Response to the recommender

We would like to first thank you for accepting to evaluate our work and for the quality of the reviews received. Indications were clear and constructive. We tried to address all the concerns of the reviewers about the structure and content of the article. Please, find below our answers, and find at the following link the new version of the manuscript (lines indicated in the answers refers to the lignes of the last version of the article) : https://www.biorxiv.org/content/10.1101/2020.04.14.038893v2 . We also provided a pdf with the tracked modifications.

Response to reviewer 1.

We first want to thank the reviewer for this constructive review of our paper. We tried to take into account all comments and found our manuscript to be greatly improved.

>In the second paragraph of the results section you mention that only B. rupestris represents a parasitic species amongst the species with elevated dN/dS. This is not completely true as Sphecodes is a parasitic species of Halictidae and Colletidae.

This is entirely true and we apologize for missing this. We modified our text to mention this second case of parasitism (l.152).

>In the section Anthophila bees and high rates of protein evolution you claim to have 5 eusocial species represented while these are effectively only four species and one socially parasitic species (B. rupestris). This should be corrected as socially parasitic bumblebees show a range of different levels of gene evolution (e.g. Erler et al. 2014 Infect Genet Evol; Helbing & Lattorff 2016 Infect Genet Evol; Fouks & Lattorff 2016 Front Ecol Evol).

We consider *B. rupestris* as an eusocial species while it is a socially parasitic species which, as you showed in your work, is most probably under different new selective pressures due to this peculiar ecology. The postulate we made here is that given the resolution of the phylogeny we use, the terminal branch leading to *B. rupestris* encompasses the evolution of the whole genus, which is ancestrally eusocial. The evolution of social parasitism being quite recent, relative to the age of the genus, this branch is thus supposed to carry information about evolution under eusociality in terms of long-term *Ne*.

>For differences in recombination rates you might also cite Jones et al 2019 Mol Biol Evol.

We now cite this work when discussing the potential effect of recombination (l.106,208).

>The argument of inbreeding as prerequisite for creating high relatedness and thus favouring the evolution of sociality needs a valid explanation in the light of the genetic load at the sex locus. Single locus sex determination (SISD) has been show to occur in a range of species with the exception of the few parasitoid species that might have a whole genome sex determination system. As also a range of very basal members of the Hymenoptera have the SISD, it might be the ancestral state.

It is true that haplo-diploid species showing such a sex determination system should pay an extra-cost to inbreeding. The loss of allelic diversity at the sex locus would lead to an increased proportion of usually non-functional diploid males, a fact that has been used to partly explain the evolution of strong inbreeding avoidance mechanisms in Hymenoptera. However, this extra-cost is not specific to this hypothesis but also valid for the evolution of eusociality and reduced effective population size in general. We agree that this is an important part of the question, that we chose not to mention in the initial version of the paper. We follow your advice and mention this rationale in the new version of the Introduction (l.74-77) and Discussion (l. 243).

>My major point here is, why you have not chosen more eusocial taxa to be sampled? Quite a range of sequenced genomes are available which could have been utilized (okay, the gene annotation is a bit more complicated, but for some of them also transcriptomes are available).

We chose to strictly stick to the dataset of Peters et al. (2017) and not to increment it with additional data for various reasons. First, manually adding data from eusocial species may bias the dN/dS analyses. Indeed, the transcriptomic data produced by Peters et al. (2017) was treated in a homogeneous, specific, and hardly reproducible way, and we fear that adding differently treated data (because of different sequencing technologies, with different bioinformatic treatments or software versions) will introduce many biases (data heterogeneity in terms of error rates, gene length depending on sequencing technology, gene family/orthology definition…). As analyses of dN/dS can be very sensitive to data heterogeneity, we believe that relying on such an homogeneous dataset is one of the main strengths of our work, leading to more conservative results. By adding new data from other eusocial species, we expect higher chances of alignment or orthology errors between these species and the main dataset, which will invariably be translated into higher dN/dS, leading to an overestimated effect of eusociality. Our last reason not to add additional genomes to the analysis stems in the balance between the methodological biases it will bring (described above) and the real gain in terms of statistical power given the currently available data. We estimated that the gain in statistical power would be shallow at best, as all available genomic data on fully eusocial hymenoptera are restricted to three eusocial groups that are already represented in our dataset: Formicidae, Corbiculate bees and Polistinae/Vespinae wasps (we added a list of available genomes at the end of this document). As phylogenetic control is used in our statistical analysis, adding new taxa to these groups would probably not yield a great gain in statistical power.

>Certainly there will be variation within the eusocial species as well and they might differ in important aspects as annual vs perennial colonies, colony size, at what age sexuals are produced (generation-time effect, some ants produce sexuals after years, bumblebees produce them after a few months). Thus, increased sampling within the eusocial lineages might cover more of these aspects.

Adding new eusocial species would surely make it possible to cover more aspects. It would allow to test for other predictions of population genetics regarding the effect of some eusocial LHT, such as a colony longevity, size or generation-time, on molecular evolution. We believe this to be an important and ambitious prospect, that we are also contemplating ourselves, and that we tried to introduce in the subsection "Molecular consequences of eusociality" (l.222-224). There, we briefly described ants as an ideal group to address these questions, as they are a most ecologically diverse group with many genomes available. We had not considered bumblebees at first but now that you mention it they would surely represent important data points is such a study. However, it represents a secondary and different question, which we believe should be addressed in another study and would require a tailor-made dataset.

Additionally, we would like to stress that in our opinion, the most novel and convincing aspect of this work is more the extremely high dN/dS ratios in both social and non-social pollinating bees than the effect of eusociality. We therefore acknowledge that it was not clear at all in the previous version of the manuscript, and hope that the changes in the title, abstract, results and discussion makes it now clearer.

> Furthermore, I wondered that there is no link to the distribution of species. For Anthophila you use the argument that pollen collection might constraint the Ne because of the limited
availability of this resource. Agreed, but this might only be true for species of temperate regions, while in the tropics/subtropics pollen might be a less limited resource. Furthermore, the number of eusocial species (relative number) is higher in tropical/subtropical habitats compared to temperate regions (e.g. stingless bees only occur in such regions, the hotspot of honeybee evolution and diversification is in South-East Asia). There is also an effect for the occurrence of nest parasites of social bees, wasps, and ants, which are more frequent in temperate regions (Wcislo 1987 Biol Rev) probably to factors like seasonality and hence predictability of developmental stage of a colony to attack. Hence external environments might affect certain characteristics that in turn might have an influence of sociality but also pollen collection. I suggest to infer the distribution data for at least the species of interest or a least discuss this issue.

We agree that this is an important point and have tried to link our results with some proxies of species distribution. For each present species, we have retrieved GBIF occurrences and computed mean latitude as well as different versions of distribution areas. However, as none of these simple proxies correlates with dN/dS, as GBIF is notoriously prone to error, and because the genetic data available here is insufficient in terms of represented ecologies and geographical distributions, we have chosen not to mention this analysis in the first version of our paper. As your suggestions make a lot of sense and made us understand that it could be important to at least report our limited results, we have modified the manuscript to include them (l.159-162, 260-264, 360-370, new table S5).

Response to reviewer 2

We first want to thank the reviewer for this constructive review of our article. We tried to take into account these comments and found our manuscript to be greatly improved.

> I am not sure it is fair to say most invertebrates have larger and less variable long-term Ne. I would wager that eusocial insects have the most genetic data, and they are also likely to have the lowest Ne because of their reproductive division of labor.

We agree that this sentence was confusing as written and removed this part from the manuscript (l.45)

> Inbreeding has not been invoked to explain the evolution of eusociality. The maintenance of high relatedness within groups has been raised as a prerequisite, but not inbreeding.

It is true that ultimately, it is not inbreeding itself but the maintenance of high intra-group relatedness which favors the evolution of eusociality by means of kin selection. However, one of the direct effects of inbreeding is to raise said intra-group relatedness. This has led several authors to recognize inbreeding as one of the indirect mechanisms facilitating the evolution of eusociality (see Hughes et al., 2008; Bourke, 2011; Tabadkani et al. 2012). We agree that we presented this idea in a confusing way and modified our text in the abstract (l. 19), introduction (l. 63-67) and discussion (l.238-244) to clarify the links between Ne, intra-group relatedness, and inbreeding .

>With the exception of some social spider species, most insects (and vertebrates and other taxa), AVOID inbreeding because the costs of doing so generally outweigh the benefits.

We agree and this should be particularly true in Hymenoptera, as many species in this group are expected to rely on single locus sex determination, which does not play well with inbreeding (we added some mentions to this idea in the manuscript (l.74-77,243) following the demand of another reviewer). However, the existence of inbreeding avoidance mechanisms does not mean that inbreeding is always avoided, and there is a wide spectrum of inbreeding/outbreeding strategies even in insects. See for example the study Schrempf et al. (2005), suggesting that a large part of matings in the flightless monogynous ant *Cardiocondyla batesii* occur between brothers and sisters. Also, following the fig. 1 of Tabadkani et al. (2012) that you can see below, while inbreeding is expected to first favours the evolution of eusociality, we completely expect that it will be generally avoided after this first step to counter-balance general issues related to low-Ne (which can also be counter-balanced by inbreeding preferences, as in social spiders).

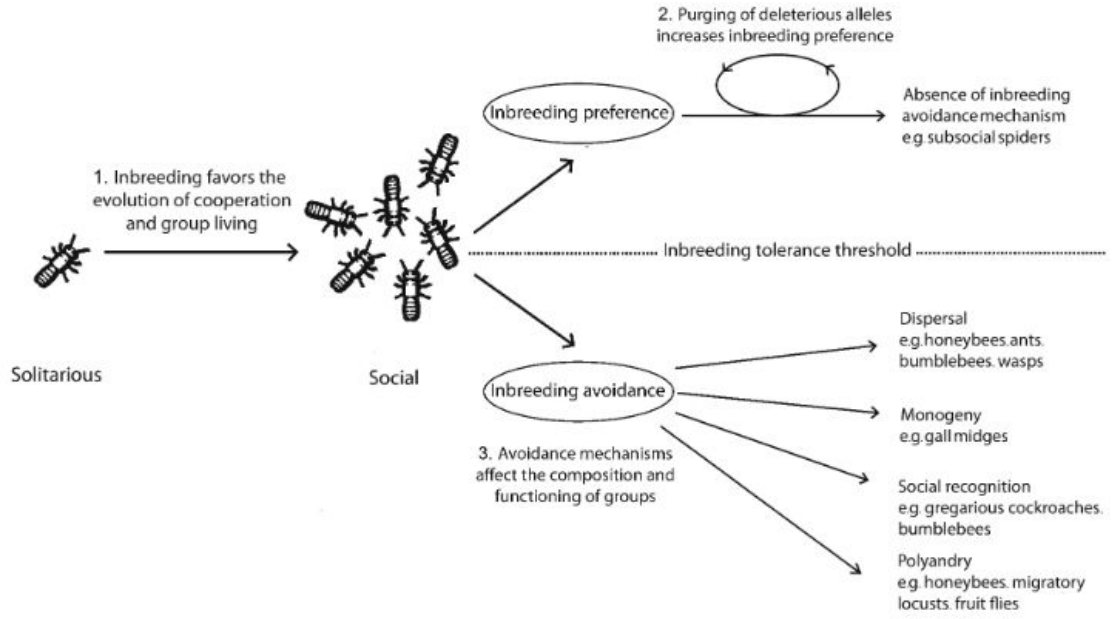

Fig. 1 Theoretical diagram illustrating how inbreeding could affect the evolution of social behaviors. *1* In the early stages of socialization, inbreeding drives the evolution of cooperation and group-living via the action of kin selection. However, as animals go from solitarious to social, their risks of incurring high costs of inbreeding depression increase as closely related individuals are more likely to encounter and mate. We suggest that inbreeding tolerance by individuals, as defined by the balance between the kin-selected benefits of inbreeding and the costs of inbreeding depression, determines two pathways for sociality. *2* If the costs of inbreeding depression are lower than the benefits of kin selection, a long history of inbreeding and the purging of deleterious alleles should further promote the evolution of sociality by kin selection. This scenario has been proposed to explain the transition from subsociality to sociality in spiders. *3* Alternately, if the costs of inbreeding depression are higher than the benefits of kin selection, individuals should evolve mechanisms to avoid (or reduce) inbreeding as observed in various arthropods (Tables 1–3). These mechanisms have a profound impact on the composition and the functioning of social groups and reduce the evolutionary potential of sociality via kin selection, thus shaping the social biology of species. In arthropods, inbreeding avoidance mechanisms include dispersal, kin recognition, polyandry, and monogeny

Fig 1 of Tabadkani et al. (2012)

>Hymenoptera are also “special” because they are haplodiploid, which could result in lower Ne. You mention this later in introduction, but it is a bit misleading to not mention this when you introduce Hymenoptera earlier.

As requested, we now mention this earlier in the text (line 45-49).

>I think the end of the Introduction should briefly describe the approach taken to compare these species and look at purifying selection. You should also make it clear that only 12 of the 169 species examined are eusocial.

We added such a description at the end of the introduction and, as suggested, tried to be more specific about the low number of eusocial species in the dataset.

>I am actually surprised by how low this number. Certainly, there are many more than 12 complete genomes published for ants, bees, wasps. Why have the authors only chosen 12 of these is unclear. It would certainly improve you paper to use genetic data from more eusocial species, which is definitely available as of 2020.

Reviewer 1 has made the same discerning point. We agree that ultimately, more data would have improved our paper, but it would also have created additional biases and constraints. See the detailed answer in response to reviewer 1.

>You can't claim a branch is eusocial because sample you have not sampled are eusocial. You really should work with the data (and species) you have here.

We assume this refers to the use we made of the available genetic data for the solitary Halictine bees *Lasioglossum xanthopus* and *Halictus quadricinctus*. According to some authors (Cardinal and Danforth 2011), these species originate from an primitively eusocial ancestor but lost this trait secondarily. We thus postulated that terminal branches leading to these species could carry some information about evolution under eusociality, and chose to apply our analyses both with and without these species as eusocial ones. We then proceeded to show that considering these two species as eusocial would reinforce the estimated effect of eusociality. We see now how confusing this postulate can be, and we agree that in a sense, this is equivalent to working with data we do not have. We thus choose to follow the advice of the reviewer and treat these species as strictly solitary. We modified all the Results section accordingly. As presented in the original version of the manuscript, the effect of eusociality becomes non-significant when considering the two halictine bees (mentioned before) as solitary and including all Anthophila in the analyses. This is because eusocial corbiculate bees do not show an increase in dN/dS relative to other Anthophila. The effect of eusociality becomes significant only after removing all Anthophila bees from the dataset. Also, as stated in one answer to Reviewer 1, we want to stress the point that we do not believe that the main and most interesting result of this article is the effect of eusociality. The most novel and convincing aspect of this work is the extremely high dN/dS ratios in both social and non-social pollinating bees We acknowledge that it was not clear at all in the previous version of the manuscript, and hope that the changes in the title, abstract, results and discussion makes it now clearer.

>In the analyses in the "Anthophila bees displays highly accelerated relative protein evolution rates", are you basically saying that this group is driving all of your results? I have trouble following this section, and not sure it needs to be its own sub-section.

As mentionned is the previous paragraph, the large increase in dN/dS in Anthophila is masking the smaller effect of eusociality, rather than driving it. To make it clearer that the effect of eusociality is not driven by Anthophila bees, we added a model testing for the effect of eusociality within only non-Anthophila species. This model shows that if the Anthophila effect is put aside, the effect of eusociality is clearly visible and significant. This is confirmed also by the more sensitive, gene-by-gene Hyphy RELAX analysis, which includes all samples and still shows an important relaxation of selection for many genes on eusocial branches (halicitine bees excluded). We modified the abstract (l.15-17) and results (l.64-74) to make it clearer that the Anthophila effect is a separate effect, which does not drive the results regarding eusociality. We actually believe that this surprising effect represents a result by itself, at least as interesting as the effect of eusociality and for sure more novel. We now emphasize this point in the text and make it clearer that this is not just a confounding effect.

>Rather than having an entire section "Controlling for branch lengths, biased gene conversion or species sampling", why not just include branch length in all of your prior models? Similarly, why not use corrected genomic dN/dS values from the outset, rather than saying doing so gives similar results? In other words, rather than replicating analyses

multiple times throughout this paper, I think it would be much improved if you simplify this by explaining all of this in the methods and then presenting fewer models that control for various factors and use standardized values.

We agree that it is much simpler, and just as effective, to describe all corrections and present only the results from the last full model using all corrections together. We applied this advice and find our manuscript to be much clearer this way. We thank the reviewer for this constructive advice.

> In the analyses in “Effect of body size and parasitism on relative protein evolution rates” it is unclear how you control for phylogenetic non-independence. In the methods you only say that you use PICs, but no information on how you actually do this. Currently, it seems like these are straight linear models, but these should all be done in a phylogenetic framework. I would only present the results from the phylogenetic analyses.

We used PIC, which are a transformation of the initial data (both response variable and covariables), yielding a dataset that is free of phylogenetic autocorrelation, ready to be used within a straight linear model. The use of straight linear models on this transformed data thus does not mean that analyses are not done in a phylogenetic framework

.

>I also think you might have the order of events backwards when you discuss the relationship between Ne and eusociality. Eusocial species have low Ne because of a reproductive division of labor and high reproductive skew within groups. Ne is therefore a consequence of the social system rather than a cause. When you say things like “…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.”, I think your logic is flawed. I don’t know of a single person studying social evolution in insects or any other organism that would agree with this statement. I think your directionality is backwards and you need to think that high skew results in low Ne. The selective forces that driving group influence skew, which in turn influences Ne.

We understand these rightful concerns about this really nonclassical hypothesis, taking the traditional causality between eusociality and small Ne backwards. We make a questioning analogy between our results and some theoretical work about the evolution of multicellularity. Studying the evolution of multicellularity, Lynch (2007) has made the strong postulate that a reduction in Ne, and thus in the relative strength of selection, might be the main effect allowing organisms to develop multicellularity, as strong selection would not allow such complexity to evolve. We found this hypothesis to coincide interestingly with our data, where we observe a taxon-wide drop in Ne in all Anthophila, the very group which counts the more convergence towards eusociality. Of course this analogy is highly debatable, and by no means represented our main proposal. Moreover, testing this hypothesis would require a much more refined analysis and dataset. We removed this part of our manuscript because it was indeed pure conjecture, and also removed these prospects from the new versions of the introduction and conclusion. We thought that this idea deserved to be at least mentioned, as it would not be the first time that analogies between multicellularity and eusociality yield interesting new ideas, but understand the reviewer’s concerns.

annotated genomes of fully eusocial species in orthoDB

Acromyrmex echinatior	Formicidae
Atta colombica	Formicidae
Camponotus floridanus	Formicidae
Cardiocondyla obscurior	Formicidae
Cyphomyrmex costatus	Formicidae
Dinoponera quadriceps	Formicidae
Harpegnathos saltator	Formicidae
Lasius niger	Formicidae
Linepithema humile	Formicidae
Monomorium pharaonis	Formicidae
Ooceraea biroi	Formicidae
Pogonomyrmex barbatus	Formicidae
Pseudomyrmex gracilis	Formicidae
Solenopsis invicta	Formicidae
Trachymyrmex septentrionalis	Formicidae
Vollenhovia emeryi	Formicidae
Wasmannia auropunctata	Formicidae
Bombus impatiens	Corbiculate bees
Melipona quadrifasciata	Corbiculate bees
Polistes canadiensis	Polistine/Vespine wasps

genomes of fully eusocial species in genbank genome database

Apis cerana	Corbiculate bees
Bombus terrestris	Corbiculate bees
Bombus vancouverensis	Corbiculate bees
Lepidotrigona ventralis	Corbiculate bees
Bombus vosnesenskii	Corbiculate bees
Bombus bifarius	Corbiculate bees
Tetramorium parvispinum	Formicidae
Tetramorium simillimum	Formicidae
Tetramorium bicarinatum	Formicidae
Tetramorium immigrans	Formicidae
Tetragonula mellipes	Corbiculate bees
Tetragonula hockingsi	Corbiculate bees
Tetragonula davenporti	Corbiculate bees
Tetragonula clypearis	Corbiculate bees
Tetragonula carbonaria	Corbiculate bees
Odontomachus brunneus	Formicidae
Polistes dorsalis	Polistine/Vespine wasps
Polistes fuscatus	Polistine/Vespine wasps
Polistes metricus	Polistine/Vespine wasps
Temnothorax longispinosus	Formicidae
Atta texana	Formicidae
Cataglyphis niger	Formicidae
Solenopsis fugax	Formicidae
Aphaenogaster picea	Formicidae
Aphaenogaster floridana	Formicidae
Aphaenogaster miamiana	Formicidae
Aphaenogaster rudis	Formicidae
Aphaenogaster fulva	Formicidae
Aphaenogaster ashmeadi	Formicidae
Temnothorax curvispinosus	Formicidae
Ooceraea biroi	Formicidae
Pseudomyrmex gracilis	Formicidae
Trachymyrmex zeteki	Formicidae
Trachymyrmex septentrionalis	Formicidae
Trachymyrmex cornetzi	Formicidae
Cyphomyrmex costatus	Formicidae
Atta colombica	Formicidae
Dinoponera quadriceps	Formicidae
Monomorium pharaonis	Formicidae
Wasmannia auropunctata	Formicidae
Formica selysi	Formicidae
Vollenhovia emeryi	Formicidae
Formica exsecta	Formicidae
Polistes canadiensis	Polistine/Vespine wasps
Cataglyphis hispanica	Formicidae
Apis dorsata	Corbiculate bees
Lasius niger	Formicidae
Melipona quadrifasciata	Corbiculate bees
Frieseomelitta varia	Corbiculate bees
Acromyrmex echinatior	Formicidae
Bombus impatiens	Corbiculate bees
Camponotus floridanus	Formicidae
Harpegnathos saltator	Formicidae
Solenopsis invicta	Formicidae
Atta cephalotes	Formicidae
Apis florea	Corbiculate bees
Linepithema humile	Formicidae
Pogonomyrmex barbatus	Formicidae