

Dear Recommender,

Thank you very much for the detailed review of our manuscript. We address below the points raised.

I thank the authors for their additional work that has clarified a number of points. Although the reviewers both consider the new version as sufficiently revised, I must confess I still have some concerns that I try to summarise below. In brief, with the clearer overall picture we now have in hands, this genomic change looks very much to me like a very recent mutational event, with little, if any, evolutionary implications. It is of interest to note that some HEG can occasionally insert somewhere in the nuclear genome, and I acknowledge that one had to investigate this insertion in more details. But the finding that it is most likely on its way to pseudogenisation arguably mitigates the significance of this observation. I would find it unfair to simply not recommend this study for this reason, but I suggest below some further revisions to try and clarify this point, and avoid giving excessive expectations to the reader. I apologise in advance if some of the issues raised below stem from my too limited knowledge of the system. If this is the case, may be some additional clarifications are needed to make the paper most accessible to a wide readership.

R: We are thankful the detailed suggestions, and have revised our manuscript accordingly. We do not want to stress the adaptive importance of this event, we agree that further analyses would be needed to definitively conclude on this matter. We hope that the revised manuscript no longer suggests any excessive expectation. We do think, however, that the reported mutational event is rather singular, and the fact that it did happen and the type of mutation, of sufficient interest in itself.

I will start with a summary of what I understood, so that the authors can correct me wherever there is a misunderstanding:

*One sequence (called UMAG_11064), present near a telomere of chromosome 9 of the fungal maize pathogen *Ustilago maydis*, has a mitochondrial origin, as evidenced by:*

** a strongly AT biased composition*

** strong similarity with mitochondrial sequences of close relatives (but not present in *Ustilago maydis* itself)*

** More specifically, this sequence is an insertion of a mitochondrial homing endonuclease. It is not clear how this insertion got there specifically.*

R: This is, indeed, what we wanted to say.

(I have one question on this point, that can hopefully be addressed in the paper. The question is general at first: how do HEGs integrate IN THE FIRST PLACE at their homing site? I do understand how HEGs can spread through gene conversion following cleavage; BUT how does the first insertion occur at the target site? In my understanding, HEGs do not themselves carry the capacity to insert; they are DNA cutting enzymes. Strangely, I did not find an answer to that seemingly basic question in the literature; did I misunderstand something? => The question then becomes more specific: how did this HEG insert at this site in the nuclear genome. The authors mentioned in their rebuttal letter that not much can be said on the insertion site; this should then be stated explicitly, saying that we have no idea how the HEG got there; obviously, the answer to the general question above will affect the answer to this specific one).

R: This is indeed a good question, for which we do not have a definitive answer. So far, HEG insertions are known to work via template-based DNA repair: the HE recognizes a cleavage site and initiates a DSB. The DSB is then repaired using the original HEG containing sequence as a

template, leading to the insertion of the HEG. When the insertion occurs at a non-orthologous site, it is not obvious which template is used. As far as we know, there are only speculations how this could occur. One is that the mRNA is retrotranscribed and the cDNA is used as a template. It has also been proposed that the RNA itself could be used as a template. Eventually, one could think that some mtDNA may have leaked into the cytoplasm during processes of mitochondrial fission and fusion events.

We have added the following sentences in the discussion, with the aim to clarify this aspect: “Several questions remain unanswered regarding the mechanisms of insertion into the nuclear genome, providing it happened as a homing event. For the event to happen, a template sequence containing the HEG is required for repairing the break initiated by the HE. This template must have, therefore, “leaked” from the mitochondrion. The recognition motif of the original *UMAG_11064* HEG is unknown, and the very short flanking regions surrounding the insertion site do not allow any comparison with known motifs, preventing further conclusions to be made regarding the nature of the insertion event of *UMAG_11064*.”

The gene is no longer acting as an HEG (anyway, it cannot be, since it is not inserted in a homing site; right?). It has lost the HEG active sites and its start codon. It does have a potential start codon, but it is not transcribed. The dN/dS is hard to estimate because the branch is very short, but it may be high. Most likely, it is on its way to pseudogenisation.

R: we do not know whether the gene is inserted at a homing site or not. We report several evidence that the HEG is pseudogenized, in particular owing to a deletion in the active site. There are, therefore, three possible scenarios: 1) the HE was active at the time of insertion and the insertion happened as a homing event. The HE then got inactivated (suggesting that the insertion is not too recent for mutations to have occurred). This is the scenario that we find the most parsimonious, but in the revised version of our manuscript we now mention two alternative scenarios where 2) the HE was active but the HEG was inserted by an unknown mechanism, different from a homing event and 3) the HEG was already degenerated when inserted, and the insertion occurred by an unknown mechanism.

In the discussion, we added the following paragraph:

“The insertion of the HEG in the nuclear genome poses the question of the underlying mechanisms, independently of the origin of the HEG. First, the HEG could be encoding a fully functional HE and the [HEG]_{nuc} allele contained a HE recognition sequence. Under this scenario, the insertion event was a homing event and the inactivation of the HEG occurred after the insertion. Therefore, several generations must have passed since the insertion event in order for the inactivating mutations to occur. Alternatively, the [HEG]_{nuc} allele might not have contained a recognition sequence and the insertion of the HEG occurred by an unknown mechanism. Lastly, a possibility is that the inserted sequence already encoded an inactivated HE, which was then inserted by an unknown mechanism. In this latter case, the insertion could have occurred very recently, possibly a few generations in the past.”

*The integration of this sequence correlates with the possible truncation of a neighbouring gene (*UMAG11065*), which includes many paralogs in the genome. Some paralogs are coding for fully functional helicase enzymes. Maybe *UMAG11065* was truncated upon the insertion of *UMAG11064*, maybe it was truncated before (but the later hypothesis is not considered by the authors?). Now *UMAG11065* is expressed a little bit.*

R: Here, we would like to amend the above statements: the sequence alignment clearly shows that the sequence starting immediately after the end of *UMAG_11065* is highly similar to other (longer) HEG genes. We therefore think that the insertion of the HEG is the most likely cause of the truncation. The level of expression of *UMAG_11065* is significant, in particular, it is similar to that of other helicase genes.

Experimental knockout of UMAG11064 has no phenotypic consequence; the knockout of UMAG11065 has some effects in some particular experimental conditions, that may be related to the loss of its activity; (which does not mean this activity is a function maintained by selection). In brief, there is no evidence that UMAG_11065 is functional either.

R: we would like to stress that we only perform a double mutant of the two genes simultaneously. Because we do not see any significant expression level of *UMAG_11064*, we conclude that the observed phenotype is due to *UMAG_11065*.

*The UMAG11064 is only found in this particular strain. This suggests it could be a very recent mutational event. Based on the phylogeny of UMAG11064 and its homologues, the authors suggest the insertion may have predated the split between *Sporisorium reilianum* and *Sporisorium scitamineum*, but in fact, (1) the (*S. reilianum*, *S. scitamineum*) node is poorly supported and (2) it may as well be that *UMAG_11064* came very recently from another source, not sampled here (from a strain that that would indeed branch at that position in the tree). The fact that it is not present in other natural populations argues against an ancient insertion, and against a new functional role.*

R: We agree that this phylogenetic argument was rather weakly supported by the data and we have removed the corresponding paragraph from the discussion to avoid confusion.

Could the authors comment on this summary to help us reach a decision? In the additional comments below, I suggest some modifications on the basis of my understanding of the story.

R: we are thankful for the suggestions to improve the clarity of our manuscript. We agree that the event that we report here is an “accident”, in the same way that all mutations are. The mutation reported here, however, is a rather unusual one. Because we have here a snapshot of evolution taken very quickly after the mutation happened (which we now emphasize in the conclusions), we think that the study of this locus (for instance by sampling more wild strains) has the potential to shed light on the mechanisms of HEG insertion and transfer between mitochondria and nuclei. As we conclude, we propose that such mutational events potentially have evolutionary consequences as they can lead to rather significant modifications of the local gene architecture.

Title

The term “transfer” tends to suggest that the “mitochondrial selfish element” has retained its capacity to act as a such. I would suggest the following: “Accidental insertion of a mitochondrial selfish element to the nuclear genome and its consequences”

R: We changed the title to “The insertion of a mitochondrial selfish element into the nuclear genome and its consequences”. We agree that the term “insertion” is more appropriate than “transfer”. We would prefer, however, to drop the term “accidental”, as all mutational events are accidents.

Abstract

Some revisions of the abstract would reduce risks of misunderstandings

L18: “Some HE genes are found within Group I introns, where they further facilitate their excision”. Without further explanations, one wonders here in what sense do the HEGs facilitate excision of the introns. The sentence would also suggest that HEGs alone are not selfish invasive

elements, although they are. Being in an intron just reduces harmful effects. Overall, I would thus suggest to remove this sentence from the abstract.

R: This sentence has been deleted.

L22: “HE that integrated into”; I suggest adding again the adverb “accidentally”

R: Done.

L24: “or a horizontal transfer”. In some sense, transfer from mtDNA to the nucleus is already a horizontal transfer, even if this occurs from your own mitochondria. I would thus suggest “or a horizontal transfer from a different species”

R: Done.

L25: “acquired a new start codon,” => in fact, the start codon is just a remnant of a methionine codon that happened to be there; the current phrasing would suggest a new start codon was selected for. Something like this may be more appropriate: “The telomeric HE underwent mutations in its active site and lost its original start codon. A potential other start codon was retained downstream, but we did not detect significant transcription of the newly created open reading frame, suggesting the inserted is not functional.”

R: The abstract was corrected as suggested.

L29: the last two sentences starting with “This unusual homing event” are problematic in my view. “Creation of two new genes seems inappropriate. The ‘homing’ term can also be questioned because it should be restricted, in my understanding, to the conversion event following a DSB at the homing site. Here we rather have an accidental insertion event, through an unknown mechanism. Instead of those two sentences, I would suggest a more modest ending of the abstract. First, the absence of the insertion in other strains should be mentioned, as an indication that this event is likely recent and, in any case, not fixed. The abstract could end by stating that such mutations may be important in some cases, although this is not apparently the case here. Something like: “The absence of this insertion in other field isolates suggests it likely represents a recent mutational event, and brings no support for a putative adaptive significance. These findings indicate that mitochondrial HEGs can occasionally insert in the nuclear genome, a particular mutational event that may constitute a source of adaptation, although we found not support for such evolutionary implications in that case.”

R: While we agree that it cannot be proven, we do not see any reason why a homing event should be excluded. Since the detailed mechanisms of insertions are beyond what we can study with the data at hand, we prefer to remain open regarding this question and do not favor any scenario. We now mention the absence of the insertion in other strains. We note, however, that if the mutation is indeed very recent, its lack of fixation does not argue against a putative adaptive role. We have rewritten the end of the abstract as:

*“The truncated helicase is expressed during infection of the host, together with other homologous telomeric helicases. This unusual mutational event altered two genes: the integrated HE gene subsequently lost its homing activity, while its insertion created a truncated version of an existing gene, possibly altering its function. As the insertion is absent in other field isolates, suggesting that it is recent, the *U. maydis* 521 reference strain offers a snapshot of this singular mutational event.”*

Main text

L58: “using the HEG itself as a template”; if I understood correctly, what is used as a template is the homologous chromosome, which happens to carry the HEG. I find it slightly unclear to state that the HEG is used as a template: it just happens to be part of the template.

R: This sentence was rephrased as “The resulting double-strand break is subsequently repaired by recombination using the homologous sequence containing the HEG itself as a template, resulting in its insertion in the target location”

L119: “...which suggests that UMAG_11064 is an authentic nuclear gene.” It seems to me that the nuclear location of the gene is already well established at that stage by the genome assembly + PCR control. So, to me, this new piece of data (the fact that there is no copy of this gene in the mitochondria) should rather be seen as an argument that the gene was either lost from mitochondria following its transfer, or acquired from a different species.

R: We have moved this sentence to the section where occurrence in other species is discussed.

L136: “The amino-acid sequence of UMAG_11064 matches the N-terminal...” may be an indication of the level of identity at the protein level would be useful here.

R: the level of nucleotide identity was added (54%).

L163: “To further assess the possibility that the UMAG11064 gene is evolving under positive selection, ...”. I would suggest a rephrasing with something like: “the possibly high dN/dS seen in the UMAG11064 branch could be explained both by relaxed purifying selection or positive selection. To assess the validity of these two explanations...”

R: We rewrote this sentence as “To further assess whether the high dN/dS ratio measured in the UMAG_11064 gene could be explained both by relaxed purifying selection or positive selection, we fitted a branch-site model allowing specifically for sites in the UMAG_11064 gene to evolve under positive selection (foreground branch), which we contrasted with a null model where all sites evolve under purifying selection or neutral evolution.”

L189: “the cox1 gene seems to be a hotspot of Group I introns in smut fungi”; to make the idea more explicit, I would suggest: “... of Group I introns encoding HEG in smut fungi”

R: We rephrased this sentence as “...the cox1 gene seems to be a hotspot of HEG-encoding introns in smut fungi.”

L190: “Lastly, intron 1 in S. reilianum was not detected in U. maydis.” The next bit of this paragraph deals, if I understood well, with the UMAG_11064 gene alignment. But strangely the paragraph starts with this sentence about the absence of this gene in U. maydis (presumably in the CO1 gene?). It seems to me this information should not be presented here, or not in this way.

R: this paragraph is about the closest homolog of UMAG_11064, the intron 1 of the cox1 gene in *S. reilianum*. We first recall that this intron has no homolog in the mitochondrial genome of *U. maydis*, and then discuss that the HEG in this intron is most likely no longer functional. We have rephrased this paragraph with the hope to make it clearer:

“Lastly, no homolog of intron 1 in *S. reilianum* was detected in the mitochondrial genome of *U. maydis*. A closer inspection showed that the ORF could be aligned with related HEGs in other fungi (Figure 3). This alignment revealed an insertion of four amino-acids, a deletion of the first glycine residue in the active site plus several frameshifts at the beginning of the gene, which suggests that this gene has been altered and might not encode a functional HE any longer.”

L199: "...detected 13 homologous sequences..."; may be use "paralogous" here instead of "homologous" to make it clear that you are looking homologous sequences in the same genome?

R: This was corrected as suggested.

L202: Could the authors state why they go for phylogenetic reconstruction here? What do they want to know with this analysis? I can see why it was required for the inserted gene, but it is not see clear for this gene. If it is only part of the dN/dS analysis, it should not pbe presented and discussed in details, and should not make a figure.

R: The goals of the phylogenetic analysis are two-fold: first, we aimed at understanding the evolution of this gene family. The main result is that most duplications seem to have happened in the *U. maydis* lineage. The second goal was to identify the closest "relative" of the *UMAG_11065* gene, in order to get insights about the gene structure before the insertion of the HEG. We find that helicase genes vary extensively, but that the closest relative, which shows a high sequence similarity, has a much longer ORF. This hints at a longer pre-*UMAG_11065*.

L208: "... but rather to its truncation": but this assumes *UMAG04486* corresponds to the ancestral form of *UMAG11065*. Are there good reasons to think this is the case?

R: This is the conclusion we reached from the phylogenetic analysis, where *UMAG_04486* was found to be the closest relative to *UMAG_11065*. The next closest relative is *UMAG_06506*, which is also much longer, suggesting that the ancestor of *UMAG_11065* was indeed longer.

L215: "Our results suggested, however, that the *UMAG11065* gene evolved under purifying selection (dN/dS ratio equal to 0.342)"; but it seems to me you do not know if this selection regime still holds after the insertion of *UMAG11064*. The question is: is *UMAG_11065* still functional and subject to purifying selection following this truncation. I suspect it is difficult to answer this question with the data in hands.

R: This is indeed a very good point. We have amended the text as: "We inferred a dN/dS ratio equal to 0.342, which suggests that the *UMAG_11065* gene evolved mostly under purifying selection since divergence from the *UMAG_04486* gene. The insertion of *UMAG_11064*, therefore, was not followed by positive selection, or was too recent for sufficient positively selected substitutions to occur."

A remark on figures and supplementary figures legends: it is currently very difficult to follow which figure is which, since the figure numbers are not in the pdf.

R: The figures are now directly inserted into the manuscript.

*Figure 6: It is slightly disturbing not to see *UMAG11065* on these pictures, considering it the closest to *UMAG11064*?*

R: The reason is that we only tested here for the presence of *UMAG_11064* using PCR, using *UMAG_11072* as a positive amplification control. We do not have population data permitting to assess the presence/absence pattern of other genes in the region. In order to make it clearer, the legend was edited as "Presence of the *UMAG_11064* gene and structure of the *cox1* gene in several *U. maydis* and *S. reilianum* strains, as assessed by PCR, together with their phylogeny. The *UMAG_11072* gene, located 90 kb downstream the *UMAG_11064* gene on chromosome 9, was used as a positive control."

L233: “an ancestor of the two strains 518 and 521”; shouldn’t that be “an ancestor of the three strains 518, 521 and SG200” ? But in any case, all these are coming from a single spore in the lab? I think this should be emphasised: only one occurrence of this insertion was found, likely very recent.

R: These strains indeed come from a single diploid teliospore, sampled from a natural population. So in theory, the insertion event could have happened in the individual that generated this teliospore. Unless we assumed that the transferred sequence was already inactivated, such a very recent insertion (1 generation ago) is incompatible with the accumulation of mutation events inactivating the HEG after its insertion. Again, we would like to recall that we do not have a sample of the original population, only one spore. Therefore, we cannot exclude that the mutation was segregating in this population.

L318: “potentially had non-neutral effects”; but this is highly hypothetical; the neutral explanation proposed for the other story by Louis and Haber seems to correspond rather well to the current one.

R: this is indeed hypothetical. The difference between this occurrence and the one reported by Louis and Haber is that, here, the insertion occurred within an expressed protein coding gene.

L321: “However, an alternative start codon was detected,..”; yes, but the gene is not expressed. This would tend to suggest it is not functional.

R: We rephrased this section with the goal to make it more explicit: “However, a putative alternative start codon was detected, downstream the active site, followed by an uninterrupted peptide sequence containing the helix-turn-helix binding domain of the original HE. Furthermore, we could not detect any significant level of expression of the *UMAG_11064* gene in various laboratory conditions. Comparative sequence analysis further suggests that *UMAG_11064* is evolving under relaxed purifying selection, indicating that it might be undergoing pseudogenization. These results, therefore, suggest that the *UMAG_11064* gene is not functional.”

L333 and following ones: in my view, arguing for an adaptive role is too speculative based on the data in hands.

R: In this section, we are testing whether the truncation could have adaptive consequences. We report results of reverse genetics experiments that we conducted, and remain cautious in our conclusions, which we indeed present as speculations. In the revised version of the manuscript, we further added the following sentences at the end of the discussion: “Furthermore, a possibility remains that the observed phenotype of *UMAG_11065* is ancestral and not due to the truncation itself, which could be neutral. In order to elucidate the putative adaptive role of the truncation of *UMAG_11065*, knowledge of the ancestral, non-truncated *UMAG_11065* allele is needed, as well as its distribution in natural populations.”

Reviews

Reviewed by Jan Engelstaedter, 2020-03-24 06:05

I think the authors have done an excellent job addressing the comments made by myself and the other reviewers. In particular, the new figures 2 and 6 are very helpful and I appreciate that the authors have clarified many aspects of their work and have taken on board my suggestion of an alternative evolutionary scenario for HEG transfer. The article raises many fascinating questions

and I hope it will motivate more studies elucidating the evolutionary genetics of mitochondrial HEGs.

Reviewed by Yannick Wurm, 2020-04-01 08:35

The concerns have largely been addressed - the paper looks great. As mentioned previously, a density plot of genome sequencing coverage across the insertion, or comparing the insertion to other parts of the genome would add further support (not PCA indeed). As also mentioned, it would strengthen the story but not change it.

R: We are thankful to the reviewers for their insights and suggestions on how to improve our manuscript. We recall that the *U. maydis* genome (published in 2006) was obtained using Sanger sequencing using a combination of shotgun and mapping-based strategies. Unfortunately, we do not have access to detailed coverage information along the original genome.