



**HAL**  
open science

## Some evolutionary insights into an accidental homing endonuclease passage from mitochondria to the nucleus

Sylvain Charlat

► **To cite this version:**

Sylvain Charlat. Some evolutionary insights into an accidental homing endonuclease passage from mitochondria to the nucleus. 2020, pp.100101. 10.24072/pci.evolbiol.100101 . hal-02612297

**HAL Id: hal-02612297**

**<https://hal.science/hal-02612297>**

Submitted on 19 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Some evolutionary insights into an accidental homing endonuclease passage from mitochondria to the nucleus

Sylvain Charlat based on reviews by Jan Engelstaedter and Yannick Wurm

### Open Access

A recommendation of:

Julien Y. Dutheil, Karin Münch, Klaas Schotanus, Eva H.

Stukenbrock and Regine Kahmann. **The insertion of a**

**mitochondrial selfish element into the nuclear genome and its**

**consequences (2020), *bioRxiv*, 787044, ver. 4 peer-reviewed and recommended by Peer Community in Evolutionary Biology.**

[10.1101/787044](https://doi.org/10.1101/787044)

*Submitted: 30 September 2019, Recommended: 15 May 2020*

**Cite this recommendation as:**

Sylvain Charlat (2020) Some evolutionary insights into an accidental homing endonuclease passage from mitochondria to the nucleus. *Peer Community in Evolutionary Biology*, 100101.

[10.24072/pci.evolbiol.100101](https://doi.org/10.24072/pci.evolbiol.100101)

Not all genetic elements composing genomes are there for the benefit of their carrier. Many have no consequences on fitness, or too mild ones to be eliminated by selection, and thus stem from neutral processes. Many others are indeed the product of selection, but one acting at a different level, increasing the fitness of some elements of the genome only, at the expense of the

Published: 17 May 2019

Copyright: This work is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/4.0/>

“organism” as a whole. These can be called selfish genetic elements, and come into a wide variety of flavours [1], illustrating many possible means to cheat with “fair” reproductive processes such as meiosis, and thus get overrepresented in the offspring of their hosts. Producing copies of itself through transposition is one such strategy; a very successful one indeed, explaining a large part of the genomic content of many organisms. Killing non carrier gametes following meiosis in heterozygous carriers is another one. Less known and less common is the ability of some elements to turn heterozygous carriers into homozygous ones, that will thus transmit the selfish elements to all offspring instead of half. This is achieved by nucleic sequences encoding so-called “Homing endonucleases” (HEs). These proteins tend to induce double strand breaks of DNA specifically in regions homologous to their own insertion sites. The recombination machinery is such that the intact homologous region, that is, the one carrying the HE sequence, is then used as a template for the reparation of the break, resulting in the effective conversion of a non-carrier allele into a carrier allele. Such elements can also occur in the mitochondrial genomes of organisms where mitochondria are not strictly transmitted by one parent only, offering mitochondrial HEs some opportunities for “homing” into new non carrier genomes. This is the case in yeasts, where HEs were first reported [2,3]. In this new study, based on genomic experimental data from the fungal maize pathogen *Ustilago maydis*, Julien Dutheil and colleagues [4] document one possible evolutionary pathway for which little evidence existed before: the passage of a mitochondrial HE into the nuclear genome. The GC content of this region leaves little doubt on its mitochondrial origin, and homologs can indeed be found in the mitochondrial genomes of close relatives. Strangely enough, *U. maydis* itself does not appear to carry this selfish element in its own mitochondria, suggesting it may have been acquired from a different species, or be subject to a sufficiently rapid turnover to have been recently lost. Many elements of the story uncovered by this study remain mysterious. How, in the first place, was this HE gene inserted in a nuclear genomic region that shows no apparent homology with its original insertion site, making typical “homing” a not-so-likely explanation? This question may in fact be generalised to many HE systems: is the first insertion into a homing site always the product of a typical homing event, which implies the presence of an

homologous template DNA fragment, or can HE genes insert through other means? But then, why specifically in regions that would be targeted by the nuclease they encode? What is the evolutionary fate of this newly inserted element? The new gene may well be on its way to pseudogenisation, as suggested by the truncation of its upper part, precluding its functioning as a HE, and the lack of evidence of selective constraints through dN/dS analysis; but the mutation generated by the insertion event may have phenotypic implications, possibly through the partial truncation of another gene, encoding a helicase. How old is this insertion? The fact that it has accumulated some mutations makes a very recent event rather unlikely, but this insertion has been detected in only one isolate of *U. maydis*, suggesting it is not so frequent in natural populations. Whatever the answers to these open questions, that will hopefully be addressed by further work on this system, the present study has revealed that horizontal transmission enlarges the scope of possible evolutionary consequences of HE genes, that may move not only between mitochondrial genomes, but also occasionally into a nucleus.

## References

- [1] Burt, A., and Trivers, R. (2006). *Genes in Conflict: The Biology of Selfish Genetic Elements*. Belknap Press. [2] Coen, D., Deutch, J., Netter, P., Petrochillo, E., and Slonimski, P. (1970). Mitochondrial genetics. I. Methodology and phenomenology. *Symposia of the Society for Experimental Biology*, 24, 449-496. [3] Colleaux, L., D'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F., and Dujon, B. (1986). Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. *Cell*, 44, 521–533. doi: [10.1016/0092-8674\(86\)90262-X](https://doi.org/10.1016/0092-8674(86)90262-X) [4] Dutheil, J. Y., Münch, K., Schotanus, K., Stukenbrock, E. H., and Kahmann, R. (2020). The insertion of a mitochondrial selfish element into the nuclear genome and its consequences. *bioRxiv*, 787044, ver. 4 peer-reviewed and recommended by PCI Evolutionary Biology. doi: [10.1101/787044](https://doi.org/10.1101/787044)

## Revision round #2

2020-04-08

Decision on “The transfer of a mitochondrial selfish element to the nuclear genome and its consequence” by Dutheil et al ([doi.org/10.1101/787044](https://doi.org/10.1101/787044)).

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.

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.

(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).

- 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.
- 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. May be *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.
- 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.
- 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.

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.

## 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”

## 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.

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

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”

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.”

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 adaption, although we found not support for such evolutionary implications in that case.”

## 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.

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.

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.

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...”

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”

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.

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?

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.

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?

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.

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.

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

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.

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.

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

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

Preprint DOI: <https://doi.org/10.1101/787044>

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.

### Author's reply:

[Download author's reply \(PDF file\)](#)

## Revision round #1

*2019-12-19*

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?

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.

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?

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?

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

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

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

=> 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.

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.

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

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?

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?

L197: “The UMAG\_11072 gene...” what does this information tell? Is this a positive control? Of what exactly?

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

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?

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?

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

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

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.

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?

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.

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

**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."

*Preprint DOI:* <https://doi.org/10.1101/787044>

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.

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.

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.

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.

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.

Language: - l287: "singularly homogenous" - what does this mean? - l223 a word missing

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? 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?

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.

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?

Specific comments:

I.16: I think this is wrong. As described correctly by the authors further down, HEGs 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.)

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?

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.

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

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?

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

I.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?

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

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 I.298 as "within-group correspondence analysis" but as it stands this figure is very cryptic.

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?

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.

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?

### **Author's reply:**

[Download author's reply \(PDF file\)](#)