EDITOR's COMMENTS

(1) All three reviewers ask for more clarity of Method and Figure description. You will find specific comments and suggestions in the respective review reports.

Response: Thanks for all the pertinent suggestions; see below detailed responses to reviewers' comments. In short we :
- added new analyses of life history traits
- performed a new analysis of the distribution of substitutions across coding seuence length – but did not include it in the manuscript
- included one additional Table (model parametrisation) and one additional Supp Figure (Akaike weights)
- clarified various aspects of our approach and results
- discussed methodological and biological aspects more deeply

(2) Reviewer #3 suggests an alternative hypothesis to explain the weak clustering of WS substitutions within Hominadae genes, which I find very interesting. This reviewer also suggests an hypothesis to explain observations of the RSD analysis, which I think could be worth exploring.

Response: These are excellent points, which are now discussed in the revised version.

(3) Personally, I am wondering how the contribution of ancestral and lineage-specific polymorphisms could bias the estimation of substitution rates in closely related species (see e.g. doi: 10.1093/molbev/msz203)?

Response: Fair point. A paragraph discussing this and other potential biases with our substitution inference approach was added (line 230-241).

(4) Besides, the author might find this reference on DSB evolution in mice useful in relation to their own observations in mice (doi: 10.1093/molbev/mst154).

Response: This article is cited as reference [10].

REVIEWER 1's COMMENTS:

In the manuscript entitled "Fine-scale quantification of GC-biased gene conversion intensity in mammals." by Nicolas Galtier (DOI https://www.biorxiv.org/content/10.1101/2021.05.05.442789v3), the author investigates how to measure gBGC strength in 4 clades of mammals (Hominidae, Cercopithecidae, Bovidae and Muridae).

The study is well designed as it compares observed statistics such as Moran's I or substitution density to simulations with nested/increasing number of parameters. The study uses a maximum-likelihood approach to reject the simplest models as it should be done. I don't have comments concerning the experimental design. The analysis and interpretation of results are clear as well. The flaws of this manuscript rely on the text or figure's legend which do not always help the reader to

properly understand what was done and why. The presence of the scripts to make figures helped me doing the review.

Here are the specific comments:

1. Fig1: Please explicit in the legend what are the green branches (I assumed it is the studied branches but I'm unsure).

Response: Legend to figure 1 now clarified.

2. Fig2: I missed out why you need branches with at least 100 genes with each of them having at least 3 substitutions of each type. I would have expected this criterion will emphesize any signal of clustering. If you are willing to test whether there is a cluster, shouldn't you keep branches where there is a minimal number of substitution to have a signel (here, 300) regardeless they are well dispersed accross genes?

Response: This is because clustering (Moran's *I*) is measured within genes, then averaged across genes (see section 5.3). Genes carrying a very small number of substitutions have a meaningless Moran's *I* estimate and just add noise to the mean. To analyse substitution clustering across genes could be interesting too but requires a different approach. Here we aim at measuring substitution clustering at the scale of a gene conversion track length, which of the order of 100bp.

3. In several figures you wrote there is 1 dot per species. I thought you were also looking at internal branches. Do you calculate Moran's I solely at tips or in all the branches suitable for the study? Specifically, sup fig 1: you said 1 dot = 1 species but there are 8 dots in Bovidae and only 7 species. Moreover, I see only 7 blue dots and 8 red ones. Is it possible 2 blues dots are overlapping on the right? You could add a darker perimeter for the circles such that we see if 2 overlaps or not.

Response: The reviewer is correct: we analysed all suitable branches, not only terminal branches; "species" was replaced with "lineage" in the legend of Supplementary Figure 1 and Supplementary Figure 2. There are actually two overlapping blue dots in the Bovidae panel of Supp Fig 1 (the right-most two).

4. Supp Fig 1: Why is there differences between line blue and red in the simulations? Given the equations, I thought we would estimate the exact same value of B for both SW and WS substitutions.

Response: Actually simulations of WS and SW substitutions were run independently and accounting for the distribution of GC3 across genes, as mentioned in section 5.4, line 380-383. A gene with GC3=90% was 10 times more likely to host a SW substitution than a gene with GC3=10% in our simulations, as we now clarify.

5. Table S1: Please explicit what is the number of genes_cleaned. Given the explanation in the mat and methods (X->Y substitutions : All the descendants must be X on one side of the tree and Y on the rest), I don't get why we don't have the same number of genes within a clade.

Response: This reflects the fact that genes in which the estimated per base pair substitution rate was higher than ten times the across-genes median rate were discarded. A sentence was added to clarify that this can result in variable number of genes among lineages within a

family (section 5.2, line 346). This filtering step aims at minimising the impact of misaligned regions.

6. In general, many sentences are long which does not help the reader. Please rewrite at least the sentence line 349-352. It is too long and the idea is complex (I had to re-read it 6 times before getting to the point).

Response: We agree that this sentence is long and a bit difficult to follow; on the other hand this is the description of an algorithm, which can hardly be shortened. We modified the text so that the two main parts of the sentence are itemised, hopefully making the reading easier.

7. Do you have an idea why there is an outlier in Bovidae at 3.86. If I get it right it is the capra-ovis ovarie branch (see supp table). It has a huge sd (~6). Do you think it is because there is no info, few substitution and so it is difficult to estimate a signal or do you think it is a biological signal, like a huge hotspot of gBGC intensity along that branch? I would have been interested to read about it in the discussion part (like if it's biological, is it realted to anything know about their evolutionary history?).

The *B*=3.86 lineage is indeed the *Capra-Ovis* ancestral branch, as specified in the legend to figure 3. The discussion mentions that in Bovidae ancient branches are associated with a higher estimated B than recent branches, in agreement with the hypothesis of a large ancestral Ne in this group (see referenced papers by Figuet). The high estimate of *B* in the *Capra-Ovis* ancestral branch contributes to this pattern, but as the reviewer says, the associated standard deviation is rather high, so, not sure this data point deserves to be specifically discussed.

8. Models are complex and while they are well explained in the methods, I think you could help the reader by making a table summarizing the models. I had to dig to understand why f or z in models' names. In the same topic, could you make a table summarizing how many branches are rejected per model ?

A new Table 1 recapitulating the parametrisation of the 7 models used in this study was added. There is a fairly large number of possible model comparisons (M2 vs. M1, M3z vs. M1, M3s vs. M2, ...), all of which can be easily done based on Supplementary Table S1, which provides the likelihood of each model in each lineage. For instance below is the R code implementing the M2 vs. M1 likelihood ratio test:

```
> d=read.csv("gBGC_mammals_2021_tableS1.csv")
> dL=2*(d$lnL_M2 - d$lnL_M1)
> pval_M2_M1=1-pchisq(dL, df=1)
```

9. Which branch is not rejected using M3sh compared to M3h ? Is it in humans where there is small gBGC or is it somewhere else ? Do you have any comments to make on this branch ?

The branch in which M3h rejected M3sh is the *Bison bison* terminal branch (Bovidae), as now specified.

10. What do you mean by averaging the Akaike Information Criterium of each model in practice? I couldn't find it in the scripts and I am interested in understanding what you did there.

Here the across-model weighted average B is calculated, where models are weighted by the so-called AIC weights, taken as a measure of the probability that a model is the correct one (e.g. see Posada & Buckley 2004 Syst Biol). This calculation is performed at line 150-200 of script MLres.R. A new Supplementary Figure S2 was added showing the results of this analysis. The legend of this Supplementary Figure S2 explains the underlying calculation. Reference to Posada & Buckley 2004 also added.

10b. In equation 16 and 17. What is « k »: the genes or the AIC values ?

Response: *k* in these equations denotes gene category, as now clarified.

11. Fig4 : I think you represent the correlation of B and dN/dS using a log transformed scale (but the y axis and x axis values are still the values of B and dN/dS). The title is misleading.

Response: The reviewer is correct, legend to figure 4 modified.

12. Fig5 : The second sentence in the legend is unclear : « in for ».

Response: Typo corrected, thanks ("in" -> "is").

REVIEWER 2's COMMENTS:

In this work, N. Galtier estimated the strength of gBGC and investigated its relationship with Ne acrosss 4 different families of mammals. To do this, he analysed nucleotide substitution patterns in coding sequences of 40 mammalian lineages using a maximum likelihood approach. The results of the study suggest that gBGC is prevalent in these mammalian families, estimating that large proportion of WS synonymous substitution can be attributed to this process and that its strenght varies across lineages and genes depending on Ne and the dynamics of recombination hotspots.

This work joins a large body of literature that demonstrates that gBGC is a major force shaping patterns of molecular evolution. The article is well writen and I enjoyed reading it. The potencial limitation that came to mind while I reviewd the study where properly discussed, such as the fact that these results are dependent on assuming a constant mutational process across species. So, I do not have major comments for improving this manuscript.

In terms of novelty of the work, other studies have tried to estimate B accross mammalian lineages. However, most studies have estimated B from site frequency spectra. Few studies, which are cited in this manuscript, have already tried to estiamate B using substitution patterns. Specifically, Lartillot (2012) proposed an integrated Bayesian model for reconstructing the evolutionary history of gBGC, and for estimating its correlation with life-history and karyotypic traits. Nonetheless, this maximum-likelohood framework is an anlternative model that confirmed many previous studies and seems very valuable for the research field and community.

Minor comments:

line 102: I suggest editing: "As far as SW substitutions were concerned, " to "The centered Moran's I for WS substitutions "

Response: edited as suggested.

line 110: Why is there a discrepancy between the the bp scales used by the author when calculating Moran's I (400 bp) and the one used in the simulation (40 and 500 bp?) If the aim is to assess the amount of clustering needed to explain the observed values, real and simulatated data should have the same bp scale.

Response : Deciding on the exact scale for Moran's I calculation is somewhat arbitrary. Here I used the 100, 200 and 400 bp scales, while only showing results at the 400bp scale. The simulation was conducted using parameters as close as possible from the experimental literature on gBGC, so that the inferred proportion of substitutions within gBGC clusters be biologically as meaningful as possible. We made sure that the scale used for calculating Moran's *I* based on simulated data was the same as the one used for real data, thus ensuring comparability.

Figure 4: The sample size was here too small to investigate the within-family relationships. To further investigate the relationship of B and Ne, the author could show if there is a correlation betnween B and Ne-related life history traits and assess this within-family relationship. This would help strengthening the argument given that even this relationship (putative correlation between B and Ne) as judged by the correlation between B and dN/dS is weakly convincing as there are few data points within each family. Moreover, the family with the largest number of lineages is the one that shows no significant relationship.

Response : We agree that not detecting a significant relationship between Ne and dN/dS in Cercopithecidae was disapointing, and is so far unexplained. Following the reviewer's suggestion, we collected data on species body mass and longevity. These variables were significantly correlated with *B* (longevity : r^2=0.36, *p*-val=0.0026 ; log-body mass : r^2=0.22, *p*-val=0.017), while not as strongly as heterozygosity π. No significant relationship between *B* and longevity or body mass was found within Cercopithecidae. These new analyses are reported in the revised version (lines 168-169). They do not help understand the lack of a within-Cercopithecidae significant relationship, though. It should be noted that the variance in life history traits among species of Cercopithecidae is not high (see Supplementary Table S2).

Supplementary Figure 1: It was difficult to understand this plot.
It is not clear if numbers in red are shared between Bovidae and Muridae or if they are missing for Bovidae panel (same for the two upper panels).
The last sentence of the legend: "accounting for substitutions that were lost because appearing within introns or flanking regions." It is not clear in the methods how these were accounted for.

Response : The legend now says that numbers in red are not provided for Bovidae and Cercopithecidae by lack of space, but are very similar to numbers in Muridae and Hominidae. Regarding exon boundaries, the Material and Method says (line 377) : "If the sampled location of the (n+1)th substitution reached beyond the boundaries of the exon carrying the (n)th substitution, then the (n+1)th substitution was ignored."

Section 5.4 Need clarification. Contrary to the rest of the manuscript, this part was not clear.

line 344: It is uncelar what does the author mean by "randomly sample the location of the first substitution " What substitution? For the first substitution in a 4 species alignment?. For this the location in the hypothetical branch was randomly sampled ? It is also unclear what do these authors mean by "across genes and exons;" was this done once for genes (incudng introns) and once for exons? (I assume this has to do with my previous question for Supp Fig.1 so some clarification hee is needed).

Response : What we simulate here is the process of substitution in a single lineage, or branch. Conceptually, the initial state is a set of coding sequences, and the final state the same set of coding sequences with a number of substitutions at various positions (third codon positions only). We control the total number of substitutions and randomly draw their types (WS, SW, or GC-conservative) and positions. Substitution locations are drawn iteratively, hence the mention of the "first", "(n)th" and "(n+1)th" substitutions. We tried to clarify this by adding the words "initiation" and "iteration" to the description of the algorithm. We also had"across genes and exons" replaced with "among the third codon positions of all genes"

line 360: "Two parameters of the simulation procedure were varied among conditions, namely the per third codon position density of substitutions, and the probability pclust for two successive substitutions". It is not clear in the text what where the values for these parameters accross different similations.

Response : We now mention that these parameters take values in {0.0003, 0.001, 0.003, 0.01, 0.03} and {0, 0.1, 0.2, 0.3, 0.4}, respectively (line 385-386).

line 399: Empirical estimates of mutation rate in humans were used. It is possible that mutation rates vary between the investigated taxon families. Could a real difference in mutation rates between lineage lead to the observed patterns attributed to differences in B?

Response : This is discussed in the third paragraph of section 3.1. The existing literature suggests that the relative SW and WS mutation rates in the mouse (Muridae) and rhesus macaque (Cercopithecidae) are similar to the human ones, while no data is available in Bovidae.

Figure 4 and 5 could be placed in the appropiate sections.

Response : Not so easy to achieve with LaTeX, sorry for this !

REVIEWER 3's comments

In this manuscript, Galtier quantifies the strength of GC-biased gene conversion (gBGC) and its impact on protein evolution in 4 families and 32 species of mammals. He founds a substantial impact of gBGC on AT > GC synonymous substitutions (explaining ~60% of the variance). I've divided this revision into 4 sections.

1. Is the science sound, with a logical narrative and well-supported results and conclusions?

The manuscript follows a logical narrative and the methods are sound. The literature context provided in the introduction is very helpful. However, there is a key question regarding the interpretation of an important result that should be addressed before recommendation: I agree that if

recombination hotspots are more ephemeral in Hominadae than in other groups then this could explain the weak clustering of WS substitution within genes. However, there is another alternative hypothesis. Could the weak clustering of WS substitutions within Hominadae genes be due to their lower diversity? The more distant the segregating sites are, the less likely would be for gBGC to generate a cluster of substitutions in a given gene. Is there a correlation between heterozygosity at the gene level and Moran's I for WS substitutions? Hence, maybe the substitution clustering within genes is conditional on B intensity + gene heterozygosity + (local & global) Ne. I am also assuming that the gene conversion tract length is not negatively correlated to the Ne. I am not sure there is literature regarding the correlation between gene conversion tract lengths and Ne.

Ideally, genes' heterozygosity and Ne should be decoupled to assess this hypothesis (by comparing genes with different mutation rates within a genome?), which is hard. But maybe this alternative can be further discussed or elaborated by the author.

Response: Thanks for a highly sensible comment. We modified this part of the discussion (line 295-313) and in particular added a sentence stating that what determines the prevalence of clustered WS substitutions is the duration/strength of recombination hotspots and the standing heterozygosity – which also varies among taxa in a Ne-dependent way (line 308). Actually the reviewers' hypothesis is probably the simplest and most natural explanation to consider.

The reviewer's idea of testing this hypothesis by comparing small-Ne to large-Ne genes within a genome is appealing. However please note that variation in recombination rate in large part determines the variation in Ne (sensu lato) among genes via the dissipation of Hill-Robertson effects. So the situation is complex, recombination being causative on the two effects we want to correlate (gBGC sttrength and heterozygosity). All in all I have the feeling that to embark on this could be very interesting and fruitful (for instance I guess we might come up with an upper bound for the lifespan of recombination hotspots in primates, knowing their strength), but this is a different study, involving some additional modelling and the analysis of within-species polymorphism data.

2. Is there enough info to allow verifying and reproducing the data?

The supplementary information, plus the scripts, are easy to access and rerun.

3. Are there obscure passages that a potential reader can't go through?

So far the paper is easy to follow, but of course, there are always things that can be clarified. For example:

3.1 It would be good to have a table (in the main text?) with all five models (M1, M2, M3z, M3h, and M3sh), their number of parameters, and the average lnL across species. I can not find in supplementary table 1 the info regarding model M3h and the p-value of the LRTs commented in the main text.

Response: A new table recapitulating model parametrisation was added (Table 1). The average or summed lnL across species can easily be obtained from SupplementaryTable S1, in which results from model M3h were added.

3.2 I don't quite understand model M3sh. If q (Is q equivalent to the number of hotspots per gene?) approaches zero then does this mean that there are no hotspots within genes? or that hotspots occur in a very tiny fraction of the gene? Maybe the definition of model M3sh can be extended or rephrased?

Response: M3sh does not fundamentally differ from M3h; it rather reflects the observation that under M3h, there is a subset of the parameter space (q<<1) where the predicted numbers of WS and SW substitutions can be approximated with good accuracy by a simplified formula. For instance if 1% of the sites in a gene are under strong gBGC and 99% are neutrally evolving, then equations (14) and (15) hold. These equations have a simpler interpretation, also save one degree of freedom, and offer the opportunity to test the q<<1 hypothesis - which turn out to be almost never rejected with this data set. We rewrote the whole paragraph and added a sentence explaining that M3sh is a special case of M3h assuming that the fraction of sites affected by gBGC within a gene is small (line 444-445).

3.3 Line 135-136. "These were very similar to estimates obtained by averaging $\bar{B}$ across 136 the M1, M2, M3z, M3sh, and M3h models, weighting by the AIC of each model." Could it be possible to add to supplementary table 1 the AIC weights too? and have a supplementary figure equivalent to figure 3 but with the AIC weighted parameters? Just to back up this sentence with figures and tables.

Response: A supplementary figure was added as suggested (Supplementary Fig. S2). AIC weights can easily be calculated from the content of supplementary table 1, as also shown line 150-200 of script MLres.R.

4. Potential extra analysis only if interesting enough to the recommender and/or author:

4.1 Regarding the across genes RSD analysis. Is the recombination map in Muridae also more uniform than in Bovidae, Hominidae, and Cercopithecidae? That could explain the results, but I understand that the recombination map for all these groups might not be available.

Response: At a large scale (Mb) the mouse and rat recombination maps have a coefficient of variation similar to the human recombination map (Jensen-Seaman et al 2004 Genome Res). The genome-wide analysis of linkage disequilibrium also revealed similar numbers and distribution of recombination hotspots in humans and mouse (McVean et al. 2004 Science ; Brunschwig et al. 2012 Genetics). A sentence discussing this aspect was added (line 293-294).

4.2 Again regarding the clustering of mutations within genes. Would it be possible to assess whether most WS clustering is happening at first exons (the ones closest to CpG islands)? As far as I know, at least in humans, recombination hotspots tend to occur at CpG islands at the starting of genes.

Response: To address this reark, we analyzed the distribution of the location of substitutions along gene coding sequences, separately for each lineage. For each gene of length 1000bp or longer, we collected the position of substitutions and divided these by gene length. Then we merged these numbers across genes, thus obtaining a set a relative positions taking values in the [0,1] interval. We fitted a Beta distribution to this set of values using the fitdistrplus R package. The Beta distribution, which takes value in [0,1], is defined by two parameters, usually called α and β, and takes various shapes (e.g. bell shape, J shape, U shape, L shape, uniform) depending on α and β. The Beta distribution converges to a uniform distribution when $\alpha=\beta=1$. Below we show the joint distribution of α and β among the 40 analysed lineages :

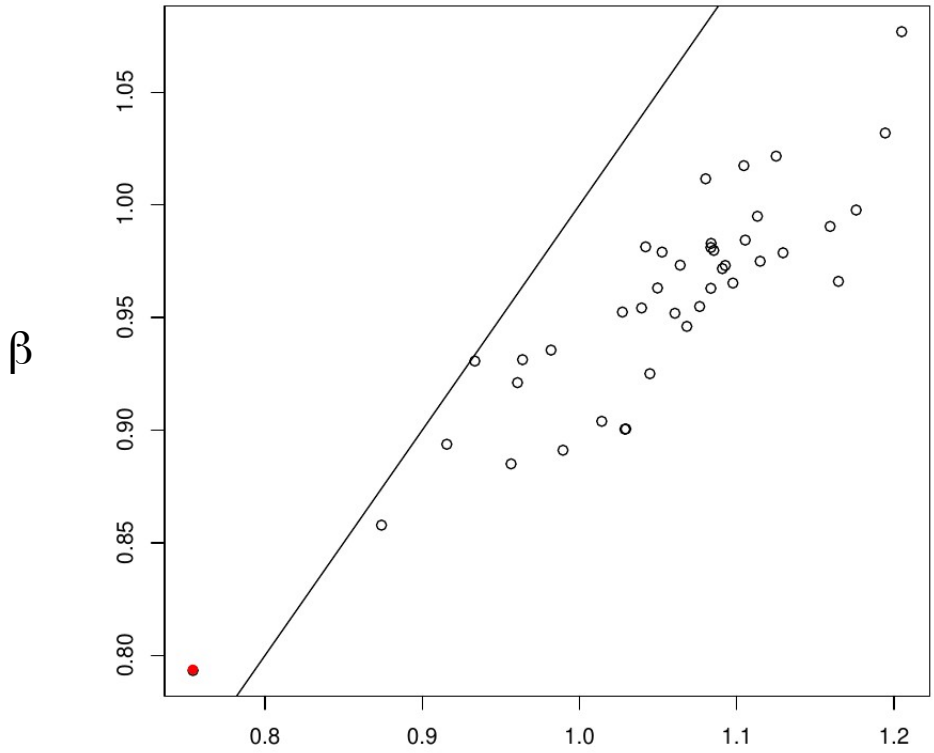

Figure R1 : Estimated parameters of a Beta distribution fitted to the relative position of infered substitutions. Each dot is for a lineage. Red dot : *Rhinpithecus bieti* terminal branch.

This figure shows that, in most lineages, α and β are very close to one, indicating that the distribution of substtitutions across coding sequences is close to uniform. β was sligthly lower than α in most branches, inicating that the typical pattern was a slight excess of substitutions near the 3' end of coding sequences.

One lineage behaved a bit differently, namely the *Rhinopithecus bieti* terminal branch (red dot in Fig. R1), in which both α and β were below 0.8 - a U-shape distribution. This peculiarity is difficult to interpret, and perhaps casts some doubts on the quality of the alignments for this species. However *Rhinopithecus bieti* did not behave differently from other lineages of Cercopithecidae in any of our analyses of gBGC. Figures very similar to the above Fig. R1 were obtained when we considered WS and SW substitutions separately.

Although this analysis might potentially be of interest, it is still quite preliminary. More work would be required to make sense out of it and disentangle the underlying biological processes from potential artefacts. We refrained from including these results in the manuscript since we thought they would mainly distract the reader from our main message.

4.3 Relative to the genome-wide excess of WS mutations due to gBGC. Would it be possible to estimate the defect of SW mutations too? In other words, it would be interesting to know the overall impact of gBCG on substitution rate taking into account that the absolute number of WS and SW substitutions might be different? Maybe controlling by GC conservative substitutions across species?

Response: Under model M3sh, which was favored in almost all lineages, no depletion of SW substitution is expected. Under this model, gBGC is locally strong but affects a vanishingly small minority of sites, so that the rate of SW substitution is unaffected (see equation 15 and the comment below it). This might be an extreme assumption to make, and therefore an extreme conclusion to draw. Some additional modeling effort might lead to a more nuanced picture. We added a sentence indicating that no depletion of SW substitutions is expected under M3sh.