

Comments to the authors

The work presents a novel example of viral genome integration into the genome of parasitoid wasps (in this case from the *Leptopilina* genus). Contrary to previous publications, the authors were able to identify an extant dSDNA that still infects *Leptopilina boulandi* (LbFV), previously published by the authors (10.1093/gbe/evw277), as a close relative to the original viral donor. Through the use of phylogenetics and comparative genomics, they are able to provide strong evidence for a single event of integration of the viral sequences found in the analysed wasp genomes. Additionally, the authors explore the development of the venom glands and the production of the Viral-like particles (VLPs) and the expression and amplification of the virally-derived genes (VDGs) in *L. boulandi*. Finally, extrapolating from the behavioural changes in *L. boulandi*'s egg-laying (preference to laying eggs in already parasitised larvae), the authors propose that this is a likely mechanism that could have been used by the virus to spread to other wasp lineages and could have been instrumental in the birth of the symbiotic association. I believe the article is generally OK-written (needing some restructuring and clarifications), well presented, and greatly contributes to the knowledge about how these symbiotic associations have impacted and contributed to the host biology. Most experiments and interpretations are well presented and discussed. The authors really did a well-rounded job in investigating this viral HGT to the wasp host. It deserves to be considered for publica., sorry, recommendation after some corrections/modifications/clarifications.

Major comments

page 5 line 184: The authors state that “*The evolutionary history of the thirteen genes is consistent with an horizontal transfer from an ancestor of the virus LbFV (or a virus closely related to this ancestor) to Leptopilina species (Figure 3)*”. However, Figure 3 does not precisely show that. The only phylogenies that are “consistent with a horizontal transfer from an ancestor of the virus LbFV” are those of ORF58, ORF60, ORF68, ORF78, ORF85, ORF92, and ORF96 (7 genes). For the rest of the genes, the lack of outgroups (I'm sure the authors did not found any suitable ones in the databases) does not allow the identification of the VDGs as monophyletic. The choice of LbFV as the outgroup of VDGs in those trees forces the rest of the genes to be monophyletic. Thus, if no suitable outgroup(s) to LbFV + VDGs is found, the hypothesis that the authors tried to test (VDGs monophyletic and/or sister to LbFV) is not testable. So please correct this in the results section and omit these phylogenies from the figure.

page 4 line 133: Regarding the blast results, please provide full settings (e.g. evaluate threshold, percent identity) and version in the methods section. Also, please specify what you mean with “highly significant” (provide evaluate or relative bitscore vs self hit or other metric). Additionally, is it correct that you only used LbFV to do the BLAST searches. How do you make sure no proteins from Nudiviridae or others that have previously been identified in other wasps are not present?

Figure 2: I find this figure is a bit confusing. First, the TEM images shown beside the *Leptopilina* seem to all be the same. I find this a bit deceiving since it gives the impression those TEMs are from each one of the species. From reading the article they are all from *L. boulandi*. So, please remove or replace by a cartoon or other symbol. Also, the diagrams under “Wasp chromosomes”, what exactly are they. Are they based on actual data or are they cartoons?. I believe this need to be corrected to make it clearer what the authors are trying to convey here.

Minor comments

Clarifications

page 2 line 38: For the phrase “However, in a number of cases”, the authors should clarify which cases by citing them. Otherwise remove the “number of cases” part.

page 2 line 41: For the sentence “The high frequency and relevance of such phenomenon has been recognized for decades for bacteria but was considered to have had a marginal impact on the evolution of metazoans” please provide citation(s) or remove.

page 3 line 85: In the sentence “However, close relatives of the donor viruses do not infect present-day wasps, nor infect their hosts.”, unless very strong evidence such as very large population surveying of these kind of viruses in a number of different species and populations is cited, authors should rephrase to something like “have not been identified, either because the “donor” viral lineages went extinct....”.

page 4 line 122: The authors talk about repetitive sequences. How were these estimated (e.g. RepeatModeller?).

page 5 line 151: Again, what do you mean by “highly significant”, please provide numbers.

page 5 line 171: “In addition, several typical intron-containing eukaryotic genes were predicted in the vicinity of these genes (Fig. 1).”. Where exactly are these shown in figure 1?. If they are clearly not in these figure, please make a new one were it can be easily discerned.

page 6 line 210: In the phrase “The phylogeny obtained after the sequencing of the PCR products was consistent with the species-tree obtained with the ITS2 sequences (Fig. S3B).”, I just don’t see this. While the two phylogenies show some congruency, they are not perfectly congruent. For example, the clade (*L. clavipes* + *L. boulandi*) + *L. guineaensis* is indeed recovered in both phylogenies. However, the position of *L. victoriae* and *L. heteromona* are not congruent between the two phylogenies. Please rephrase this.

page 7 line 243: In “The venom gland produces the VLPs that are released in the lumen (Fig. 6) and that finally reach the reservoir where they are stored until the emergence (Fig. 5E).” I don’t believe that the VLPs reaching the reservoir can be appreciated in Fig. 5E. Please clarify or remove.

page 8 line 268: Please just clarify in a couple of words if the primers used are internal to the genes or external.

page 11 line 399: The authors state “Several recent publications suggest that large, possibly full-genome insertions of symbiont into their host DNA do occur in the course of evolution, including from dsDNA viruses.”, but fail to cite the “several recent publications. Please cite these.

page 11 line 413: Again, for “Indeed, from a function point of view, the domestication we document here is very similar to what has been described in the microgastroid complex in **B**raconidae, in **C**ampopleginae, and in **B**anchinae” add references (and capitalise).

page 12 line 430: For “[...] of the PolyDNAviruses described above) but instead proteins”, please add references.

page 14 line 508: In “We extracted the DNA of a single female abdomen using Macherey-Nagel columns, similarly to what was performed for *L. boulandi*.”, I could not find this in the text. If it is actually not in there, please cite.

page 14 line 514: Please specify the insert size of the library.

page 14 line 528: Please state the BUSCO Arthropoda database version.

page 15 line 556: Please add the version of the software used, so as to know the defaults, and/or the full list of parameters chosen.

page 15 line 561: I am uncertain about the use here of the term “species tree”. I would rather use “concatenated protein phylogeny”.

page 16 line 589: Please just specify if the primers are internal to the gene.

page 16 line 598: Where the sequences reverse-aligned with a certain software? (Pal2Nal) or an in-house script (If so please include in supplementary material).

page 18 line 667: Please specify accessed date as to know which version of database and software was used by the server.

Corrections

page 1 line 20: “[...] either because no closely related descendant infect the wasps, [...]” authors should add the possibility that the virus lineage has not yet been identified/found.

page 1 line 26: Please rephrase “Intriguingly, the contemporary [...]” to “ Intriguingly, **this** contemporary [...]” to make it clearer that you are talking about the previously referred close relative.

page 2 line 45: In “[...] leading to genetic innovation” authors could reference 10.1038/nrmicro.2017.137 (e.g. “reviewed in X”), a nice review of functional HGT from bacteria to eukaryotes.

page 3 line 81: “whereas a beta nudivirus **has**”.

page 3 line 94: “so-called”.

page 4 line 138: In the phrase “[.] strong homologies [...]”, please correct to strong identity or similar. Sequences are either (putative) homologous or not.

page 4 line 141: Rephrase “ Two of them (**ORF** 27 and 66) are predicted” to make it more easily readable.

page 4 line 144: “In the following **section**, we will focus on the second class of genes identified by this blast analysis.”.

page 5 line 168: “by analysing”.

page 5 line 169: “easily detect”.

page 5 line 177: “Taken together, **these**”.

page 9 line 298: Rephrase “[...] deriving from **either** a direct ancestor of LbFV **or** from a closely related one.”.

page 9 line 318: Rephrase “, **and thus** other denomination **has** been proposed **in lieu** of VLP [26].”.

page 9 line 333: “In humans [...]”.

page 10 line 338: Reference 51 is weirdly located inside the parenthesis. Please check these throughout the text, as I found a couple located at weird spots in the text (e.g. ref 17).

page 17 line 621: Correct “actine”.

Modifications

page 4 line 124: Either provide a citation for “ which is most likely sufficient to get the whole gene set” or just remove it. I don’t think this explanation is necessary since authors state the coverage and the BUSCO results.

page 6 line 213: This whole paragraph constitutes a conclusion, it does not represent a result. So please either move to the discussion or to a specific conclusion section.

page 6 line 228: In “showing that they are all as essential [...]” please modify to a more appropriate wording such as “**suggesting** that they are all as essential [...]”. Unless you knock them out or show otherwise they are indeed essential, “showing” is an overstatement.

page 7 line 262: I believe the second sentence in “Interestingly, among “early” virally-derived genes, we identified a putative DNA polymerase (ORF58, see table 5). This opened the fascinating possibility that the DNA encoding those genes is amplified during this biological process.” belongs in the discussion. I suggest to leave a sentence stating the results, and the rest to be treated in the discussion.

page 8 line 287: Please add to this section the discussion sentence “ORF85 is an homologue of Ac81, a conserved protein found in all Baculoviruses” with its citation or your result from searching.

page 9 line 328: May I suggest joining this and the following paragraph. It reads nicer.

page 14 line 528: The whole phrase “For the four genomes analysed, the proportion of “missing genes” was < 3 .5%. The statistic was even better for the three *Leptopilina* genomes (“missing genes” < 1 .9%), and the proportion of fragmented genes was also reduced compared to *Ganaspis xanthopoda* (<1 .5% for *Leptopilinas* versus 18% for *Ganaspis*.” belongs in the results section.

Tables 2, 3, 4: May I suggest to do a summary table for the main article (ORF gene and scaffold ID from the hit along with identity and alignment length). I pretty sure all of these can be summarised in a single table and the full tables can be included as a plain text (tab-separated columns) file in supplementary material.

Table 5: Again, I suggest to move this to supplementary material as a plain text file.

Figure 1: Can you please specify what are the grey “brackets” (eukaryotic genes surrounding the virally-derived genes?).

Sincerely,

Alejandro Manzano Marín