Dear PCI recommender,

please find below the responses to the reviewer's comments. Regarding the main issue raised by reviewer one (the interpretation of gene trees with or without virus outgroups), we followed your own suggestion and replaced our verbal argumentation by a quantitative one. We measured the mean divergence of LbV with wasps species and compared this to the divergence among wasp species for all 13 phylogenies. This relative divergence was the same for both groups (phylogenies having other viral outgroups versus phylogenies without such viral outgroups), suggesting that they do indeed have the same evolutionary history. We believe that this new argument strengthens our initial conclusion based on (among others) verbal arguments. We hope that this new version will fits your expectation.

Sincerely

Julien Varaldi

The authors have made several revisions based on my comments. The authors agree that a viral origin does not discount the potential status as an organelle; however, this agreement is not reflected within the paper. An example of this is in the following paragraph:

Because 323 the proteins wrapped within the VLPs have a eukaryotic origin and because 324 neither viral transcripts nor viral proteins had been identi_ed from venom 325 gland analysis, it has been claimed that VLPs do not have a viral origin [56], 326 and thus other denomination has been proposed in lieu of VLP [29]. On the 327 contrary, our data strongly suggest that the VLPs found in Leptopilina do 328 have a viral origin and derive from a massive endogenization event involving 329 a virus related to an ancestor of the behaviour manipulating virus LbFV(Fig 330 2B).

This sentence is puzzling given the fact that the authors contend in their response document: “Nowadays, VLPs are eukaryotic structures (organelles) even if some of the key genes involved in their production derive from virus genes.”

Taken together, these statements are confusing. Origin of VLPs is discussed without a clear description of our current knowledge of VLPs from various Leptopilina species. Have VLPs been described from the Leptopilina species studied here? What proteins are present in VLPs and do the results in this paper have anything in common with any described VLP proteins? Even a negative result would be worth stating and discussing.

Similarly, in line 401, the authors write: “All together, our data show that VLP production is possible thanks to the domestication of 13 virally-derived genes, captured from an ancestor of LbFV.”

We have modified the sentence as : “All together, our data *strongly suggest* that VLP production is possible thanks to the domestication of 13 virally-derived genes, captured from an ancestor of LbFV”

Thus, it is clear, the authors are convinced of their idea, even though they have stepped back (superficially) by changing the title of their paper.

In this reviewer's view, the authors have not shown that the 13 virally-derived genes in the Leptopilina genomes studied are involved in VLP production. Their data demonstrate the existence of these genes in wasp genomes and their spatial and temporal expression in venom gland extracts.

As such, these results do not link LbFV genes to venom production, VLP production, or venom/VLP function. It is commendable that the team tried RNA interference experiments. However, in the absence of such results, it is advisable to wait to get the necessary evidence that will shed light on the function(s) of these interesting wasp genes/proteins. Only two sequences have any sequence similarities, but no further experimental data on these or any of the wasp LbFV proteins is available. Thus, there is a significant gap between evidence and interpretation.

Ref 56 has interesting ideas that differ from the ones proposed by the authors. It is worth stating them clearly with underlying evidence. Instead of taking an oppositional view (as in lines 320-330), a more balanced view of pertinent ideas would improve this manuscript and benefit the quality of discussions in this growing field.

We do think that our data strongly suggest that VLPs do have a viral origin. That's why we propose a different scenario as the one proposed in Heavner et al. (2017) or in Poirié et al (2014) for instance. However, to give a more balanced view on this topic, we included the main arguments proposed by the authors favoring this non-viral origin hypothesis (in the discussion). But again, although we believe that our data strongly suggest that VLPs do have a viral origin, this is not contradictory to the eukaryotic origin of the proteins that are inside the VLPs (the virulence proteins). The viral genes are "only" responsible for the production of the membrane surrounding virulence proteins thus favoring their delivery to *Drosophila* immune cells.

Other comments:

(1) I do not understand the reluctance to show alignment of wasp ORFs with viral ORFs. This information would be informative in understanding, for example, where the primers (for expression and copy number studies) bind.

We included the requested alignments as supplementary figures (S2-S14).

(2) If you have done experiments to check the copy number of shake and actin (used as controls for copy number), provide your evaluation of their copy number in the supplement. In the same context, provide appropriate citations showing that these genes are single copy genes in related genomes.

We added some details in the material an methods to justify the conclusion that they are single copy genes:

Shake and actin genes were chosen as single copy genes. This was checked by looking at the blast results using each primer set (a single 100\% match was observed for both pairs of primers). Accordingly, a single band of the expected size was observed on a gel and the expected sequence was obtained after Sanger-sequencing for both loci.

(3) Regarding the eukaryotic origin of LbGAP, the reference cited is incorrect. Ref. 17 is the correct reference for this. Please make sure all statements are correctly corroborated with appropriate references.

Thank you for that. We corrected the error.

(4) The species for the Ganaspis wasps is changed in this revision. Identification of these wasps is quite difficult. So it is important to say how verification of species used was carried out. Please do this for all species. What criteria were used? Looks like G.x (line 525)—is carried over from the previous version of the paper? Please correct this.

We now mention the origin of the strains and corrected the Gx error. The identity of the Ganaspis strain has been determined by the laboratory that kindly sent us the line. The Leptopilina strains have been captured in France by our group and we used classical criteria to distinguish the two species (L. boulardi/ L. heterotoma). This is quite easy since they are the sole *Leptopilina* species in this geographical area. We do not think that this detail is necessary in the manuscript but please let us know if you think it is.

(5) Figure S1 legend: last sentence requires a full stop at the end. Dr. Shubha Govind's name is misspelled in the acknowledgements. Please review the paper for similar errors.

Thank you for that. We apologize again and corrected the error.

Reviewer 1

Indeed I do agree the phylogenies of the 6 loci in question are not in any way inconsistent with the authors' hypothesis. However, they are not consistent with it either. Again, with the lack of an outgroup they just do not provide evidence for a "horizontal transfer from an ancestor of the virus LbFV". I understand using midpoint rooting, however this just indicates that LbFV is very distant to the genome-insertion-event copies, and that these are closely-related. What thie ephylogenies do provide evidence for, is for a putative single origin, all these being phylogenetically very closely related. So, I suggest rephrasing as such (or similar):
"The evolutionary history of 7 genes is consistent with a horizontal transfer from an ancestor of the LbFV virus (or a virus closely related to this ancestor) to Leptopilina species (Figure 3B-D[etc..]). For 6 genes (ORFs 58, 78, 92, 60, 68, 85, 96), no homologs were available in public databases apart from their homologs in LbFV. However, the three copies from wasp genomes always formed a highly supported monophyletic clade."
I would argue the topology within the monophyletic clade is not very stable, having 3 with Lb as sister to (Lc + Lh), 2 with Lh sister to the other two, and one with Lc sister to the other two. Authors could also reorder the phylogeny panels so as to group the ones that had no other homologues but LbFV.

The major comment concerns the interpretation of the 6 "unrooted" phylogenies as opposed to the "rooted" with other viruses. To add another quantitative argument to this debate, we calculated the mean divergence of LbFV-Leptopilina relative to the mean divergence among Leptopilina species for all 13 loci. We then compared this index between groups (the 7 "rooted" versus the 6 "unrooted"). There was no difference, further suggesting that all 13 genes have the same evolutionary history. We hope that this quantitative argument, which complements other arguments (based on the similar

topologies of the phylogenies, and on the co-occurrence on the same scaffolds), will be enough to convince the reviewer. We modified this part of the text accordingly.

I understand, I assumed that is why you did not performed the searches. But I would recommend to include it in the text as a goal of the article. If not, It might leave the reader with the impression that the authors only indeed found LbFV hits and not from other viral lineages (from Nudiviridae or others), especially when some headers state things like this: "Leptopilina species captured 13 viral genes". More appropriately it should read "Leptopilina species captured 13 viral genes from an LbFV-like virus"
My I suggest including a phrase as such:
"In order to identify putative events of integration from an LbFV-like virus to the wasp genomes, we blasted the 108 proteins encoded by the behaviour-manipulating virus that infects L. boulardi (LbFV) against the Leptopilina and Ganaspis genomes (tblastn)."

We modified the text according to the suggestion (with a slight modification to account for the fact that we were looking both for endogenization into wasp genomes and gene capture by the virus).

I would argue the topologies are generally congruent (since it is only congruent in the well-supported clades). The well-spoorted clades (going from the authors' definition of >=0.95) I see in phylogeny A are L. australis+L. clavipes, L. orientalis+(L. freyae+L .boulardi [this clade is missing support value]) and L. guineaensis + (L.victoriae+L. heterotoma[not well-supported]). The ones I see in phylogeny B are L. freyae + L. boulardi [missing support value, I am assuming it is well supported; bipartition not well supported in phylogeny A], L. clavipes + L. clavipes [same species], L. guineaensis + (L. clavipes + L. freyae +L. boulardi), (L.victoriae+L. heterotoma [not well-supported in phylogeny A]). So, I would say the topologies are generally congruent.

We agree on that. Just a clarification: the clades without support values have low aLRT (<0.7) and thus are note reliable.

Indeed, such a concatenated protein phylogeny is possibly a very good (if not the best) approximation of the species-tree, but "for sure the species tree", not. There are many approximations to a true "species-tree" (using the lax definition of any tree where several genes [protein-coding and not] are used for phylogenetic inference). The authors themselves are using yet another definition of "species-tree" in their article in figure S4, were they, in my opinion, wrongfully use the term species-tree for a phylogeny based on ITS2 sequences (single locus). So, I suggest to correct the naming of species-tree for the phylogeny based only on ITS2 and to use " a species tree was approximated".

We agree an modified the text accordingly (line 222).

The thing with coverage is that, while it might give you a sense (and I mean a sense in the subjective opinion sense) about having "all" your genome sequenced, it tells you more about the quality of the base calling than of the completeness of your assembly. Completeness of your assembly (with the technology you chose for sequencing) is best estimated with k-mer analysis checking for saturation (therefore you know no new k-mers are discovered with further sequencing using the same technology) and secondarily by a BUSCO result that tells you you have all conserved genes. To my knowledge, there is no study that analyses across several eukaryotic genomes and correlates a certain minimum coverage with "genome completeness", or "sufficiency to get the whole gene set". So, I would abstain of making such a definite statement as "which is most likely sufficient to get the whole gene set".

We added a reference of such a work on fish genomes. They tested the relationship between coverage and “completeness” measured as the quantity of BUSCO genes found and observed that above 15x they were done for gene content. We also have similar data on 35 hymenoptera genomes and found similar trend with slighty higher value but still below the coverage obtained in the present paper.

M. Malmstrøm, M. Matschiner, O. K. Tørresen, K. S. Jakobsen, and S. Jentoft. Whole genome sequencing data and de novo draft assemblies for 66 teleost species. Scientific Data, 4:160132, Jan 2017.