INSTITUT de BIOLOGIA EVOLUTIVA CSIC upf.

Dear Dr. Roulin,
We would like to thank you for editing our manuscript.
We are also thankful for the comments of the two reviewers that we believe have help us improved the manuscript. We provide below a point-by-point response to the reviewers.
We are also providing a version of the manuscript with track-changes.
Best
Josefa González, on behalf of all co-authors

This paper constitutes a valuable resource to understand the genomic bases of local adaptation to the urban environment in Anopheles mosquitoes. It represents a substantial sequencing and analytical effort. I also appreciated the care brought to provide a detailed functional annotation of TEs regulatory and promoter sequences. My main comments are as follows:

1) The authors acknowledge and discuss this, but it remains difficult to draw strong conclusions about the role of TEs in adaptation due to the rather low sample size. However, the discussion on potential mechanisms is rather long. It may be seen as hand-waving, and I worry that it is edging toward story-telling as defined in (Pavlidis, Jensen, Stephan, & Stamatakis, 2012). It may be worth toning down these parts.
>>>We agree with the reviewer and we have revised the manuscript to avoid overstating the claims about the role of TEs in adaptation. We completely agree that the sample size is small and that more data is needed to reach any conclusions. We have also changed the title so that it better reflects all the analysis performed in the manuscript. We do think, however, that our data is indicative of the potential role of TEs at least in the two biological processes analyzed: immunity and insecticide resistance.

2) However, a possible way to provide more support for a role of TEs in adaptation may be to examine patterns of SNP variation at their flanking genomic sides. This may require substantial analytical time, and I would understand it if the authors declined to go forward with this suggestion. A possible way would be to run a software such as ARGWeaver (Hubisz, Williams, & Siepel, 2020; Rasmussen, Hubisz, Gronau, & Siepel, 2014) to extract times since coalescence (TMRCA) along the genome, and test how these times differ at the vicinity of insertions that may be candidate for local adaptation. With six genomes, the algorithm should run rather fast, and this would bring a lot of information to at least check whether TMRCAs drop drastically near a candidate insertion.
>>>We thank the reviewer for his suggestion. The number of genomes analyzed is too small to estimate statistics such as iHS or H12 that we currently used for looking for signatures of selection in regions flanking candidate TE insertions (see for example Rech et al 2019). We have followed the reviewer suggestion and we have evaluated whether ARGweaver would be a good tool to look for evidence of selection in our dataset. We finally decided to not include the ARGWeaver results as only one of our genomes is an individual genome. The other five genomes were generated from a pool of six larvaes. Another caveat is that we generated the genomes using long-reads, and as such the quality of reads is not as good as the Illumina short-reads that is normally used to run programs such as ARGWeaver.

Edifici CMIMA
Psseig Marítim de la Barceloneta 37-49
08003 Barcelona

Tel: 93 230 95 07
FAX: 93 221 10 11

Edifici PRBB
Doctor Aiguader 88
08003 Barcelona

Tel: 93 316 0805
FAX: 93 316 09 01

3) In relation to point 2), the authors do not genotype insertions nor SNPs but look at presence/absence and reconstruct a haploid genome assembly if my understanding is correct. It might be useful to call variants and TEs (maybe with MELT? (Gardner et al., 2017)) to examine, for example, variation in the allele frequency spectrum of TEs in relation with features such as recombination. Given the potential importance of background/linked selection in shaping the average frequency of TEs, examining correlations between TE density/frequency with recombination might be relevant. This would help discussing whether indirect effects of selection may impact the identification of TEs under positive selection.

>>>We did genotype the insertions that were located nearby genes. In this new version of the manuscript, we have extended our genotyping to all insertions in the genome (see below).

We agree that examining the variation of the TE allele frequency spectrum in relation with features such as recombination rate would be interesting. However, estimates of recombination rate, have only been recently reported for *Anopheles coluzzii* and *Anopheles gambiae* (bioRxiv preprint doi: https://doi.org/10.1101/2021.02.04.429659), and these estimates are still not available (we have contacted the authors but we have not got a response yet). Although ARGweaver estimates recombination rates, the authors warn that the estimation of the recombination rate "is not particularly well estimated by the SMC (sequentially Markov coalescent)" (https://github.com/mdrasmus/argweaver/blob/master/src/arg-summarize.cpp). Besides, as mentioned above, our data is not the ideal data to run ARGWeaver. Thus, we can not perform a detailed analysis of the correlation between the recombination rate and the TE density/frequency. What we can do is test if the density of TEs is higher in heterochromatin compared with euchromatin as it is expected because the recombination rate in heterochromatin is lower. We found that this was the case (chi-squared test *p-value* < 0.05 in all seven genomes). We have now added this test to the manuscript in line 236.

I also have a few minor comments listed below.

L100: citation not in the same format than other citations

>>>Thanks for pointing this out to us. It is now corrected.

L194: It may be worth explaining what sort of threshold is used to group TEs in a family (check Methods section)

>>We have re-written the explanation of the thresholds used to group TEs into families in the Methods section (line 613).

We also now specify that we require consensus sequences to have a minimum length of 100bp (line 601).

L165: I am a bit lost here, sorry. Should not there be an overlap between these 85 and the 435 families identified previously?

>>>We have re-written this section to clarify this issue (Line 167). To construct a TE library as complete as possible, we first identified 435 TE families using a fully *de novo* strategy using REPET. Additionally, using an homology-based approach (Repeatmasker), we also found that 85 previously described families from other members of the *Anopheles* genus available in *TEfam*, were also present in our genomes. These 85 TEs were either not predicted by REPET or were discarded during the manual curation stages. We thus

considered both the 435 families detected by REPET and the 85 families identified by RepeatMasker to annotate our genomes.

L182: Can you perhaps examine whether these TRIM elements have the said impact on Anophele genomes, since you have long reads?
by acting as target sites for retrotransposon insertions, alter host genestructure, and transduce host genes
>>> To assess whether TRIM elements are likely to alter host gene structure, we tested whether the number of TRIM insertions in exons and introns is higher than expected compared with the rest of the insertions in the genome. We found that TRIM insertions are underrepresented in gene bodies (chi-squared test, *p-value* < 0.05) thus suggesting that they do not affect gene structure. We also test whether TRIM elements could be acting as target sites for retrotransposon insertions. We found that TRIM elements were found to be nested more frequently than expected compared with other TE families (chi-squared test, *p-value* < 0.05).
We have added these two tests to the manuscript (line 188).

L204: What about other features such as recombination rate?
>>>Please see our answer above

L208: I understand why, but it may be worth explaining that you want to focus on well-assembled regions, where TEs may have a functional impact. Especially because you just said that most of the variation is in heterochromatin, so why looking at euchromatin?
>>>Following the reviewer's suggestion we have now added to the manuscript that we focused on euchromatic regions because we are mostly interested in the functional impact of TEs (Line 218).

L209: p-value -> p in italics
>>>Thanks for pointing this out to us. It is now corrected

L281: I feel like this whole part could come earlier. For example the sentence L285 makes it clear that LTRs have been recently active and this echoes their large variance in copy number shown L210.
>>>We agree with the reviewer's that there are several ways of organizing the results. We thought it might be a good idea to bring together all the sections of the manuscript that deal with potential impacts of the TE insertions annotated in this work. We now point out that the recent activity of LTRs echoes their variance in copy number mentioned before in the manuscript (line 310).

L302: This is where I thought that it would be nice to have the frequency
>>>We agree with the reviewer, and that is why we manually curated all the insertions from active families that are nearby genes so that we could estimate their frequencies. In the previous version of the manuscript, we focused on transferring only those TEs located nearby genes. In this new version, we have transferred all the euchromatic TE insertions to the reference genome. We were able to transfer to the AcolN1 genome 71.93% to 75.29% of the regions surrounding the TEs in each of the six genomes leading to a total of 67,548 regions with present or absent TE insertions. These regions were transferred from the AcolN1 reference to the remaining six genomes and we were able to

Edifici CMIMA
Psseig Marítim de la Barceloneta 37-49
08003 Barcelona

Tel: 93 230 95 07
FAX: 93 221 10 11

Edifici PRBB
Doctor Aiguader 88
08003 Barcelona

Tel: 93 316 0805
FAX: 93 316 09 01

confidently transfer 53,893 (79.78%) of these regions to at least three of the genomes with 32,185 (47.65%) transferred to all six genomes. We checked whether the TEs that failed to be transferred were enriched for nested TEs and we found that this was the case: while 68.58% and 51.18% of the TEs located more and less than 500 bp away from other TEs, respectively, were transferred, only 38.26% of the TEs overlapping other TEs were transferred to the six genomes.
We have updated the material and methods accordingly (Line 632).

L548: Maybe explain whether these are the default parameters, or why you chose these specific ones.
>>>We have added this explanation in the Material and Methods section (Line 569).

L692: A/B scores > support? Maybe provide more details.
>>>We have added these details in the Material and Methods section (Line 721).

Figure 1: It would be interesting to see how abundant each new family is when adding new genomes. In other words, are we mostly adding families with a single element? Maybe provide a graph with the total length or number of TEs, in relation to the number of genomes added, irrespective of family?
>>>We explored this possibility and observed that a median of 76% of the total copies were identified when using a single genome. This percentage rapidly increases when adding other genomes thus suggesting that families that are only identified when using more than three genomes have a low number of copies. We have added this information to the manuscript and we have added the figure as supplementary material (Line 1222).

Figure 2: I do not find this representation very clear nor informative. In panel A, maybe provide the length or a scale for each element? Panel B is hard to interpret. Panel C could benefit from a phylogenetic representation of relationships between species. In general, the figures I had were too compressed, hard to decipher.
>>>We have followed the reviewer's suggestions to improve Figure 2. We have added the length of the insertions in Panel A. We have modified the legend of Panel B and we have added the phylogenetic representation of relationships between species.
We have replaced Figure 2 with an updated version.

Thank you for your work, I hope you find the comments constructive.
All the best,
>>>Thanks so much for taking the time to evaluate our manuscript. Your comments were indeed constructive.

References

Gardner, E. J., Lam, V. K., Harris, D. N., Chuang, N. T., Scott, E. C., Pittard, W. S., … Devine, S. E. (2017). The Mobile Element Locator Tool (MELT): Population-scale mobile element discovery and biology. Genome Research, (410), gr.218032.116. doi: 10.1101/gr.218032.116

Hubisz, M. J., Williams, A. L., & Siepel, A. (2020). Mapping gene flow between ancient hominins through demography-aware inference of the ancestral recombination graph. PLoS Genetics, 16(8), 1–24. doi: 10.1371/JOURNAL.PGEN.1008895

Edifici CMIMA
Psseig Marítim de la Barceloneta 37-49
08003 Barcelona

Tel: 93 230 95 07
FAX: 93 221 10 11

Edifici PRBB
Doctor Aiguader 88
08003 Barcelona

Tel: 93 316 0805
FAX: 93 316 09 01

Pavlidis, P., Jensen, J. D., Stephan, W., & Stamatakis, A. (2012). A critical assessment of storytelling: Gene ontology categories and the importance of validating genomic scans. Molecular Biology and Evolution, 29(10), 3237–3248. doi: 10.1093/molbev/mss136
Rasmussen, M. D., Hubisz, M. J., Gronau, I., & Siepel, A. (2014). Genome-Wide Inference of Ancestral Recombination Graphs. PLoS Genetics, 10(5). doi: 10.1371/journal.pgen.1004342

Impact of transposable elements on the genome of the urban malaria vector Anopheles coluzzii
Carlos Vargas-Chavez et al

GENERAL COMMENTS

In this work, authors have analyzed the TE content of the malaria vector An. coluzzii, a sister species to the well characterized species An. gambiae. Their work has one major originality: the use of long-read sequencing strategies that allow to span full TE insertions. Authors have used these data to annotate An. coluzzii TEs, including the discovery of new TE families, have studied their distribution throughout the genome, and shown their presence at chromosomal inversion breakpoints. They have also studied TEs associations with genic regions as well as the presence of transcription factor binding sites and promoters in some TEs, and shown that some insertions were associated with insecticide-associated or -responsive genes, or with genes putatively involved in immunity.

While this work has thus provided a wealth of data that are potentially very useful and open roads for further studies, I fear that the manuscript contains too many flaws to be recommended as it stands for a support by PCI Genomics.

My major criticism is that authors constantly overstate the meaning of their results. The work is presented from the start as an analysis of the role of TEs in An. coluzzii adaptative processes and of their impact on genic functions, and sentence like "To better understand the role of TEs in rapid urban adaptation, we sequenced..." as well as "we found that TEs have an impact in ...the regulation of functionally relevant genes", are even found in the Abstract.
Yet nothing in the results provides any hints on the role of TEs in rapid urban adaptation, nor demonstrate any impact on gene regulation.
>>>We thank the reviewer for the comments on the manuscript. We agree with the comments and we have revised the writing throughout the text to ensure that we are not overstating our results. We have also modified the title of the manuscript so that it better reflects all the analysis performed in the manuscript.
However, we do think that the presence of regulatory regions in TE insertions located nearby genes that are known to affect the vectorial capacity of Anopheles is indicative, although not conclusive, of the potential functional effect of these insertions. We also agree that further analysis are needed before reaching a conclusion on the adaptive role of these insertions, as we mention in the discussion and the conclusion sections.

Edifici CMIMA
Psseig Marítim de la Barceloneta 37-49
08003 Barcelona

Tel: 93 230 95 07
FAX: 93 221 10 11

Edifici PRBB
Doctor Aiguader 88
08003 Barcelona

Tel: 93 316 0805
FAX: 93 316 09 01

More specifically:

1) Authors have sequenced different larvae from two sampling sites, and interpret differences as true genomic differences and TE variability throughout the manuscript. For instance, authors have found large differences in the gypsy content. However authors have not evaluated whether these differences could simply be due to variabilities during the sequencing process (such as the large coverage differences between samples or other technical aspects).

>>> To discard a direct link between TE content and technical sequencing variability we used linear regression models to evaluate whether several parameters reflecting differences between the samples such as differences in long-reads coverage, long-reads N50, number of contigs in the assembly, N50 of the contigs, complete BUSCO values and number of successfully transferred genes have an effect on assembly statistics, scaffold statistics, TE content, or TE family abundance. These analysis have been added to the manuscript and are provided in Supplementary Table S1. We did not find that any of the sequencing parameters significantly affected the observed differences in total TE content or in superfamily abundance.

Some sequencing replicates for individual larva samples should have been performed to adress this issue, but are mentioned nowhere. In the absence of such replicates, these inter-sample differences in global TE contents or in specific superfamilies have little meaning. More generally, all inter-sample quantitative aspects of the results, including TE insertions distribution throughout the genomes should be considered very cautiously. In the Discussion, authors merely mention the possibility of heterochromatic TE differences resulting from of possible differences in the quality of the genome assembly, but discard this possibility for euchromatic TEs, for no clear reason.

>>> Given the characteristics of our samples (larvaes were collected in natural breeding sites and kept in ethanol), it was impossible to obtain enough DNA to perform sequencing replicates. We agree with the reviewer that results should be considered with caution. However, while sequencing-related variables such as long-reads coverage or long-reads N50 are marginally significant for TE content in heterochromatin (long-read coverage: p-value = 0.01 and FDR p-value = 0.08), these variables were not significant for TE content in euchromatin (long-reads coverage p-value = 0.55 and FDR p-value: 0.74). We have revised the manuscript accordingly.

Exemples of TE variations identified between sequenced genomes are provided as references, but many of those variations were actually analyzed via split-reads strategies, allowing to securely determine that insertions are truly absent or present from determined genomic positions.

Similarly, what is the point of analyzing the TE landscape separately for each larva ?

>>>There are several strategies to determine if TE insertions are truly absent or present in a given genome. In our case we *de novo* annotated TEs in each of our genome assemblies. To validate that a TE was indeed absent or present in our genomes, we manually inspected the regions which contained TEs nearby genes of interest, thus we are confident that the TEs we report as present are indeed present and those reported as absent are truly absent. Additionally, we analyzed the TE landscapes independently for each breeding site as these represent different individuals collected several kilometers apart.

Edifici CMIMA
Psseig Marítim de la Barceloneta 37-49
08003 Barcelona

Tel: 93 230 95 07
FAX: 93 221 10 11

Edifici PRBB
Doctor Aiguader 88
08003 Barcelona

Tel: 93 316 0805
FAX: 93 316 09 01

2) Furthermore, the study of TEs association with genic regions, and the presence of transcription factor binding sites and promoters in these elements, is merely descriptive and rather anecdotical, and does not really provide any evidence of any impact on genes. A true first step towards such evidence would have been the demonstration of an enrichment of TEs and their TFBS/promoter regions in insecticide-associated or immunity genes, which is not the case in this paper. Only the presence of TEs near such genes is shown, but these TEs could also be present or even enriched near other gene classes. Such larger scale enrichment analysis could be done, as authors have the genomic data.
>>>We have revised our writing to make sure that it is clear that we are not claiming that insecticide and immunity are more affected by TEs. Indeed, we have observed no enrichment of TEs located nearby genes involved in insecticide resistance (chi-squared test, p-value = 0.24) or located nearby immunity associated genes (chi-squared test, p-value = 0.34). Although overall there is not an enrichment, this does not preclude that some of the insertions might be affecting the expression of their nearby genes as has been previously reported. Our objective is to pinpoint a subset of TEs located nearby functionally relevant genes that might have a potential role in the regulation of these genes.

In summary, the characterization of TE families from long-read sequencing is quite interesting, authors mostly need to water down excessive claims, make a better use of their data and reorientate their paper.

In addition, I have noted a number of other flaws that will need adressing (some more technical comments at the bottom).

OTHER COMMENTS

3) Most primary Figures are very poor quality (letterings are fuzzy and barely readable), better quality figures should be uploaded in the manuscript. Supplementary Figures are OK.
>>>Thanks for pointing this out. We have now made sure that the quality of the figures is better.

4) The material sampling is quite confusing. As I understand from Material and Methods, in addition to a single larva (LVB11) used for PacBio, six larvae from each of the five other breeding sites were used for Nanopore sequencing.
- Why then mention that 25 larvae were initially sampled at each site if only six (or one in the case of LVB11), plus one used for complementary Illumina, were actually used from each sampling site?
>>>We agree with the reviewer and we have now removed this information from the manuscript.

- line 535, were the 5 larvae from each site barcoded separately? This would agree with the fact that DNA was extracted separately from each larva. Yet results are only presented per sampling sites. Or were the 5 larvae DNA pooled, with each sampling site barcoded for pooled Nanopore sequencing (in this case, why extract DNA separately from each larvae ?)
Clearly some clarity is needed here.

Edifici CMIMA
Psseig Marítim de la Barceloneta 37-49
08003 Barcelona

Tel: 93 230 95 07
FAX: 93 221 10 11

Edifici PRBB
Doctor Aiguader 88
08003 Barcelona

Tel: 93 316 0805
FAX: 93 316 09 01

>>>Yes, the larvae from each site sequenced using Nanopore were barcoded separately. However, the individual throughput was too low to perform individual genome assemblies, thus we joined all data from each site to perform the genome assemblies. The DNA was extracted separately given that we wanted to use the minimum number of larvaes possible while also keeping track of every individual extraction in case the quality of the data was too low and we had to remove it before assembling. We have added this information to the manuscript (line 529).

5) Some explanation on the PCR used to identify bona fide An coluzzi is required (I understand it is based in the presence/absence of a given SINE insertion, is that right ?)
>>>The reviewer is correct, the identification of An. coluzzii is performed using a PCR to identify a SINE insertion. We have revised this information in the manuscript (Line 520).

6) Between 172 and 294 TE families are detected within each genome, but clustering TE libraries from several genomes leads to 435 TE families. This raises many questions and should be better discussed. What is exactly the definition of a TE family in this work? Were families defined as in the Wicker's classification (80-80-80)? Could additional families be merely slightly divergent versions?
>>>The reviewer is right, families were defined using a slightly modified Wicker's classification as we required that minimum length was 100bp, thus insertions were considered as members of the same family if they were longer than 100bp, and shared at least 80% sequence identity over 80% of their length. We have now added this information to the manuscript (Line 599 and 613).

7) Some major data discrepancies are found between Figure 2 and Figure S1: both contain the two new TE families AcoI_LTR_Ele3 and AcoI_LTR_Ele4, but plot coverages differ strongly, as well as species distribution (as far as I can tell, since species abbreviations are not readable on Figure 2). Please check again that all TE identifications and corresponding data are OK.
>>>Thanks for pointing this out to us. It has now been corrected.

8) Lines 328-329: what is exactly the meaning of "...the number of genomes where the gene was correctly transferred."? From Table 1 the vast majority of genes (>95% except DLA155B) were detected in the genome assemblies. Yet from Table 3 it seems that the vast majority (all but AGAP002633) of the TE-targeted genes were not detected in all seven assemblies. This actually strenghten the hypothesis that most of the variability between samples is due to technical artefacts.
>>>This refers to the number of genomes where the whole gene structure (correct number of exons) was correctly transferred. The reviewer is correct to point out that most of the genes were detected in most genome assemblies, however this transfer also takes into consideration partial hits. For this analysis, we manually tried to identify missing exons whenever possible to rescue the whole gene structure before comparing between genomes. However in many cases this was not possible. If as we suggest these TE copies are indeed active they might be present in only some of the individuals used to perform each genome assembly, thus complicating the assembly of these regions. We have modified the legend of Table 3 (Line 337).

Edifici CMIMA
Psseig Marítim de la Barceloneta 37-49
08003 Barcelona

Tel: 93 230 95 07
FAX: 93 221 10 11

Edifici PRBB
Doctor Aiguader 88
08003 Barcelona

Tel: 93 316 0805
FAX: 93 316 09 01

9) Abundant new TRIM families - Plot coverage patterns in Figure 2 suggest a higher abundance of LTRs versus full copies. Have full length autonomous version of these TRIMs, that would share similar LTRs, been found ? Alternatively, do these plot coverage patterns indicate an abundance of solo-LTRs ?
>>> The reviewer is right, the plot coverage patterns suggest an abundance of solo-LTRs. Regarding the full length autonomous version of the TRIMs, we did not identify the full length element for any of them. The LTRs of the six TRIMs were not found in any other element in our library. We have now modified figure 2 legend (Line 1234).

10) Table S13, Promoter motifs in TE insertions: what is the meaning of "promoter motifs at the correct distance". The correct distance for expression of the gene? Retrotransposons are supposed to contain promoters (and TFBS) for their own expression. More globally, it is very likely that TEs that are NOT in close proximity to genes also contain various TFBS and promoters.
>>>The supplementary table stated "a pair of the necessary motifs at the correct distance" which refers to the TATA box and either the Initiator or TCT motifs where the TATA box is at the expected distance from either of the other two elements.
We have corrected the header in supplementary table S13 to better describe the information it contains.

11) Identification of active families: a more straighforward way of identifying active retrotransposons and their relative age would be to evaluate the divergence between LTRs of the same copy.
>>>We appreciate the suggestion of the reviewer. We evaluated the divergence between the LTRs of the same copy of potentially active families. The mean identity of the pairs of LTRs in all families was higher than 97% in all cases. Additionally, we observed several pairs that were identical which would suggest that these copies are in fact very recent. We now reported this values in a new supplementary table and we have added the information to the manuscript (Line 305).

12) Lines 353-366: It is unclear why authors focused only in intronic or upstream insertions, as downstream insertions can also affect gene expression, notably RNA stability.
>>>We did not focus only on intronic or upstream insertions. However, this information was only included in supplementary table S14. We now make this information also clear in the main text (Line 375).

13) Figure S5: out of the 43 genes analyzed, only 23 contained at least one insertion:
- what is the point of showing diagrams for the 20 genes that do not contain any insertion? This is quite superfluous and non informative.
>>>We agree with the reviewer, we have removed figures without insertions.

- which one of those genes were differentially expressed in An. gambiae when exposed to insecticides ? As I understand, some of the genes are insecticide resistance genes, some are differentially expressed when exposed to insecticides, but are not necessarily insecticide resistance genes.
>>>The reviewer is correct, while most of the genes included in the list are differentially expressed when exposed to insecticides some of them are also known to participate on the

Edifici CMIMA
Psseig Marítim de la Barceloneta 37-49
08003 Barcelona
Tel: 93 230 95 07
FAX: 93 221 10 11

Edifici PRBB
Doctor Aiguader 88
08003 Barcelona
Tel: 93 316 0805
FAX: 93 316 09 01

insecticide resistance. We now mention this in the manuscript and in the header of supplementary table S14.

- this differential expression was described in An. gambiae. Was it correlated to any TE insertion in An. gambiae ?
>>>No, this has not been correlated to TE insertions in *An. gambiae*. However, there is an example in An. funestus which is mentioned in line 372 of the manuscript.

OTHER MORE TECHNICAL POINTS

14) Line 299: Additional file 1: Table S ? (Table number is missing)
>>>We thank the reviewer for pointing this out, we have corrected this.

15) Additional file 2 (Figure S1) is difficult to analyze, especially in C diagrams (TE distributions in other species): species names are indicated only below a few diagrams, and only in an abbreviated form that is nowhere explained in the manuscript. I would suggest to add species abbreviations to each diagram, and to add full species names somewhere in the legend.
Same remark for Figure 2
>>>We have followed the reviewer's suggestions and improved both Figure 2 and Fig. S1.

16) Supplementary Table S12 contain color highlights that are not explained
>>>Thanks for pointing this out to us, we have corrected this.

17) Table S15: "columns with the genome names can be TRUE or FALSE if the insertion is present or not" is quite obscure, replacing TRUE/FALSE by YES/NO would be more clear.
>>>We have followed the reviewer's suggestions and fixed this.

18) Lines 428-429: "four TRIM families previously undescribed in anopheline genomes that are likely to be important players in genome evolution". While it is known that TRIMs are often important players in genome evolution, this particular sentence suggests that this has been demonstrated for the four TRIM families identified here. This is not the case. This turn of phrase needs to be corrected.
>>>We agree with the reviewer, we have now corrected this.

19) The number of genes detected in An coluzzii samples is provided only in Table S1 (and thus hard to find) and only a completeness ration is provided in the manuscript core (Table 1). Gene numbers should be added to Table 1.
>>>We have added this information to Table 1.

20) Lines 265-266: It is not very clear what authors mean by "our genomes have the standard conformation for all five inversions" on lines 265-266. It seems to me that from Table S8 some breakpoints are missing from several samples?
>>>The reviewer is correct, we were unable to identify all breakpoints in all the individuals we assessed. However, all of the individuals that we were able to analyse were in standard conformation. We have revised our writing to ensure we are not misleading the reader (Line 275).

Edifici CMIMA
Psseig Marítim de la Barceloneta 37-49
08003 Barcelona

Tel: 93 230 95 07
FAX: 93 221 10 11

Edifici PRBB
Doctor Aiguader 88
08003 Barcelona

Tel: 93 316 0805
FAX: 93 316 09 01