguier, Lourenco, Popadin et al. 2014; Popadin2007; Romiguier et al. 20072014; Figuet et al. 2016; Botero-Castro et al. 2017; Rolland et al. 2020). However, most known examples are clustered within a few vertebrate taxa: mainly mammals, birds or reptiles. To date, only few examples of such patterns have been found in invertebrates, which cast doubt on the existence of a general relationship between life history strategies and the efficiency of natural selection in metazoan. Various reasons might explain the difficulty to demonstrate such relationships in invertebrates: There is relatively less genomic data available than in mammals or birds, and most taxa are expected to feature larger and more variable long-term Ne. Additionally, gathering life-history data in a large number of non-model invertebrates can be difficult as they have generally received less attention than vertebrates. Effective population size comparisons among invertebrates clades can also be particularly difficult, as the existence of reproductive systems such as haplo-diploidy affect Ne estimations (Wang et al. 2016).

Among invertebrates, Hymenoptera conveniently share the same haplo-diploid system, while displaying display—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 other Arthropods (parasitoids) (Mayhew 2016), which could shape their demography, as hosts population structure and size of the host can influence that of the parasite the parasite’s (Mazé-Guilmo et al. 2016). Second, the Hymenoptera order contains Hymenoptera contain 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 scattered in time and space. Finally, eusociality, which is a rare lifestyle in animals, is relatively common in the Hymenoptera order with at least 9 independent apparitions (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 of long-term Ne, and of the efficiency of natural selection, is often believed to be a general consequence of eusociality (Bromham and Leys 2005; Romiguier, Lourenco, et al. 2014; Settepani, Bechsgaard, and Bilde et al. 2016). Maintenance of high relatedness within low-Ne inbred group Without excluding the previous hypothesis, low Ne and inbreeding has also been raised as a prerequisite to the evolution of eusociality because it favors altruistic behaviors through kin-potential prerequisites, facilitating evolution towards eusociality by inflating within group relatedness and the benefit of kin selection (Hamilton 1972; Hsueseneneder et al. 1999; Hughes et al. 2008; Tabadkani et al. 2012). Ancestral population bottlenecks bottlenecks associated with such inbreeding episodes could thus be a typical feature of taxa in which eusociality frequently evolves.-

Some early studies have tried to detect Ne decreases in eusocial lineages by comparing their heterozygosity (Owen 1985; Berkelhamer 1983; Reeve, Reeve, and Pfennig Berkelhamer 1983; Owen 1985; Reeve et al. 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 to rely on a single locus sex determination system (Aspl en 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’s results are to be taken with care—however, as the number of compared taxa is low, and as none of the solitary reference taxa are Hymenoptera, while almost all compared
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.

We use Here we estimated the efficiency of purifying selection in a phylogenomic dataset of 3256 genes in 169 Hymenoptera species (Peters et al. 2017), including 10 eusocial species distributed among 4-5 independent origins of eusociality (Formicidae: 3 species; Polistinae/Vespinae wasps: 3 species; Stenogastrinae wasps: 1 species; Corbiculate bees: 3 species; Halictidae: 2 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. We estimated the efficiency of purifying selection by estimating mean genomic $dN/dS$. We compared these estimations between eusocial and solitary species, parasitic and free-living species, and correlated them to life-history descriptor variables. We also confirmed the links between increased $dN/dS$ and drops in the efficiency of purifying selection in eusocial species and solitary bees, via specialized analyses differentiating positive selection from relaxed purifying selection.

**RESULTS**

dN/dS distribution across the Hymenoptera phylogeny

We estimated $dN/dS$ in 3241 gene alignments of 169 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 2018). 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 (Duret and Galtier 2009; Lartillot 2013; 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 Hymenoptera tree on figure 1 (Distribution 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). As expected for conserved coding regions, the distribution of genomic $dN/dS$ ratios is close to 0 (overall average of $0.0947\pm0.003$), indicative of the large prevalence of purifying selection. We observed above average $dN/dS$ ratios in all 4 available eusocial clades: Formicidae (0.1068 ± 0.0093, 3 species), Polistinae/Vespinae wasps (0.1033 ± 0.0088, 3 species), Stenogastrinae (0.0951, 1 species only), the *Apis/Bombus/Tetragonula* clade (0.1086 ± 0.0352). 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, Colettaeidae, Andrenidae, and Melittidae) show high $dN/dS$ ratios (0.1190 ± 0.0302, 41 species) with no dependence on their social organization. Finally, only two purely solitary taxa displayed comparable $dN/dS$ ratios: Siricoidea (0.1025 ± 0.0251, 3 species) and Cynipoidea (0.1005 ± 0.0175, 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, which 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 \( T = 3.0567, \text{df} = 9.3549, p\text{-value} = 0.01305 \) over-representation of these eusocial species in alignments (mean number of alignments available per species: \( 2732.40 \pm 88.09 \)) as compared to solitary species \( (2276.7 \pm 90.74) \). To control for potential bias 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). Average corrected genomic dN/dS estimated using the 88-species subsampled dataset are displayed along the Hymenoptera tree on figure S2.

We estimated dN/dS in 3241 gene alignments of 169 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 2018). Average genomic dN/dS values are displayed along the Hymenoptera tree on Figure 1. We used this tree obtained by Peters et al. (2017) and its topology throughout all analyses. The largest and smallest mean ratios were inferred for Euceria nigrescens (0.0745) and Macrocetinus marginator (0.0318). As expected for conserved coding regions, the distribution of genomic dN/dS ratios is close to 0 (overall average of 0.0424), indicative of the large prevalence of purifying selection. We observed above average dN/dS ratios in all 5 available eusocial clades: Formicidae (0.0463 ± 0.0019, 2 species), Polistinae/Vespinae wasps (0.0492 ± 0.0023, 3 species), Stenogastrinae (0.0447, 1 species only), the Apis/Bombus/Tetragonula clade (0.0511 ± 0.0057), and Halictidae (0.0613 ± 0.0039, 2 species). These last two clades of bees do not clearly stand out however, as most bees in the dataset (Antophila, species characterized by pollen feeding of larvae: Apidae, Megachiliidae, Halictidae, Colettaeidae, Andrenidae, and Melitildae) show high dN/dS ratios (0.0522 ± 0.0073, 41 species) with no clear dependence on their social organization. Finally, only two purely solitary taxa displayed comparable dN/dS ratios: Siricoidea (0.0466 ± 0.0027, 3 species) and Cynipoidea (0.0447 ± 0.0059, 5 species). We further used simple linear modelling to try and relate variation in dN/dS ratios to life history traits. When needed, phylogenetic independent contrasts were used to account for phylogenetic relationships (Felsenstein 1985).

**Effect of body size and parasitism on relative protein evolution rates**

As in birds and mammals (Figuet et al. 2016; Botero Castro et al. 2017), we found a significant \((df=131, F=5.2647, p\text{-val} = 0.0234)\) and positive (slope = 1.844e-04) effect of body size on dN/dS ratio in Hymenoptera. This effect only explains a very small part of the observed variation \((R^2 = 0.03864)\) but remains significant when taking phylogeny into account \((df=130, F=4.3121, p\text{-val} = 0.03981, \text{slope} = 1.542e-04, R^2 = 0.0321)\). When testing for a difference in dN/dS ratios between parasitic (parasitoid or parasites) and free-living Hymenoptera, we found a significant effect \((df=167, F=46.327, p\text{-val} = 1.715e-10, R^2 = 0.2172)\), but which completely disappears when taking phylogeny into account \((df=166, F=1.2111, p\text{-val} = 0.2727, R^2 = 0.007243)\). We thus interpret this as being a confounding effect of sampling disequilibrium, as groups with elevated ratios completely lack parasite (with the exception of the cuckoo bumblebee *Bombus rupestris*). We further tried and tested 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 litterature (Jervis et al. 2003; Traynor and Mayhew 2005a; Traynor and Mayhew 2005b; 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.

**Eusociality is associated to accelerated relative protein evolution rates**

The social status of a terminal branch (eusocial or solitary species at tip) significantly explains dN/dS variations (df = 167; F = 15.403; p-val = 0.00012, R² = 0.0844) and this effect gets clearer when accounting for body size in the model (df = 130; F = 17.7127; p-val = 4.757e-05, R² = 0.1153). Depending on the exact definition of eusocial branches, this effect can be robust to phylogenetic corrections (df=129, F=9.2684, p-val = 0.002827, R²=0.0648) or not (df = 129, F=0.17, p-val=0.68083, R²=0.0012). More specifically, the effect is still significant only when considering terminal branches leading to Halictus and Lassioglossum genera as eusocial. We argue that these branches are likely to reflect eusocial evolutionary dynamics, because Halictus and Lassioglossum are known to originate from an eusocial ancestor (Cardinal and Danforth 2011) in spite of the secondary loss of eusociality in the specific species of our dataset. Table 1 presents results obtained when constructing full models accounting for all mentioned effects jointly, as well as results obtained when considering Halictus and Lassioglossum genera as solitary.

**Antophila bees displays highly accelerated relative protein evolution rates**

High dN/dS ratios in Antophila bees is by far the main pattern observed in our results. Treating appartenance to Antophila 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 = 0.5129). Despite Antophila being only one monophyletic group, this effect is still present when accounting for phylogeny and adult size (df = 139; F = 6.4645; p-val < 0.01219). As five eusocial species in the dataset belong to Antophila and were susceptible to drive the eusociality effect, we also validated the presence of an effect of eusociality when accounting for the Antophila effect (see table 1).

**Controlling for branch lengths, biased gene conversion or species sampling**

Short terminal branches are known to bias dN/dS estimations upward because they yield more inaccurate estimations of this parameter, which real value is often close to its zero boundary at a genomic scale. There is some association between dN/dS ratios and branch length in this study (df = 167; F = 4.0492; p-val = 0.0458, R²=0.02367) but it can not be responsible for the Antophila and eusociality effect as taking branch length into account does not nullifie any effects (see Table 1).

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

Unlike 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, we found a significant effect (df= 167, F= 46.327, p-val = 1.715e-10, R² = 0.2172), but which completely disappears when taking phylogeny into account (df= 166, F= 4.2111, p-val = 0.2727, R² = 0.007243). 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 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 literature (Traynor and Mayhew 2005a, [b] 2005; Jervis, 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 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 appartenence 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 = 0.5129). 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).

As eusocial Hymenoptera are known to have high recombination rates (Wilfert et al. 2007; Sirvio et al. 2011; Wallberg et al. 2015), which in turn are known to inflate dN/dS when associated to biased gene conversion in Vertebrates (Duret and Galtier 2009; Lartillot 2013; Galtier et al. 2018), we further validated these results using dN/dS estimated considering GC conservative substitutions only, and which 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 Hymenoptera tree on figure S1. All previous effects were recovered using these dN/dS ratios, with the exception of the adult size effect which is not significant when taking the phylogeny into account (table S2).

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. 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 +- 0.018) is not different than within genes without relaxed selection (2.463 +- 0.024) (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.

Four of the 12 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 \textit{Nasonia vitripennis} (Werren et al. 2010). This translated into a better power for gene prediction by Peters et al. (2017) in eusocial species, and thus into an over-representation of these eusocial species (mean number of alignments available per species: \(2732.40 \pm 88.09\)) as compared to solitary species (2276.7 \(\pm 90.74\)) that is significant (\(T = 3.0567, df = 9.3549, p \text{ value} = 0.01305\)). To control for a potential bias originating from varying precision in estimations, we validated all our results on a balanced subsampled dataset containing 134 alignments, each of them containing data for the same 88 species (most represented half of the species). Average corrected genomic dN/dS values are displayed along the Hymenoptera tree on figure S2. Both the effect of eusociality and of Antophila were recovered with this dataset (table S3).

**Eusocial taxa and Antophila bees experienced a relaxation of purifying selection**

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 a phylogeny and by comparing the distribution fitted on a focal group of branches (eusocial taxa and Antophila 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 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 S3. 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 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 doesn’t 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 \(\pm 0.018\)) is not different than within genes without relaxed selection (2.463 \(\pm 0.024\)) (\(F=2.11; p\text{val}=0.146\)).

These verifications are needed because 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 Antophila 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 eusociality**

We observed a significant higher accumulation of non-synonymous substitutions in eusocial genomes. 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, Glémín, and Webster 2015; Wilfert et al. 2007; Sirvö et al. 2011; Wilfert, Gadau, and Schmid-Hempel 2007; Jones Wallberg et al. 2019), these results could have been imputed to biased gene conversion, 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 not qualitatively changed by the use of dN/dS ratios accounting only for computed using only GC-conservative substitutions. 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).

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 labour 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.

Ecological and molecular predisposition to eusociality in bees

High genomic dN/dS ratios in all social and solitary bees unexpectedly 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 prerequisites 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. This dependence to large amounts of pollen to feed their larvae is 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.

Besides their implication regarding the evolution of eusociality, our results have important consequences for the conservation field. Indeed, pollination 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, Kuhlmann, and Potts 2009; Arbetman et al. 2017).

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 labour 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 eusociality complexification on selection efficiency. We observed a clear effect of adult body size on relative non synonymous substitution rates, an observation already made in vertebrates and usually interpreted as the effect of life history strategies on demographic parameters (Ne) and natural selection efficiency (dN/dS). In this study, we show that the emergence of a new level of organismic complexity (eusocial life history strategy) has a similar effect. One logical next step will be to investigate the existence of such associations within eusocial organisms, along a gradient of complexity in colonial organisation.

**Conclusion**

This study brings 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. Interestingly, the observation of a widespread reduction of selection efficiency in both eusocial and solitary also brings support to the hypothesis that some ecological characteristics associated with low Ne might facilitate evolution towards eusociality. Altogether, this study suggests that, as in Vertebrates, purifying selection efficiency in invertebrates is constrained by the ecology and life-history strategies of species.

However, the hypothesis of such an effect of the level of sociality on natural selection efficiency matches somewhat badly with the observation of a smaller increase of dN/dS ratios in highly eusocial and long lived ants than in less highly eusocial wasps. This hypothesis alone also fails in explaining why the dN/dS ratio shows no additional increase in social bees when compared to solitary ones.

**Acknowledgment**

Ecological and molecular predisposition to eusociality in bees
A high genomic dN/dS ratios in all social and solitary bees is a clear pattern in our results. Interestingly, Antophila (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 literature, 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). Interestingly, 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. While the idea is controversial, it has been suggested that inbreeding itself might directly favor the evolution of eusociality by strengthening the advantages of within-group altruism (Hamilton 1972; Hussedener et al. 1999, Tabadkani et al., 2012, but see Nowak et al. 2010). Whether it be direct or indirect, inbreeding is thus clearly believed to be associated with the evolution of eusociality, but few genomic evidences supporting such a link have been observed so far. By showing a striking increase in dN/dS ratio in all Antophila 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 high inbreeding rates might have preceded and/or favoured evolution towards eusociality.

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

The fact that the increase of dN/dS precedes eusociality in bees, the hymenopteran group where eusociality has appeared the most frequently, also leaves one to wonder whether the relaxation of selection itself could have had a role in eusociality emergence. This might be seen as an argument supporting the idea that analogously to some hypotheses for the transition from unicellularity to multicellularity (Lynch and Conery 2003; Lynch 2007), eusociality can be seen as a complex trait that can evolve only in taxa with low Ne, where selection is not strong enough to maintain simpler organisations.

**Data Availability**

Besides many independent transitions toward eusociality, Antophila are also characterized by their pollen-collecting behaviors. This dependence to large amounts of pollen to feed their larvae is believed to be a potential constraint on Ne, and this particularly in specialist species, an argument which has been made previously (Zayed and Packer 2007). Pollen is a resource which is scattered in space and time, and which can require a large energetic investment to come by and exploit (through progressive provisioning), thus constraining the very fecundity of female individuals. If pollen-collecting is linked to Ne variation in Hymenoptera, it might help explaining our results, where dN/dS are similarly high between social and solitary pollen-collecting species. Indeed, group living might enhance the productivity of pollen collecting and metabolizing, thus effectively compensating the decrease of Ne linked to eusociality in Antophila. A global survey of Hymenoptera fecundity would be necessary to test this hypothesis, even if such tests would be quite limited by the fact that Antophila is the only pollen-collecting taxa in Hymenoptera.

Besides their implication regarding the evolution of eusociality, our results have important consequences for the conservation field. Indeed, pollination 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. 2009; Arbetman et al. 2017).
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.

**Conclusion**

This study brings new and genomic evidence supporting the existence of an association between eusociality and reduced selection efficiency, through reduced effective population size. Interestingly, the observation of a widespread reduction of selection efficiency in eusocial and solitary bees suggests that the picture might not be as clear-cut as thought previously. It is indeed possible that this reduction in effective population size is caused in part by ecological innovations that are not strictly eusocial characteristics, or that these reductions themselves have favored the complexification of social organisations. Besides the new questions raised by this study, this study brings the first genomic evidence suggesting that, as Vertebrates, invertebrates are constrained in their molecular evolutionary dynamics by their ecology and reproductive strategy.

**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*) and Halictine bees (*Lasioglossum xanthopus* and *Halictus*).

**Data cleaning**

Each amino-acid alignment was first checked for potential false homology using HmmCleaner (Di Franco Philippe et al. 2019; Philippe 2017; Di Franco et al. 2017, 2019) 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 alignment, 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 occurrence, 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.

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.

**RELAX analyses**

We used the RELAX procedure (Wertheim et al. 2015) from the HyPhy program suite (Pond, Frost, and Muse et al. 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 (\(\omega_f = \omega_b^k\)). 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 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 alignments didn’t allow the procedure with AnthophilaAnthophila branches as background branches.

**ACKNOWLEDGEMENTS**

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

**REFERENCES**


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

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.
Figure S1: **Uncorrected** Corrected genomic dN/dS ratios for 169 Hymenoptera species. dN/dS ratios estimated on terminal branches using 3241 genes and GC conservative substitutions are
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 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.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Number of genus with information</th>
<th>correlati on</th>
<th>p-val</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of mature eggs at emergence</td>
<td>Maximal observed number during dissections of emerged females</td>
<td>11</td>
<td>-0.048</td>
<td>0.889</td>
<td>Jervis &amp; Ferns (2011) Mayhew (2016)</td>
</tr>
<tr>
<td>Maximal number of mature eggs</td>
<td>Maximal observed number during dissections</td>
<td>10</td>
<td>0.146</td>
<td>0.688</td>
<td>Jervis &amp; Ferns (2011) Mayhew (2016)</td>
</tr>
<tr>
<td>Ovigenic index</td>
<td>Ratio of the last two variables</td>
<td>10</td>
<td>0.114</td>
<td>0.711</td>
<td>Jervis &amp; Ferns (2011) Mayhew (2016)</td>
</tr>
<tr>
<td>Oviposition rate</td>
<td>Average number of eggs laid by one female in one day</td>
<td>10</td>
<td>0.176</td>
<td>0.627</td>
<td>Jervis &amp; Ferns (2011)</td>
</tr>
<tr>
<td>Egg length</td>
<td>Propagule size</td>
<td>3</td>
<td>-0.5</td>
<td>0.667</td>
<td>Traynor &amp; Mayhew (2005a)</td>
</tr>
<tr>
<td>Clutch size</td>
<td>Average number of eggs in one clutch</td>
<td>5</td>
<td>-0.707</td>
<td>0.182</td>
<td>Traynor &amp; Mayhew (2005a)</td>
</tr>
<tr>
<td>Fecundity</td>
<td>Average number of eggs laid by one female until death (continuous)</td>
<td>14</td>
<td>0.415</td>
<td>0.14</td>
<td>Jervis &amp; Ferns (2011) Traynor &amp; Mayhew (2005a)</td>
</tr>
<tr>
<td>Development time : egg -&gt; adult</td>
<td>Proxy of developmental speed</td>
<td>11</td>
<td>-0.2</td>
<td>0.55</td>
<td>Jervis &amp; Ferns (2011)</td>
</tr>
<tr>
<td>Time as an egg</td>
<td>Proxy of developmental speed</td>
<td>4</td>
<td>0.2</td>
<td>0.8</td>
<td>Traynor &amp; Mayhew (2005a)</td>
</tr>
<tr>
<td>Time as a pupa</td>
<td>Proxy of developmental speed</td>
<td>5</td>
<td>-0.3</td>
<td>0.624</td>
<td>Jervis &amp; Ferns (2011) Traynor &amp; Mayhew (2005a)</td>
</tr>
<tr>
<td>Time as an adult</td>
<td>Proxy of developmental speed</td>
<td>14</td>
<td>0.226</td>
<td>0.436</td>
<td>Jervis &amp; Ferns (2011) Traynor &amp; Mayhew (2005a)</td>
</tr>
<tr>
<td>Host body length</td>
<td>Loose proxy of resource abundance</td>
<td>8</td>
<td>0.119</td>
<td>0.779</td>
<td>Traynor &amp; Mayhew (2005b)</td>
</tr>
<tr>
<td>Number of potential host species</td>
<td>Proxy of the level of specialization</td>
<td>10</td>
<td>-0.234</td>
<td>0.515</td>
<td>Traynor &amp; Mayhew (2005b)</td>
</tr>
<tr>
<td>Egg hydropisy</td>
<td>High or low nutritional resources in eggs</td>
<td>11</td>
<td>0.375</td>
<td>0.54</td>
<td>Jervis &amp; Ferns (2011)</td>
</tr>
<tr>
<td>Gregarious behavior</td>
<td>One or more larvae on one individual host</td>
<td>12</td>
<td>1.444</td>
<td>0.229</td>
<td>Jervis &amp; Ferns (2011)</td>
</tr>
<tr>
<td>Idiobiont/Koi nobiont</td>
<td>Parasitoid types</td>
<td>13</td>
<td>1.371</td>
<td>0.242</td>
<td>Jervis &amp; Ferns (2011)</td>
</tr>
<tr>
<td>Ecto/Endoparasitism</td>
<td>Parasitoid types</td>
<td>18</td>
<td>0.022</td>
<td>0.882</td>
<td>Jervis &amp; Ferns (2011)</td>
</tr>
<tr>
<td>Hyperparasitism</td>
<td>Host is free or a parasite itself</td>
<td>12</td>
<td>0.185</td>
<td>0.667</td>
<td>Jervis &amp; Ferns (2011)</td>
</tr>
<tr>
<td>Host consumption behavior</td>
<td>Is Adult able to feed on host too ?</td>
<td>12</td>
<td>0.26</td>
<td>0.61</td>
<td>Jervis &amp; Ferns (2011)</td>
</tr>
<tr>
<td>Host development stage at infection</td>
<td>On which developmental stage is the host when eggs are laid.</td>
<td>10</td>
<td>3.836</td>
<td>0.28</td>
<td>Traynor &amp; Mayhew (2005b)</td>
</tr>
<tr>
<td>Host hiding behavior</td>
<td>How well is the host dissimulated</td>
<td>7</td>
<td>2.4</td>
<td>0.121</td>
<td>Traynor &amp; Mayhew (2005b)</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th>covariables</th>
<th>All samples residual df = 127; $R^2$=0.113</th>
<th>Non-Anthophila samples residual df = 97; $R^2$=0.159</th>
</tr>
</thead>
<tbody>
<tr>
<td>branch length</td>
<td>0.0389  5.5767  <strong>0.0197</strong></td>
<td>0.0813  9.3867  <strong>0.0028</strong></td>
</tr>
<tr>
<td>adult size</td>
<td>0.0269  3.8587  0.0516</td>
<td>0.0275  3.1808  0.0776</td>
</tr>
<tr>
<td>Anthophila</td>
<td>0.0459  6.5762  <strong>0.0115</strong></td>
<td></td>
</tr>
<tr>
<td>Eusociality</td>
<td>0.0011  0.1598  0.6900</td>
<td>0.0501  5.7797  <strong>0.0181</strong></td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>covariables</th>
<th>All samples residual df = 69; $R^2$=0.1494</th>
<th>Non-Anthophila samples residual df = 47; $R^2$=0.191</th>
</tr>
</thead>
<tbody>
<tr>
<td>branch length</td>
<td>0.0281  2.2860  0.1351</td>
<td>0.0677  3.9381  0.0530</td>
</tr>
<tr>
<td>adult size</td>
<td>0.0283  2.2963  0.1342</td>
<td>0.0524  3.0525  0.0871</td>
</tr>
<tr>
<td>Anthophila</td>
<td>0.0691  5.6131  <strong>0.0206</strong></td>
<td></td>
</tr>
<tr>
<td>Eusociality</td>
<td>0.0237  1.9294  0.1692</td>
<td>0.0714  4.1567  <strong>0.0471</strong></td>
</tr>
</tbody>
</table>

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.
Table S2: Results of simple linear modelling of corrected dN/dS ratios. Displayed results are obtained when using dN/dS ratios obtained using only GC conservative substitutions, and simultaneously accounting for all effects inside a multiple linear model.

Table S3: Results of simple linear modelling of dN/dS ratios. Displayed results are obtained when using dN/dS ratios obtained using the subsampled dataset (88 species and 134 genes), and simultaneously accounting for all effects inside a multiple linear model.

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

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)
| Mean Latitude | Mean latitude among occurrences | 132 | -0.0317 | 0.7154 | 30 | 0.1912 | 0.2944 |
| Maximal geographic distance | Maximal distance between occurrence pairs | 126 | -0.1108 | 0.2131 | 28 | -0.2650 | 0.1569 |
| Circular area | Total area contained in 100km-radius circles around occurrences | 132 | 0.0819 | 0.3466 | 30 | -0.2038 | 0.2631 |
| Convex hull area | Total area contained in the smallest convex hull containing all occurrences | 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.