

Dear PCI Evol Biol recommenders,

We are thankful for the detailed reviews and comments on our manuscript. We have deposited on bioRxiv an edited manuscript that, we believe, addresses all the points raised by the reviewers. Below, we provide a detail answer to each of these points.

Sincerely yours,

On behalf of all authors,

Julien .

| Dear Julien and colleagues,

| Many thanks for submitting your manuscript “The transfer of a mitochondrial selfish element to the nuclear genome and its consequences”.

| I now have in hands two reviews, that are both positive on your work, but also highlight some potential points of improvements.

| I concur with the suggestions they have made. In particular, I feel a dN/dS analysis, as suggested by reviewer 1, may give support to your suggestion that the inserted gene may be on its way to pseudogenisation. I also found that reviewer 2 made some suggestions that could lead to strong improvements, and would also make your paper more accessible to non-specialists. Just like him, I was wondering what kind of selective forces may turn a HEG to an intron. In that respect, it may be useful to clarify if this type I intron is still able to act as a HEG in the lineages where it is present. In other words, is it correct to denote this elements as a selfish element in its original location and in what respect? It may also make sense to ask why this particular nuclear site became a new insertion site. Can there be any prediction on the specificity of insertion sites? I also concur with Jan to say that horizontal transmission from a different species appears to be the easiest scenario; especially considering the fact that natural strains don't carry the mitochondrial or the nuclear version. Correct?

A1: We have addressed these concerns in the revised manuscript. We added new dN/dS analyses for the two *UMAG_11064* and *UMAG_11065* genes. We also tried to clarify that the HEG does not “turn” into an intron, but rather that a HEG that insert into an intron have a higher chance of survival and that the combination Group I intron + HEG becomes a selfish element. Interestingly, coevolution between the intron and the HEG has been reported by several studies, as the HEG can evolve a maturase activity facilitating the folding of the intron and its excision. HEGs are evolving very rapidly, and are often founds as pseudogenes. This is the case of the closest homologous sequences of *UMAG_11064* that we studied here in *S. reilianum* and *S. scitamineum*, which have presumably lost their homing activity.

We are unfortunately unable to provide more insights regarding the nuclear site. As we state in the manuscript, downstream *UMAG_11064* we only find a few telomeric repeats in the genome assembly, and the recognition site of the HEG is unknown. Finally, we now present the HGT hypothesis on the same ground as our proposed scenario, discussing the available evidence in respect to both scenarios equally. As suggested, we added a discussion and references about the frequent occurrence of HGT of HEG in Eukaryotes. We note, however, that the new phylogeny of *UMAG_11064* is compatible with the species phylogeny of smut fungi, grouping the intronic sequences of *S. scitamineum* with that of *S. reilianum*. The support for this grouping is not very strong and the intronic HEG of *S. scitamineum* appears to be heavily pseudogenized, leading to a rather long branch (new Figure 2), but if this phylogeny is correct, it would suggest that the insertion might not be that recent (*S. scitamineum* and *S. reilianum* diverged 13 My ago (Schweizer et al. 2018)).

I also found that not enough emphasis was given to the finding that natural strains don't appear to carry this insertion. If this was confirmed, the data may be interpreted as the result of lab rearing conditions, that may allow a slightly delirious mutation - such as this insertion - to be maintained, because of reduced population size or special environmental conditions? I would also suggest computing a tree of the various homologues; this may help the reader to understand the various plausible scenarios.

A2: We have now emphasized this aspect in the conclusion. We agree, however, that this would need to be confirmed as we do not have access to wild strains from the original population of the reference strain, which originates from a field isolate in the US. Previous studies, based on micro-satellite analyses, suggested a strong population structure between isolates from Mexico and the US (Munkacsi et al. 2008). We unfortunately do not have sequences from other field isolates from the original population of the reference genome and, therefore, cannot exclude that they contain the nuclear HEG as well. As suggested, we added a phylogenetic tree (new Figure 2).

Finally, I have some minor remarks that are listed below.

Provided that these various comments are taken into account, including those mentioned in their evaluation, but which I have not reported here, I think your manuscript can be made suitable for recommendation by PCI.

Hoping that these comments will seem relevant to you.

Sincerely yours,

Sylvain Charlat.

Minor remarks

L57: "As the recognized sequence is highly specific, the insertion typically happens at a homologous position": it should be made clearer that specificity could target any region; but that only those targeting homologous positions do invade. Correct?

A3: Our intended meaning is that recognition sequences are usually long and occur rarely along the genome. Free-to-invade sites are, therefore, in most cases located at a homologous position devoid of HEG (heg- allele). This sentence has been rephrased as: "As the recognized sequence is typically large, its occurrence is rare and the insertion typically happens at a homologous position."

L73 (and abstract) "all kingdoms of life": what living groups are referred to here? All domains (bacteria, archaea and eukaryotes) or more specifically different groups within eukaryotes?

A4: We meant "domains of life". This sentence is now corrected accordingly.

How do HEG invade mitochondrial genomes? Is there an equivalent to homologous double strand break repair in mitochondria and chloroplasts?

A5: This is a very good point. Mitochondria have been shown to be able to maintain their genome integrity by double-strand break repair and homologous recombination. The mechanistic details, however, are less known than for their nuclear counterpart, and presumably involve nuclear-encoded factors.

L126: I assume this is nucleotidic identity? Please confirm.

A6: Indeed, this is now explicitly stated.

L127: "Two other very similar sequences..." are they also HEG?

A7: The sequence matches correspond to mitochondrial genome sequences. They are located within the first intron of the *cox1* gene and given their high similarity, their HEG origin is very likely. The corresponding sequences are, however, not annotated as such.

| => A tree showing the different homologs, their assigned functions and origins may be helpful here. This tree should also show the branch where the frameshift most likely occurred.

A8: We are thankful for this suggestion. A phylogenetic tree has been reconstructed and is now presented on the new Figure 2, together with dN/dS analyses. This figure supports our extended discussion about the possible evolutionary scenarios. Interestingly, the phylogeny suggests that the transfer might not be as recent as we previously thought, supporting comment 7 by reviewer 2.

| Sup tabs: I have had difficulties when trying to visualise the tables because there are commas inside definition fields; tab delimited fields would make it easier to read.

A9: This was updated as suggested, and the files renamed as 'tsv'.

| Are there good reasons to believe that the ancestor HEG targeted this insertion site?

A10: We are unfortunately unable to answer this question, since the target sequence recognized by the HE, unfortunately, is not known.

| L177: "suggesting that the latter was truncated because of the *UMAG_11064* insertion." But why would the insertion generate the truncation; may be "following" instead of "because" would be more appropriate?

A11: Figure 2 (now 3) shows that the region showing similarity with HEG genes starts immediately after the stop codon of the *UMAG_11065* gene. We therefore think that the HEG sequence inserted within the ancestor of *UMAG_11065*, introducing an early stop codon. So the insertion of the HEG did not result in the formation of an intron in the *UMAG_11065* gene, whose predicted transcript has been confirmed by RNAseq. While the potential 3' end sequence of the original gene might only have been removed from the genome at a later time, it was effectively "removed" from the *UMAG_11065* gene at the moment of the insertion. We have edited the manuscript with the hope to make this point clearer.

| L184: "Interestingly, this gene family also contains the gene *UMAG_03394*..." In what sense is this interesting? Is there an implicit that the reader should make?

A12: This sentence was removed, as it was erroneous: chromosome 3 also contains 2 helicases, chromosome 9 is therefore not exceptional in that respect.

| L197: "The *UMAG_11072* gene..." what does this information tell? Is this a positive control? Of what exactly?

A13: The *UMAG_11072* gene is a positive control to assess that the lack of amplification of *UMAG_11064* in some strains is not due to any issue with the quality of the extracted DNA. This is now stated explicitly in the manuscript.

| L202: wouldn't a secondary loss be equally likely?

A14: This is a good point, and we rephrased the sentence as: "These results suggest that, either the *UMAG_11064* gene was ancestral to all tested strains and lost in the Mexican strains, or it inserted in an ancestor of the two strains 518 and 521, after the divergence from other *U. maydis* strains, an event that occurred after the domestication of maize and the spread of the associated pathogen, 6000 to 10000 years ago (Munkacsi et al. 2008)"

One question that could be addressed in the discussion: is there a link between the loss of this intron in CO1 and this nuclear insertion?

A15: Following a point raised by reviewer 2, in the revised version we state more clearly that the loss of the intron and the insertion in the nucleus position are two distinct events.

With regard to the scenario proposed in figure 7: it seems to me that one should highlight that there is basically no selective explanation for any of the transitions that seem to have taken place. Why did this insertion remain? Why did CO1 lose this intron? Why was UMAG_11065 gene shortened?

A16: Figure 7 explicitly mentions that the different steps may be explained by genetic drift alone, and only possibly favoured by a selective advantage of the mutated forms. This is now additionally stated in the figure legend. We note, however, that if the HEG targeted the insertion motif, its initial invasion might have been subject to genetic drive.

L244: “but the former cannot have happened...” unless the nuclear insert comes from another species?

A17: We amended the sentence to now state “under this scenario”, and added a discussion about the possibility of acquisition by HGT.

L249: not clear why this hypothesis is not presented as the most likely, on the basis of these divergence levels...

A18: In the revised version of the manuscript, we have added a discussion about the occurrence of horizontal transfer of HEG, as suggested by reviewer 2, and we now present the two scenarios without favouring one over the other.

I am not sure the arguments are strong enough to suggest that this insertion has any effect, either positive or negative, on fitness. The fact that it appears to be polymorphic does not argue for a strong positive effect anyhow.

A19: We agree and are very cautious in our interpretation of the fitness assays, which are conducted under laboratory conditions. We emphasized this aspect in the revised version. We note, however, that the lack of US field isolates does not allow us to definitely conclude on the frequency of the insertion in “natural” populations.

L287: “It likely represents a snapshot of evolution, when a mutational event occurred, but selection did not have time yet to act.” I don’t get this idea. Is it argued that there are selective effects, but very mild ones?

A20: The conclusion was rewritten. The corresponding sentence now reads as: “The absence of a GIY-YIG HEG in any field isolates of *U. maydis* sequenced so far, however, suggests that either the mutation was lost in natural populations and only maintained under laboratory conditions, or that it is only present in a so far unsampled population. ”

L289: “Its absence in any field isolates of *U. maydis* sequenced so far...” I had not noticed before that this is only found in the lab. To me, this changes substantially the take home message: it is in fact very plausible that this mutation would be selected against in the field; more generally, the fact that there is no indication of any fitness consequence of this insertion would lead me to take this as an example of a special mutational event. The fact that natural strains don’t have the intron and don’t have the nuclear insertion either also argues against the view that the intron was transferred to the nucleus => more likely a horizontal transfer.

A21: As mentioned, we have now rephrased the conclusion to reflect this possibility. In the revised manuscript, we have provided a more precise timeline when the HEG was acquired, and provide more support for the possibility that the acquisition was due to a HGT.

| Could such a transfer have happened in the lab? Are the two species kept in close contact?

A22: This is extremely unlikely. The lab strains originate from field isolates (as stated in the material and methods) and have never been crossed in the lab since they received their reference name.

Additional requirements of the managing board :

Please ignore this message if you already took these requirements into consideration.

As indicated in the 'How does it work?' section and in the code of conduct, please make sure that:

-Data are available to readers, either in the text or through an open data repository such as Zenodo (free), Dryad (to pay) or some other institutional repository. Data must be reusable, thus metadata or accompanying text must carefully describe the data.

-Details on quantitative analyses (e.g., data treatment and statistical scripts in R, bioinformatic pipeline scripts, etc.) and details concerning simulations (scripts, codes) are available to readers in the text, as appendices, or through an open data repository, such as Zenodo, Dryad or some other institutional repository. The scripts or codes must be carefully described so that they can be reused.

-Details on experimental procedures are available to readers in the text or as appendices.

-Authors have no financial conflict of interest relating to the article. The article must contain a "Conflict of interest disclosure" paragraph before the reference section containing this sentence: "The authors of this preprint declare that they have no financial conflict of interest with the content of this article." If appropriate, this disclosure may be completed by a sentence indicating that some of the authors are PCI recommenders: "XXX is one of the PCI XXX recommenders."

A "Conflict of interest disclosure" section was added at the end of the manuscript. All data and scripts necessary to reproduce the analyses are already available as both a public GitLab repository and **Supplementary File S1**.

Reviews

Reviewed by Yannick Wurm, 2019-11-25 13:32

Dutheil et al report a fascinating discovery whereby a mobile genetic element which once was an intron of a *cox1* gene seems to have excised itself and disrupted the function of a telomeric helicase gene in a strain of smut fungus. The data to support the argument are quite convincing. Intriguingly, this must be a very recent occurrence, because neither event has taken place in close related species or other strains. The manuscript is very clearly written, the methods are solid, and the story shows an elegant example of how a "jumping gene" can produce novel genetic variation in a eukaryote.

Minor comments:

The data to support the integration (genomes assembly + independent PCR) validation are quite strong. I would suggest additionally showing a plot of Illumina sequencing coverage as an additional track of Figure 1B. GC-bias for sequencing may affect this if the library was done based on PCR - an alternative may be to calculate sequencing coverage per gene and to do a PCA similar to Fig 1 but based on sequencing coverage. We would expect the focal gene 11064 to be well within the nuclear cluster but away from the mitochondrial cluster.

R1.A1: These are interesting suggestions. We would like to note, however, that the genome of *U. maydis* was obtained with a shotgun strategy and long-read sequencing technologies. We are unsure how a PCA could be conducted on coverage data only, given that there would be only a single value per gene? We further believe that the PCR independently and unambiguously confirms the genome assembly in this region.

The manuscript provides some data (and speculation) about the beneficial impact of having a truncated *UMAG_11065*. But little insight on the potential costs of the truncation having occurred. A bit more on this would help the discussion.

R1.A2: We now discuss the impact of the truncation in more details. The following sentence was added to the discussion: "The truncation likely did not have a strong negative impact, possibly because of the existence of multiple potentially functionally redundant paralogs of *UMAG_11065*, including on the same telomeric region of chromosome 9, with *UMAG_03394* being located 4 genes upstream (Table 1)."

Authors argue that 11064 is pseudogenizing based on it not being expressed and having several deleterious substitutions. Finding a high dN/dS for it would further support this argument.

R1.A3: We are very thankful for this suggestion. We have added a dN/dS analysis of *UMAG_11064*. Interestingly, we find evidence of a higher dN/dS ratio on this particular branch, as most substitutions appear to be non synonymous. We have amended our discussion to include these results, and present the pseudogenization as one possibility, the alternative being that the gene is expressed under different conditions than the ones tested. We also included a test of positive selection for the *UMAG_11065* genes, and found that it is evolving under purifying selection.

Analysis scripts are provided as supplementary - it would be beneficial to put these in a standard centralized, searchable and publicly accessible repository such as GitHub or Gitlab.

R1.A4: The scripts underlying the analyses are available on a GitLab repository. This is now stated explicitly in the manuscript.

| Some of the text within figures supposes that the reader understands shorthand such as "fw" and "rv" (eg
| Fig S2)- comprehension would be facilitated if spelling such things out. Similarly, prior to line 227, the
| meaning of the naming of the deletion strain should be specified unambiguously.

R1.A5: The significance of the “fw” and “rv” abbreviations is now made explicit in the legends of figures S1 and S2. The naming of the double deletion strain is also now unambiguously stated.

| Language:

| - l287: "singularly homogenous" - what does this mean?

R1.A6: We meant that the *U. maydis* genome display a very uniform base composition, in contrast to other organisms, in particular fungal pathogens. This made the introduced region stand out. We have deleted the sentence in the revised version in order to streamline the text.

- l223 a word missing

R1.A7: This is now corrected.

Reviewed by Jan Engelstaedter, 2019-12-18 06:40

In this paper the authors report the transfer of a homing endonuclease gene (HEG) from a mitochondrial intron onto the telomeric region of the nuclear genome of a plant pathogenic fungus. The authors use a wide range of approaches (bioinformatics, phylogenetic, DNA expression analyses, virulence assays) to elucidate this event, demonstrating that the HEG is partially degenerate, not expressed and presumably non-functional but truncated the reading frame of a putative helicase gene that is expressed during infection, with potential impacts on stress resistance. I think these results are very interesting and the methods appear sound. I enjoyed reading this paper and pondering its potential implications for our understanding of genome evolution more generally. I have three general comments and a number of specific comments which are listed below. As a disclaimer I should note that I am not an expert in either homing endonucleases or fungal genetics, so some of these comments may simply reflect my ignorance of the field.

General comments:

1) It hasn't become quite clear to me what the significance of the transferred element being an HEG embedded within a class 1 intron is. On line 26 in the abstract the authors refer to the transfer as a 'homing event' but how exactly do they think this came about? In the introduction, the authors explain how, for nuclear HEGs, the HE spreads by inducing a double-strand break in the homologous allele and then getting copied onto the other chromosome. But how does this work when the HEG is in the mitochondrial genome? And how can it effect its own transfer to the nucleus? Is the idea that both a DNA molecule containing the HEG and a HE protein itself somehow made their way from the mitochondria into the nucleus, the then HE cut the DNA in the telomeric region of chr9 which then finally gets repaired by the DNA fragment containing the HE? If so, is there any evidence for this?

R2.A1: The detailed mechanisms of transfer of HEG from one compartment to another are unknown. In principle, the scenario evoked by the reviewer seems the most sensible, but we are not aware of any evidence for this. It has also been proposed that the intron could be inserted at the RNA level by a reverse splicing mechanism, and then integrated into the genome (see Lambowitz (1989), Cell 56: 323-6). As far as we are aware of, only one other case of such transfer has been reported, in Yeast (Louis and Haber (1991), Curr. Genet 20:411-5).

Does the original telomeric region as found in other strains of *U. maydis* that don't contain the HEG feature the recognition site of the HE?

R2.A2: Unfortunately, we could not provide an answer to this question. Because the region is telomeric, we could not amplify the region where the HEG is missing.

2) Perhaps related to the previous, the paper does not seem to be as well embedded into the existing literature as it should be, specifically with respect to the horizontal transfer of introns/HEGs between species. For example, a quick search revealed that several papers, starting in the late 1990s, report horizontal transfer of a *cox1* intron between different flowering plants, with probably a fungal origin (e.g., Sanchez-Puerta et al. 2011 BMC Evol Biol 11:277 as well as papers cited therein and subsequent papers). It seems that this literature, which seems very relevant, is neglected here. It would also be good to discuss in more detail previously reported instances of transfers between plasmid and nuclear genomes.

R2.A3: We are thankful for this suggestion. We have added a new section about documented transfer of HEG in the discussion. In the revised version of the manuscript, we now consider the possibility of a HGT as likely.

3) The section comparing different *U. maydis* and *S. reilianum* strains on p.9 seems crucial for reconstructing the evolutionary history of the transfer event. I was very surprised therefore that in this section (and throughout the paper) no other species of *Ustilago* was looked at. *Ustilago* seems to be a large genus comprising several species infecting crops and with full genome sequence data available, so it seems that ascertaining which of these species (if any) also contain the UMAG_11064 HEG in either Cox1 or their nuclear genome should be feasible? Or is the taxonomy misleading in that *S. reilianum* is actually the closest relative?

R2.A4: The taxonomy of the *Ustilago* and *Sporisorium* genera are indeed polyphyletic (<https://www.ncbi.nlm.nih.gov/pubmed/27990337>). *Sporisorium reilianum* is so far the closest sequenced relative of *U. maydis*, despite a divergence time estimated to 20 My.

Specific comments:

I.16: I think this is wrong. As described correctly by the authors further down, HEs induce a double strand break and are then copied into the new position via homology-directed repair, so no excision is involved here. (The fact that they often occur within introns that are spliced out of RNA molecules seems a secondary phenomenon.)

R2.A5: Many HEG also display a “maturase” activity, which facilitates the excision of the intron. Some authors have argued that the autocatalytic activity of type-I intron is rather inefficient under physiological conditions, and the maturase guides the proper folding of the intron to enable its excision. This maturase activity is thought to have evolved as a secondary adaptation (See for instance Soddard 2005 for a review). To avoid ambiguity, we have rephrased the sentence as “Homing endonucleases (HE) are enzymes capable of cutting DNA at highly specific target sequences, the repair of the generated double-strand break resulting in the insertion of the HE-encoding gene (“homing” mechanism). Some HE genes are found within Group I introns, where they further facilitate their excision and turn them into selfish invasive elements.”

I.116-17: How well is the mitochondrial genome assembled? Could it be the the gene was missed because of a duplicate in the nucleus? If yes, perhaps the absence of the gene could also be confirmed by PCR?

R2.A6: The circular mitochondrial genome is completely assembled (NCBI entry: NC_008368.1). The absence of the gene was further confirmed by PCR in several *U. maydis* strains (Table 2).

I.120: I would suggest to change "must have occurred recently" to "is likely to have occurred relatively recently". Given that the gene is not expressed there won't be selection to adjust codon usage to the nuclear optimum, and it's also not clear how fast exactly GC content would be driven to nuclear levels.

R2.A7: The sentence was corrected as suggested.

I.137: Should it read "*S. reilianum*" here instead of "*A. bosporus*"? I thought I-AbIII-P is from *A. bosporus*, and Fig. 2 shows *S. reilianum*.

R2.A8: The sentence was indeed erroneous, and was now rephrased as “...we performed a codon alignment of the 5' region with the full intron sequence of *S. reilianum*, *T. indica* and *T. walkeri* as well as the sequence of I-AbIII-P from *A. bisporus* ...”

I.150-161 & Figure 3: I was really surprised to see that all introns in *cox1* contain an HEG, in both species. Is this typical for fungi or for the gene in general? Also, it is not clear to me why the HEG is "responsible for their correct excision". Is the HE directly involved in intron self-splicing? Aren't there type 1 introns without HEGs as well?

R2.A9: Group I introns do not necessarily include a HEG, but many do. The *cox1* gene is frequently invaded by HEG in fungi. The *cox1* gene of *A. bisporus*, for instance, holds the record of the longest *cox1* gene, with 18 HEG-containing introns (Jalalzadeh et al. 2015). It was shown that several HE proteins have a maturase activity that helps the excision of the intron, not directly catalysing the excision, but helping the proper conformation of the RNA molecule. We have updated the text and removed the “responsible for their correct excision” part.

| l.202: "very recently" is quite vague; it would be good if the authors could be more specific and provide an actual timeframe for this.

R2.A10: The sentence was rephrased as “an event that occurred after the domestication of maize and the spread of the associated pathogen, 10,000 to 6,000 years ago (Munkacsi et al. 2008).” In line of the new phylogeny presented on figure 2, we also discuss the possibility that the transfer might be more ancient than initially thought.

| l.187-205: I found this section a bit hard to digest, and Table 2 didn't help much. For example, in the table pluses indicate the presence of the 11064 ORF in some *U. maydays* and all *S. reilianum* strains, but not whether this ORF is part of the *Cox1* intron or located on chr9. Also, what does the lack of either pluses or minuses for the 11072 ORF in *S. reilianum* indicate? Perhaps this table could be replaced by a figure that shows the phylogenetic relationship between the strains and then graphically shows the configuration of both *Cox1* and the chr9 telomeric region for each strain?

R2.A11: The empty cells in Table 2 indicate that this particular configuration was not tested, this is now explicitly mentioned. The location of the gene can be deduced from the size of the regions. As suggested, we replaced this table by a new figure (Figure 6 in the revised manuscript). The original table is now part of Supplementary Figure S2.

| l.229-232: It seems that no details on the methods used for these assays are provided. Have these assays also been performed in triplicates?

R2.A12: These assay were performed according to previously published protocols, as stated in the methods section: “Stress assays were performed as described in (Krombach et al. 2018)”. They have been repeated at least three times with comparable results (now explicitly mentioned in the figure legend).

| Figure 1: Details need to be provided on the plot in panel A. It seems that some kind of principal component analysis was conducted? I guess this is what the authors refer to on l.298 as "within-group correspondence analysis" but as it stands this figure is very cryptic.

R2.A13: Figure 1A indeed represents the results of the within-group correspondance analysis, described in the methods section. The legend was updated to make this explicit.

| Figure 2: It wasn't clear to me what the shading shows exactly in this figure. The caption mentions "level of amino-acid conservation" and I guess is the result of some algorithm implemented in the Boxshade program, based on chemical similarities of amino acids?

R2.A14: This is correct, and now explicitly stated in the legend of the figure.

| Figure 3: Any reason the order of exons and introns is shown backwards in this figure? Also, the dashed lines on some of the arrows are very hard to see.

R2.A15: The *cox1* gene is encoded on the negative strand of the mitochondrial genome, hence its backwards display. A scale bar has been added to the figure showing the genomic coordinates. The contour of the arrows is now displayed with black lines, allowing to better see the dashed lines.

Figure 7: I think this evolutionary scenario is plausible. However, given that no strain carrying the HEG in both their mt and nuclear genome was identified, and that horizontal transfers of *cox1* introns have been previously reported (see general comment #2), perhaps a simpler scenario of a direct horizontal transfer of the element to chr9, without it ever being present in the *U. maydis* mt genome, would be more parsimonious?

R2.A16: We agree with the reviewer. In the revised manuscript, we have added a discussion about the occurrence of HEG HGT and present the two hypotheses (HGT and Figure 8) equally.

References:

Jalalzadeh B et al. 2015. The intraspecific variability of mitochondrial genes of *Agaricus bisporus* reveals an extensive group I intron mobility combined with low nucleotide substitution rates. *Curr. Genet.* 61:87–102. doi: 10.1007/s00294-014-0448-8.

Munkacsi AB, Stoxen S, May G. 2008. *Ustilago maydis* populations tracked maize through domestication and cultivation in the Americas. *Proc. Biol. Sci.* 275:1037–1046. doi: 10.1098/rspb.2007.1636.

Schweizer G et al. 2018. Positively Selected Effector Genes and Their Contribution to Virulence in the Smut Fungus *Sporisorium reilianum*. *Genome Biol. Evol.* 10:629–645. doi: 10.1093/gbe/evy023.