



Peer Community In Genomics

Using transcriptomics and proteomics to understand the expansion of a secreted poisonous armoury in parasitoid wasp genomes

Ignacio Bravo  based on peer reviews by **Inacio Azevedo** and 2 anonymous reviewers

Dominique Colinet, Fanny Cavigliasso, Matthieu Leobold, Appoline Pichon, Serge Urbach, Dominique Cazes, Marine Poulet, Maya Belghazi, Anne-Nathalie Volkoff, Jean-Michel Drezen, Jean-Luc Gatti, and Marylène Poirié (2023) Convergent origin and accelerated evolution of vesicle-associated RhoGAP proteins in two unrelated parasitoid wasps. bioRxiv, ver. 3, peer-reviewed and recommended by Peer Community in Genomics. <https://doi.org/10.1101/2023.06.05.543686>

Submitted: 12 June 2023, Recommended: 15 March 2024

Cite this recommendation as:

Bravo, I. (2024) Using transcriptomics and proteomics to understand the expansion of a secreted poisonous armoury in parasitoid wasp genomes. *Peer Community in Genomics*, 100248. [10.24072/pci.genomics.100248](https://doi.org/10.24072/pci.genomics.100248)

Published: 15 March 2024

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

Parasitoid wasps lay their eggs inside another arthropod, whose body is physically consumed by the parasitoid larvae. Phylogenetic inference suggests that Parasitoida are monophyletic, and that this clade underwent a strong radiation shortly after branching off from the Apocrita stem, some 236 million years ago (Peters et al. 2017). The increase in taxonomic diversity during evolutionary radiations is usually concurrent with an increase in genetic/genomic diversity, and is often associated with an increase in phenotypic diversity. Gene (or genome) duplication provides the evolutionary potential for such increase of genomic diversity by neo/subfunctionalisation of one of the gene paralogs, and is often proposed to be related to evolutionary radiations (Ohno 1970; Francino 2005).

In their recent preprint, Dominique Colinet and coworkers have explored the genetic and functional diversity of a Rho GTPase activating protein (RhoGAP) multigene family in two very divergent wasp clades within Parasitoida, namely *Leptopilina* (Figitidae) and *Venturia* (Ichneumonidae) (Colinet et al. 2024). Some members of the RhoGAP family are present in the venom of the parasitoid wasp *Leptopilina boulardi* as well as in other *Leptopilina*

species, and are probably involved in the parasitic lifestyle by binding and inactivating host's Rho GTPases, thereby interfering with the host's immune response (Colinet et al. 2007).

Venom protein composition is highly variable, even between very closely related species, and is subject to rapid evolutionary changes. Although gene duplication and subsequent neo/subfunctionalisation have been frequently proposed as the main mechanism underlying this evolutionary diversification, observations are often compatible with alternative explanations, such as horizontal gene transfer, gene co-option or multi-functionalisation (Martinson et al. 2017; Alvarado et al. 2020; Huang et al. 2021; Undheim and Jenner 2021). Furthermore, high mutation rates in venom protein-encoding genes hinder phylogenetic hypothesis testing, and venom proteomics can be needed to verify transcriptomic predictions (Smith and Undheim 2018; von Reumont et al. 2022).

Colinet and coworkers (2024) have applied a combined transcriptomic, proteomic and functional approach to i) identify potential transcripts of the RhoGAP family in *Leptopilina* species using experimental and bioinformatic approaches; ii) experimentally identify proteins of the RhoGAP family in the venom of three *Leptopilina* species; iii) identify transcripts and proteins of the RhoGAP family in the ovarian calyx of *Venturia canescens*; and iv) perform phylogenetic and selection analyses on the extant sequences of these RhoGAP family genes to propose an evolutionary scenario for their origin and diversification. The most striking results are first the large diversity of RhoGAP sequences retrieved in the transcriptomes and proteomes of *Leptopilina* and of *V. canescens*, and second the high number of branches and positions identified to have evolved under positive selection. All the retrieved hits share a RhoGAP domain, either alone or in tandem, preceded in the case of *Leptopilina* RhoGAPs by a signal peptide that may be responsible for protein vehiculation for venom secretion. Further, for some of the protein positions identified to have evolved under positive selection, the authors have experimentally verified the functional impact of the changes by reverse genetic engineering.

The authors propose an evolutionary scenario to interpret the phylogenetic relationships among extant RhoGAP diversity in the clades under study. They posit that two independent, incomplete duplication events from the respectively ancestral RacGAP gene, followed by subsequent, lineage- and paralog-specific duplication events, lie at the origin of the wealth of diversity of in the *Leptopilina* venom RhoGAPs and of *V. canescens* ovarian calyx RhoGAPs. Notwithstanding, the global relationships presented in the work are not systematically consistent with this interpretation, e.g. regarding the absence of monophyly for *Leptopilina* RhoGAPs and *Leptopilina* RacGAP, and the same holds true for the respective *V. canescens* sequences. It may very well be that the high evolutionary rate of these genes has eroded the phylogenetic signal and prevented proper reconstruction, as the large differences between codon-based and amino acid-based phylogenies and the low support suggest. Explicit hypothesis testing, together with additional data from other taxa, may shed light onto the evolution of this gene family.

The work by Colinet and coworkers communicates sound, novel transcriptomic, proteomic and functional data from complex gene targets, consolidated from an important amount of experimental and bioinformatic work, and related to evolutionarily intriguing and complex phenotypes. These results, and the evolutionary hypothesis proposed to account for them, will be instrumental for our understanding of the evolution and diversity of vesicle-associated RhoGAPs in divergent parasitoid wasps.

References:

Alvarado, G., Holland, S., R., DePerez-Rasmussen, J., Jarvis, B., A., Telander, T., Wagner, N., Waring, A., L.,

- Anast, A., Davis, B., Frank, A., et al. (2020). Bioinformatic analysis suggests potential mechanisms underlying parasitoid venom evolution and function. *Genomics* 112(2), 1096–1104. <https://doi.org/10.1016/j.ygeno.2019.06.022>
- Colinet, D., Cavigliasso, F., Leobold, M., Pichon, A., Urbach, S., Cazes, D., Pouillet, M., Belghazi, M., Volkoff, A.-N., Drezen, J.-M., Gatti, J.-L., and Poirié, M. (2024). Convergent origin and accelerated evolution of vesicle-associated RhoGAP proteins in two unrelated parasitoid wasps. *bioRxiv*, ver. 3 peer-reviewed and recommended by Peer Community in Genomics. <https://doi.org/10.1101/2023.06.05.543686>
- Colinet, D., Schmitz, A., Depoix, D., Crochard, D., and Poirié, M. (2007). Convergent Use of RhoGAP Toxins by eukaryotic parasites and bacterial pathogens. *PLoS Pathogens* 3(12), e203. <https://doi.org/10.1371/journal.ppat.0030203>
- Francino, M.P. (2005). An adaptive radiation model for the origin of new gene functions. *Nature Genetics* 37, 573–577. <https://doi.org/10.1038/ng1579>
- Huang, J., Chen, J., Fang, G., Pang, L., Zhou, S., Zhou, Y., Pan, Z., Zhang, Q., Sheng, Y., Lu, Y., et al. (2021). Two novel venom proteins underlie divergent parasitic strategies between a generalist and a specialist parasite. *Nature Communications* 12, 234. <https://doi.org/10.1038/s41467-020-20332-8>
- Martinson, E., O., Mrinalini, Kelkar, Y. D., Chang, C-H., and Werren, J., H. 2017. The evolution of venom by co-option of single-copy genes. *Current Biology* 27(13), 2007-2013.e8. <https://doi.org/10.1016/j.cub.2017.05.032>
- Ohno, S. (1970). *Evolution by gene duplication*. New-York: Springer-Verlag.
- Peters, R., S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., Lanfear, R., et al. (2017). Evolutionary history of the Hymenoptera. *Current Biology* 27(7), 1013–1018. <https://doi.org/10.1016/j.cub.2017.01.027>
- von Reumont, B., M., Anderluh, G., Antunes, A., Ayvazyan, N., Beis, D., Caliskan, F., Crnković, A., Damm, M., Dutertre, S., Ellgaard, L., et al. (2022). Modern venomomics—Current insights, novel methods, and future perspectives in biological and applied animal venom research. *GigaScience* 11, giac048. <https://doi.org/10.1093/gigascience/giac048>
- Smith, J., J., and Undheim, E., A., B. (2018). True lies: using proteomics to assess the accuracy of transcriptome-based venomomics in centipedes uncovers false positives and reveals startling intraspecific variation in *Scolopendra subspinipes*. *Toxins* 10(3), 96. <https://doi.org/10.3390/toxins10030096>
- Undheim, E., A., B., and Jenner, R., A. (2021). Phylogenetic analyses suggest centipede venom arsenals were repeatedly stocked by horizontal gene transfer. *Nature Communications* 12, 818. <https://doi.org/10.1038/s41467-021-21093-8>

Reviews

Evaluation round #2

DOI or URL of the preprint: <https://doi.org/10.1101/2023.06.05.543686>

Version of the preprint: 2

Authors' reply, 16 February 2024

Dear Recommender and Reviewers,

Once again, we thank the recommender and the reviewers for their careful and constructive comments, which helped us to improve the manuscript. We address each of the comments (reviewer 2 and recommender) in detail below.

Yours sincerely,

Dominique Colinet

Reviewer 2:

- Comment 1:

I suggest that the new BLAST and Exonerate search analyses, conducted in response to my earlier feedback, should be detailed either in the main text or Supplementary Text. This inclusion is important to transparently convey the absence of sequences similar to the new exon, even if the results are negative.

Authors' response:

L212-214 and L411-414: Some details of the analysis performed to identify the possible origin of the signal peptide of *Leptopilina* venom RhoGAPs are given. No significant sequence similarity was found between the exon(s) preceding the RhoGAP domain coding sequence of *Leptopilina* venom RhoGAPs and other sequences in the *Leptopilina* genomes that could provide an indication of the origin of the signal peptide. The sequences used for the analysis and the results of the similarity searches (using Exonerate, which is more suitable for this type of analysis than BLAST) are given in Supplementary Dataset 2.

- Comment 2:

Additionally, the authors have not presented a gene family-wide, unified phylogeny in the revised Figure 1. Such a phylogeny is essential to clearly demonstrate the independent origins of the two venom protein lineages within the same protein family. While I believe its inclusion would greatly enhance readers' understanding, the decision to omit it ultimately lies with the authors.

Authors' response:

A unified phylogenetic analysis was performed using either protein- or codon-based alignment of the RhoGAP domain, and the resulting phylogenies are shown in Supplementary Figure S4.

L457-467: The results of the phylogenetic analyses are presented in relation to the independent duplication hypothesis. The analysis confirmed a relationship between either *Leptopilina* venom RhoGAPs or *V. canescens* calyx RhoGAPs with RacGAP1 since they all form a robust monophyletic group. In the codon-based phylogeny, *V. canescens* calyx RhoGAPs formed a robust monophyletic group with VcRacGAP1, suggesting that they originated from an independent duplication event. However, in the protein-based phylogeny, the *V. canescens* calyx RhoGAPs did not form a robust monophyletic group with VcRacGAP1. Similarly, the *Leptopilina* venom RhoGAPs did not form a robust monophyletic group with *Leptopilina* RacGAP1 in either phylogeny. This was not unexpected due to the high divergence of *Leptopilina* venom RhoGAPs and *V. canescens* calyx RhoGAPs compared to RacGAP1. Nevertheless, like reviewer 1 in the round 1 revision, we are convinced of the hypothesis of independent duplication in both *Leptopilina* and *Venturia*, since these are very distant and the presence of RhoGAP family proteins in the maternal fluids of parasitoid wasps has not been described outside of *Leptopilina* and *Venturia* to date (see comparison of hypotheses made in the discussion and described below).

L625-630: A comparison was made between the independent duplication hypothesis and the alternative hypothesis of a single ancestral duplication. The latter would imply thousands or even hundreds of thousands of loss events, since (i) the last common ancestor of *Leptopilina* and *Venturia* dates back to the early radiation of parasitoid wasps (over 200 million years ago), and (ii) *Leptopilina* and *Venturia* are the only parasitoid wasps for which the presence of RhoGAP family proteins in the maternal fluids has been observed. The hypothesis of independent duplication events of RacGAP1 in both *Leptopilina* and *Venturia* seems much more parsimonious than the alternative hypothesis of a single ancestral duplication of RacGAP1.

Recommender:

- Comment 1:

I would nevertheless agree with reviewer#2 regarding the need for a higher-level phylogenetic integration of the results. Such an analysis should include in the same inference the proposed orthologs for all species studied. Otherwise the evolutionary relationships are inferred by comparison between parallel analyses, as it is the case in the present version of the text.

[...]

Notwithstanding, as reviewer #2 indicates, there is no analysis that allows for introducing polarity in the tree, and confidently establishing the independence and the order of the initial duplication events (e.g. an analysis of the phylogenetic relationships of NVIT RACGAP1, *Leptopilina* and *Venturia* RacGaps, and some representatives of the *Leptopilina* and the *Venturia* venom GAPs, chosen based on the relationships depicted in figs S2 and S4). Given the quality of the amino-acid based alignment and the increased information available at the nucleotide level (as inferred from the genetic distance matrices) I would suggest to run the phylogenetic inference at the nucleotide level (or at the codon level including codon-based nucleotide partitions), so that the fine relationships (mostly in the VcGAP analyses) can be resolved. It is also my impression that an analysis aiming at identifying and removing rogue taxa in the dataset (probably the case of VcGAP5) will help clarify the relationships.

Authors' response:

As described above in response to reviewer 2's comment 2, a unified phylogenetic analysis was performed using either protein- or codon-based alignment of the RhoGAP domain, and the resulting phylogenies are shown in Supplementary Figure S4. The results of the phylogenetic analyses are presented in relation to the independent duplication hypothesis (L457-467) and a comparison was made between the independent duplication hypothesis and the alternative hypothesis of a single ancestral duplication (L625-630) (see above).

The phylogenetic analysis included all *Nasonia* RhoGAPs and not only RacGAP1 to confirm the relationship of either *Leptopilina* venom RhoGAPs or *V. canescens* calyx RhoGAPs with RacGAP1. The phylogenetic analysis included all *Leptopilina* venom RhoGAPs or *V. canescens* calyx RhoGAPs since we could not find a suitable criterion to select among them. Furthermore, we believe that the analysis was more informative when all sequences were included.

- Comment 2:

Overall, this analysis should allow establish:

1) whether RacGAP gene-based relationships for *Leptopilina*, *Venturia* and *Nasonia* recapitulate the species relationships (it would be good to have for this the LvRacGAP, but this is a minor detail).

Authors' response:

The relationships for RACGAP1 in *Leptopilina*, *Venturia* and *Nasonia* do not perfectly recapitulate the species relationships. According to Peters et al. (2017), *Nasonia* and *Leptopilina* shared a common ancestor (more than 200 million years ago) before sharing one with *Venturia*, whereas *Venturia* and *Nasonia* RACGAP1 were grouped together in the phylogenies (although this was not robust in the protein-based phylogeny). However, we do not feel it is relevant enough to discuss explanations for this discrepancy in the manuscript.

There was an error regarding LvRacGAP1, as we could not identify its sequence from the transcriptomic data we obtained. The manuscript has been corrected.

2) whether VcGAPs and *Leptopilina* venom RhoGAPs are respectively monophyletic (this seems to be the case for *Leptopilina*, but less clear for *Venturia*), and what their respective relationships are with VcRacGAP1 and with the *Leptopilina* RacGAP1 (i.e. whether they have actually appeared after independent duplication events after the speciation between the *Venturia* and the *Leptopilina* clades).

Authors' response:

This point is discussed in the response to reviewer 2's comment 2 (see above).

3) whether Lb/mGAPs 1,3,4 and 5 are absent in Lh/c/v; should this be the case, all the subsequent duplication events in this lineage would be exclusive to the Lby/m (most likely after loss of the duplicate in Lh/c/v).

Authors' response:

See combined answer below.

4) whether Lb/mGAP and Lb/mGAPs2 and 6 are sister taxa to Lc/h/vGAPs; should this be the case, further duplication events may have occurred in the Lc, Lh and Lv lineage, e.g. (all this needs to be interpreted in the light of ITS2-based relationships in fig1).

Authors' response (points 3 and 4):

Unfortunately, the unified phylogenetic analysis performed using either protein or codon-based alignment of the RhoGAP domain does not allow the complete evolutionary history of *Leptopilina* venom RhoGAPs to be accurately described. *Leptopilina* venom RhoGAPs form a monophyletic group in the protein-based phylogeny (although not robust), but not in the codon-based phylogeny, and *L. bouleardi* venom RhoGAPs do not form a monophyletic group in any phylogeny. This could be explained by the high divergence of *Leptopilina* venom

RhoGAPs even at the codon level compared to RacGAP1. Furthermore, we are not sure that we found all venom RhoGAPs in *Leptopilina* species other than *L. boulardi* due to the quality of the transcriptomes and/or genomes we analyzed. For example, we did not find the classical RacGAP1 in the transcriptome of *L. victoricae*, so there may be venom RhoGAPs other than LvGAP1 and LvGAP2 in this species. However, our goal in including venom RhoGAPs from *Leptopilina* species other than *L. boulardi* was not to describe the complete evolutionary history of *Leptopilina* venom RhoGAPs. Our goal was to determine whether the occurrence of venom RhoGAPs was not exclusive to *L. boulardi* and likely originated from an ancestral duplication in the ancestor of the *Leptopilina* genus. The relationship of all *Leptopilina* venom RhoGAPs to classical RacGAP1 and the common domain organization with a signal peptide followed by a RhoGAP domain (whereas RacGAP1 has a different domain organization) are consistent with the hypothesis of an ancestral duplication.

Although an accurate and complete description of the evolutionary history of *Leptopilina* venom RhoGAPs was not our goal and would be challenging based on the phylogenies obtained, some specific points are already discussed in the manuscript. In addition, we have added a sentence regarding the observation that some of the *L. boulardi* venom RhoGAPs form robust monophyletic groups, suggesting that some of the duplication events following the initial duplication of the RacGAP1 gene in the ancestor of the *Leptopilina* genus appear to be specific to *L. boulardi* and would explain the large number of venom RhoGAPs found in this species (L415-417).

- Comment 3:

Regarding the mass spectrometry analyses, it is unclear from the results whether the database used for searching included only the *Leptopilina* venom RhoGAP sequences or all *Leptopilina* known CDSs. This may be important to understand the multiple matches of a same band to several protein sequences, and of a same protein to several bands, especially when considering very closely related proteins, as in LgmGAP1/2/3.x.

Authors' response:

As explained in the Methods section (L318-320) and Results section (L435-436), the coding sequences of *Leptopilina* venom RhoGAPs were used to perform Mascot searches on the mass spectrometry data. However, this has been clarified in the legend of Figure 2.

The occurrence of two or more venom RhoGAPs to the same protein band, and of two or more protein bands to the same venom RhoGAP, was a common observation in our analysis of venom proteins from various parasitoid wasps using 1D gel electrophoresis. This can be attributed to the abundance of some proteins in the venom and the high similarity between some of the *L. boulardi* venom RhoGAPs. However, our objective was to investigate whether we could hypothesize that venom RhoGAPs from *L. boulardi* and *L. heterotoma* are associated with venosomes following their secretion in the venom (and transported by these vesicles to target host cells, as previously shown for LbmGAP and LbmGAP2). We do not believe that a discussion on the occurrence of multiple venom RhoGAPs associating with the same protein band or vice versa would be relevant within the manuscript.

- Comment 4:

It would finally be interesting to project onto the phylogeny and to discuss the differential location vesicular/supernatant, which are puzzling for orthologs (LbmGAP4 and LbyGAP4; LbyGAP6 and LbyGAP), and also the vesicular-only detection of the LhGAP1/2/3.

Authors' response:

In the manuscript, we showed that all *L. bouleari* and *L. heterotoma* venom RhoGAPs were detected in the vesicular fraction, where most of them were enriched compared to the supernatant fraction. This led us to hypothesize that *L. bouleari* and *L. heterotoma* venom RhoGAPs are associated with venosomes following their secretion in the venom. The variations among some of the *L. bouleari* ISm and ISy orthologs can be attributed to differences in the abundance of venom proteins, as described in our previous works. For example, we previously evidenced that LbmGAP is abundant in the venom of *L. bouleari* ISm, while LbyGAP is present in much lower quantities in the venom of *L. bouleari* ISy, due to differences in transcription levels. The detection of *L. heterotoma* venom RhoGAPs only in the vesicular fraction can be attributed to their low abundance (described previously). However, we believe that a detailed discussion regarding such variations among some *L. bouleari* orthologs and the vesicular-only detection of some of the venom RhoGAPs is beyond the scope of this manuscript.

[Download author's reply](#)

[Download tracked changes file](#)

Decision by Ignacio Bravo , posted 25 January 2024, validated 25 January 2024

Minor suggestions for the analyses prior to recommendation.

In the revised version of their text, Colinet and coworkers have satisfactorily addressed most of the comments raised by the reviewers and by this recommender, and have provided most of the raw data, which allows for independent validation of their findings. I think the study question is pertinent and timely, that the methods are appropriate and well conducted, and that the findings will be interesting for the two communities working on venom components and on gene duplication. I would nevertheless agree with reviewer#2 regarding the need for a higher-level phylogenetic integration of the results. Such an analysis should include in the same inference the proposed orthologs for all species studied. Otherwise the evolutionary relationships are inferred by comparison between parallel analyses, as it is the case in the present version of the text. Specific suggestions follow below.

In this respect, I think the authors have confidently established:

- 1) the respective orthology of the RacGAP sequences in the Leptopilina and in the Venturia with Nasonia (acGAP1).
- 2) the presence of numerous duplication events in the Leptopilina and in the Venturia lineages that have generated the respective GAP sequences.
- 3) the differential presence of Leptopilina venom RhoGAPs in the supernatant and in the vesicular fractions.
- 4) the key role of certain amino acid substitutions for the functional interaction with Rac GTPases.

Notwithstanding, as reviewer #2 indicates, there is no analysis that allows for introducing polarity in the tree, and confidently establishing the independence and the order of the initial duplication events (e.g. an analysis of the phylogenetic relationships of NVIT RACGAP1, Leptopilina and Venturia RacGaps, and some representatives of the Leptopilina and the Venturia venom GAPs, chosen based on the relationships depicted in figs S2 and S4). Given the quality of the amino-acid based alignment and the increased information available at the nucleotide level (as inferred from the genetic distance matrices) I would suggest to run the phylogenetic inference at the nucleotide level (or at the codon level including codon-based nucleotide partitions), so that the fine relationships (mostly in the VcGAP analyses) can be resolved. It is also my impression that an analysis

aiming at identifying and removing rogue taxa in the dataset (probably the case of VcGAP5) will help clarify the relationships. Overall, this analysis should allow establish:

1) whether RacGAP gene-based relationships for Leptopilina, Venturia and Nasonia recapitulate the species relationships (it would be good to have for this the LvRacGAP, but this is a minor detail).

2) whether VcGAPs and Leptopilina venom RhoGAPs are respectively monophyletic (this seems to be the case for Leptopilina, but less clear for Venturia), and what their respective relationships are with VcRacGAP1 and with the Leptopilina RacGAP1 (i.e. whether they have actually appeared after independent duplication events after the speciation between the Venturia and the Leptopilina clades).

3) whether Lb/mGAPs 1,3,4 and 5 are absent in Lh/c/v; should this be the case, all the subsequent duplication events in this lineage would be exclusive to the Lby/m (most likely after loss of the duplicate in Lh/c/v).

4) whether Lb/mGAP and Lb/mGAPs2 and 6 are sister taxa to Lc/h/vGAPs; should this be the case, further duplication events may have occurred in the Lc, Lh and Lv lineage, e.g. (all this needs to be interpreted in the light of ITS2-based relationships in fig1).

Regarding the mass spectrometry analyses, it is unclear from the results whether the database used for searching included only the Leptopilina venom RhoGAP sequences or all Leptopilina known CDSs. This may be important to understand the multiple matches of a same band to several protein sequences, and of a same protein to several bands, especially when considering very closely related proteins, as in LgmGAP1/2/3.x. It would finally be interesting to project onto the phylogeny and to discuss the differential location vesicular/supernatant, which are puzzling for orthologs (LbmGAP4 and LbyGAP4; LbyGAP6 and LbyGAP), and also the vesicular-only detection of the LhGAP1/2/3.

Reviewed by Inacio Azevedo, 26 December 2023

The authors fully addressed the questions I raised and added other important information requested during the review process. I have no further comments and thus recommend the article.

Reviewed by anonymous reviewer 2, 21 December 2023

I appreciate the authors' efforts in addressing many of my previous comments. I suggest that the new BLAST and Exonerate search analyses, conducted in response to my earlier feedback, should be detailed either in the main text or Supplementary Text. This inclusion is important to transparently convey the absence of sequences similar to the new exon, even if the results are negative. Additionally, the authors have not presented a gene family-wide, unified phylogeny in the revised Figure 1. Such a phylogeny is essential to clearly demonstrate the independent origins of the two venom protein lineages within the same protein family. While I believe its inclusion would greatly enhance readers' understanding, the decision to omit it ultimately lies with the authors.

Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.1101/2023.06.05.543686>

Version of the preprint: 1

Authors' reply, 18 December 2023

Dear Recommender and Reviewers,

We thank the recommender and the reviewers for their careful and constructive comments, which helped us to improve the manuscript. We address each of the comments (reviewer 1, reviewer 2 and recommender) in detail below.

Yours sincerely,

Dominique Colinet

Reviewer 1:

- Comment 1:

I am convinced about the convergence of having RhoGAP recruited to venom independently in both groups, since they are very distant. However, I missed more information about the situation in the sister taxa in both cases, what would be important to understand when the recruitment occurred and what the functional meaning. For example, in the sequence phylogenies of figures 1 and 3, a NvRacGAP from *N. vitripennis* is shown as the sister sequence to both *Leptopilina* and *V. canescens* RhoGAPs. As far as I understand, *Nasonia vitripennis* is distantly related to both families, so what happens in the several genera phylogenetically placed between them? Are there other genus close to *Leptopilina* and *Venturia* that could also share the venom protein, thus representing an earlier event of recruitment? Moreover, it is not clear for me, as a non-specialist in the field, if the other genera of these families are also parasitoid. If yes, do they produce venoms that act in the process? And if so, do these venoms have any sort of RhoGAP?

Perhaps, adding more background about the group would help to understand if the reported RhoGAP recruitments were conditions to develop a major phenotype (venom mediated parasitism) or were selected to further improve it.

Authors' response:

L120-125: Background has been added in the Introduction to explain that the group of parasitoid wasps that includes *Leptopilina* species and *V. canescens* (and also *N. vitripennis*) is monophyletic and that *Leptopilina* species and *V. canescens* are distantly related (last common ancestor in the early radiation of parasitoid wasps). We also explained that the presence of RhoGAP family proteins in the maternal fluids of parasitoid wasps has not been described outside of these species. This suggests that the recruitment of RhoGAPs in parasitoid virulence was independent in these species and that these events represent an enhancement rather than a requirement for virulence.

L368-371: Some precisions were made in the use of *N. vitripennis* to include classical intracellular RhoGAPs in the phylogenetic analysis of *Leptopilina* venom RhoGAPs (and *Venturia calyx* RhoGAPs). *N. vitripennis* was the first parasitoid wasp to have its genome sequenced, and its genome has recently been re-annotated. Furthermore, the only RhoGAPs found in *N. vitripennis* are classical intracellular RhoGAPs, which are absent in maternal fluids.

L387-388: A precision was made to make clear that the robust grouping of *N. vitripennis* RacGAP1 with *Leptopilina* venom RhoGAPs suggests that *Leptopilina* venom RhoGAPs are derived from duplication of the *N. vitripennis* RacGAP1 ortholog in *Leptopilina*.

- Comment 2:

Beyond the evolutionary questions, the authors did a terrific job in finding and validating the mutations of

these proteins, further detailing biochemical aspects of the venom proteins.
The methodology used is overall appropriated and the article brings novelty to its field.

Authors' response:

Thanks !

Reviewer 2:

- Comment 1:

For better comprehension through side-by-side comparison, RhoGAPs in the Leptopilina and Venturia lineages should be compared in the same figure. This might necessitate a substantial reorganization of figures. Combining Fig. 1 and Fig. 3 would clearly show the distinct origin of their venom proteins and present their phylogenetic positions on a more sensible scale. A gene family-wide, unified phylogeny should be presented.

Authors' response:

Figures 1 and 3 have been partially combined to show the phylogeny and protein domain organization of both Leptopilina venom RhoGAPs and Venturia RhoGAPs in Figure 1. This facilitates side-by-side comparison.

- Comment 2:

Fig. 5 and Fig. 6 should also be combined to facilitate an easy comparison of how different/similar selected sites are in the two lineages.

Authors' response:

Figures 5 and 6 have been combined to facilitate comparison.

- Comment 3:

Also, the description of the protein domain and exon structures for individual genes is somewhat lengthy. While detailed descriptions are appreciated, more concise explanations would facilitate understanding overall evolutionary trends. It is recommended that these detailed texts be moved to Supplemental Information and be made more concise in the main text.

Authors' response:

The authors believe that the description of domain and exon structures is necessary for a proper understanding of the manuscript. Moving this description to supplementary data is not necessarily relevant.

- Comment 4:

L110: Clarification is needed on the "two multigene families" mentioned, as only one (RhoGAP) seems to be described.

Authors' response:

The "two multigene families" referred to the one found in Leptopilina genus and the other in *V. canescens*.
L131-137: The sentences have been clarified accordingly.

- Comment 5:

L239: Please explain what β & ω symbolize.

Authors' response:

L274-275: An explanation has been given for β & ω .

- Comment 6:

L285: Clarify whether the 25 samples were processed separately or partly pooled.

Authors' response:

L331-333: The sentence has been clarified to explain that the reservoirs obtained from the dissection of the venom apparatus of the 25 females were pooled before to collect the venom content.

- Comment 7:

L320: The term "completed" needs clarification. Does it indicate that the complete coding sequences were recovered using partial protein sequences?

Authors' response:

As described in the Methods section, the complete coding sequence of several of the *Leptopilina* venom RhoGAPs was obtained by rapid amplification of cDNA ends (RACE) for most, or by searching the transcriptome assembly and verification by RT-PCR for one.

L372-373: The sentence has been clarified accordingly.

- Comment 8:

L345: It would be enlightening to explore whether the exons encoding the signal peptide originated through de-novo gene birth or by exon shuffling with other genes. Is there any evidence of sequence similarity with other genes in the genome? Such information would be pertinent to the discussion on their origin, which is addressed later in the Discussion section.

Authors' response:

We agree that the question of the origin of the signal peptide is very interesting. However, it is also challenging. Similarity searches using BLAST and Exonerate tools were performed with the exon(s) encoding the signal peptide of *Leptopilina* venom RhoGAPs. Unfortunately, no similarity to other sequences in the genomes could be found. This is not so surprising, since divergence is observed not only in the RhoGAP domain, but also in the N-terminal part of venom RhoGAPs, although the ability to encode a signal peptide is conserved. The origin of the signal peptide in *Leptopilina* venom RhoGAPs is thus still an open question.

- Comment 9:

L421: "Mascot 5" does not appear in the table.

Authors' response:

L531: "Mascot 5" has been replaced with "Mascot" in the legend of Table 2.

- Comment 10:

L485: Providing the number of intersection, as well as union, of sites detected by different methods is suggested.

A supplementary table has been provided detailing the results obtained with the different methods for the sites detected under positive or negative selection (Supplementary Table S8).

- Comment 11:

Fig. 1B, Fig. 3A, B: Rooting method should be described. This is applicable to other figures as well.

Authors' response:

The description of the rooting method has been added to each figure showing a phylogenetic tree.

- Comment 12:

Table 3: Clarify what the parameter p provided along with ω is.

Authors' response:

L628-629: The meaning of the p parameter has been added to the legend of Table 3.

- Comment 13:

Fig. 5E, Fig. 6E: Since the site-wise relative solvent accessible surface can be computed, the current visualization method loses its quantitative information. I recommend exploring alternative forms of visualization to convey this data more effectively. This suggestion is also applicable to Fig. 6E.

Authors' response:

In the spiral view, as explained in the legend of Figure 5, the radius of the circles corresponds to the relative solvent accessibility. The quantitative information is therefore already shown in both figures. However, a supplementary figure has been added that shows the relative solvent accessibility as a histogram (Supplementary Figure S6).

Recommender:

- Comment 1:

I think the manuscript may benefit from a more thorough evolutionary analysis, proposing a timeline (or at least an order) for the duplication events described.

Authors' response:

Unfortunately, we do not believe that a timeline or an order can be proposed for the duplication events, as a number of nodes in the phylogenies obtained for *Leptopilina* and *Venturia* are not supported by bootstrap analyses. In the case of *Leptopilina*, we hypothesized in the manuscript that the venom RhoGAPs arose from a single ancestral partial duplication of the *RacGAP1* gene, followed by further duplication of the ancestral duplicated gene during the diversification of the genus *Leptopilina* (L364-411). After the ancestral partial duplication event, however, it is difficult to trace the events, especially since the *L. bouvardi* sequences form several robust groups with each other, but never with other *Leptopilina* species. In the case of *V. canescens*, the number of robust nodes is even lower. However, we could hypothesize in the manuscript that the calyx RhoGAPs probably originated from two or more imperfect duplication events of the *RacGAP1* gene (L510-559).

- Comment 2:

I also consider that a more global approach to the subject of gene duplication as a source of novelty may help conceptualise the findings. A reference to Ohno's works may also be needed.

Authors' response:

L46-52: The existing part corresponding to the importance of gene duplication in genetic and functional innovation in the Introduction has been further developed.

- Comment 3:

I think also that performing some analyses at the nucleotide level may help interpret the mechanisms promoting functional diversity, identify possible instances of gene replacement, or provide keys for understanding the possible loss of function of the enzymes. In this regard, the literature regarding a role of nucleotide

composition divergence of in-paralogs that retain a high amino acid identity may be relevant for discussion and interpretation.

Authors' response:

In our results, pairwise sequence identity was lower at the amino acid level than at the nucleotide level for both *Leptopilina* venom RhoGAPs and *V. canescens* calyx RhoGAPs. Thus, the evolution of virulent RhoGAPs appears to be the opposite of the observations in the literature you mentioned (e.g., Bourret et al. 2023). This suggests a higher divergence at the codon level at non-synonymous sites compared to synonymous sites. Therefore, we performed analyses of positive selection at the codon level to detect possible functional divergence.

L617-619: An explanation regarding the pairwise sequence identity, which is lower at the amino acid level than at the nucleotide level for both *Leptopilina* venom RhoGAPs and *V. canescens* calyx RhoGAPs, has been added to clarify the reason for performing positive selection analyses at the codon level.

A supplementary figure was added with the pairwise sequence identity calculated at the nucleotide level and at the amino acid level for both *Leptopilina* venom RhoGAPs and *V. canescens* calyx RhoGAPs (Supplementary Figure S5).

—Phylogeny of members of the Cynipoidea superfamily

- Comment 4:

please include the fraction of the alignment that was removed with Trimal, and specify the length and pattern content of the final alignment used for reconstruction.

Authors' response:

The alignment length of the ITS2 nucleic acid sequence was 837 sites before and 535 sites after the use of Trimal (36% of the alignment sites were deleted by Trimal). However, we do not believe that this information is relevant enough to be included in the manuscript. Nevertheless, the alignments obtained before and after Trimal have been deposited as a supplementary dataset for reproducibility.

- Comment 5:

Please include a proper phylogenetic tree with branch lengths.

Authors' response:

The phylogenetic tree in Figure 1A has been replaced by a phylogram showing branch lengths.

- Comment 6:

For reproducibility, I suggest to deposit or to include in SM the final alignment used for phylogenetic inference.

Authors' response:

The alignments obtained before and after using Trimal have been deposited as a supplementary dataset for each of the phylogenetic analyses we have performed (including that for members of the Cynipoidea superfamily).

—Phylogeny of *Leptopilina* venom and *Venturia* calyx RhoGAPs

- Comment 7:

Please provide accession numbers for proteins listed in TableS2.

Authors' response:

NCBI accession number was provided for RhoGAP proteins listed in Supplementary Table S2.

- Comment 8:

It is unclear whether phylogenetic inference shown in fig1b was performed at the amino acid or at the codon-based nucleotide level.

Authors' response:

L251-253: This was already mentioned in the Methods section, but has been further developed.

L473-475: This precision has been added to the legend of Figure 1.

- Comment 9:

For the sake of clarity, the abbreviation Lbm and Lby are used for the two Lb isogenic lines, but the sequences and branches do not include systematically the m or y indication, or it is included misleading, as in "LbGAPy4", instead of "LbyGAP4". The reader needs to resort to table1 for decoding.

Authors' response:

The names of the *L. bouleari* venom RhoGAPs have been changed throughout the manuscript, figures and tables by using the abbreviations Lbm and Lby as prefixes for better comprehension: LbmGAP and LbyGAP instead of LbGAP and LbGAPy, for example.

- Comment 10:

It is never explained to what integration level the different sequences named with dot numbering (as in LbGAP1.1 and LbGAP1.2) refer to. I understand that they may be different alleles found among the females sequenced, but this need to be explained.

Authors' response:

The dot numbering (e.g., LbmGAP1.1, LbmGAP1.2, and LbmGAP1.3) was used for *Leptopilina* venom RhoGAPs that share high amino acid sequence identity in the RhoGAP domain (greater than 75% identity), but were considered genes rather than alleles due to their locations in the genomic scaffolds. This numbering was chosen to facilitate the comparison of the *L. bouleari* ISm venom RhoGAPs and their orthologs in *L. bouleari* ISy in the phylogenetic analysis.

L426-428: An explanation concerning the dot numbering has been added to the legend of Table 1.

- Comment 11:

The authors need to explain/substantiate whether their analyses are sensitive enough to simultaneously address inter and intraspecies diversity, as the inclusion of deeper branches for interspecies analyses comes along with a loss of the exploitable alignment length for the intraspecies analyses.

In fact, intraspecific divergence between *Leptopilina* venom RhoGAPs can be higher than interspecies divergence. For example, only 16.5% amino acid sequence identity is observed between LbmGAP and LbmGAP1.3 in the RhoGAP domain, whereas 20.4% to 28.5% identity is observed between LbmGAP and the venom RhoGAPs of the other *Leptopilina* species. However, there was comparatively less divergence in the length of the RhoGAP domain, and the length of the alignment after using Trimal was still sufficient for phylogenetic analysis. For example, 134 sites were retained for the alignment of the RhoGAP domain of *Leptopilina* venom RhoGAPs with *N. vitripennis* classical RhoGAPs, knowing that the longest sequence was 216 amino acids long and the

shortest 170 amino acids.

- Comment 12:

The razor algorithm underlying the results in figure 2B needs to be explained, as several protein bands are matched to different sequences, but still a differential identification for a higher abundance for some proteins in certain bands is provided.

Authors' response:

Several of the protein bands contained two or more venom RhoGAPs, and several of the venom RhoGAPs were found in two or more protein bands. We used the number of peptide matches identified by Mascot software on the MS/MS data to determine (i) the number of protein bands in which each venom RhoGAP was found and (ii) the protein band in which each venom RhoGAP was the most abundant.

L345-347: An explanation was added regarding the use of the number of peptide matches identified by Mascot software on the MS/MS data to determine (i) the number of protein bands in which each venom RhoGAP was found and (ii) the protein band in which each venom RhoGAP was the most abundant.

- Comment 13:

The functional and/or evolutionary connection between the RhoGAP domain (PF00620) and the RacGAP1 domain (PTHR46199) needs to be explained and discussed. The latter is not described in the methods and appears directly on tables 1 and 2.

Authors' response:

L225-226: A precision was added on the InterPro database searched by InterProScan.

L421-424 and L526-530: Some clarifications have been made in the legends of Table 1 and Table 2 regarding the RhoGAP domain (PF00620) and the RacGAP1 domain (PTHR46199). The RhoGAP domain (PF00620) is from the Pfam database and the RacGAP1 domain (PTHR46199) is from the Panther database. Both were identified using InterProScan on the InterPro integrative protein signature database. Compared to Pfam, the Panther database provides information about the specific RhoGAP subfamily to which the domain belongs.

- Comment 14:

In the version I have evaluated, FigS2 is not accompanied by any legend. The tree inference is not described. It is unclear whether only the blue, shared domain, was used for inference, and at what level.

Authors' response:

The missing legend has been added to Supplementary Figure S2.

- Comment 15:

The topology in Fig1b and in FigS2 are different. Please discuss.

Authors' response:

The two phylogenetic trees differ in that all *N. vitripennis* classical RhoGAPs were included in the tree corresponding to Figure 1B, whereas only *N. vitripennis* RacGAP1 (together with RacGAP1 from *L. boulandi*, *L. heterotoma* and *L. clavipes*) was included in the tree corresponding to Supplementary Figure S2. Regarding the topology of the venom RhoGAPs, only one difference was found between the two trees, concerning the position of the group consisting of LbmGAP4 and LbyGAP4. This difference is most likely due to a difference in the alignment as well as in the number of aligned sites (134 aligned sites in the alignment corresponding to

the phylogeny in Figure 1B compared to 146 aligned sites in the alignment corresponding to the phylogeny in Supplementary Figure S2). However, the node grouping LbmGAP4 and LbyGAP4 with other *L. bouleardi* venom RhoGAPs was not robust in either tree according to bootstrap analysis, meaning that no relationship can be inferred from that node. In contrast, the topology was exactly the same between the two trees for each of the groups considered robust by bootstrap analysis.

- Comment 16:

The correspondence (or the lack of) between the species and the genes trees for the Leptopilina RhoGAP sequences needs to be analysed and discussed. The authors may need to address possible instances of allele replacement or gene conversion to explain the topology described in figS2.

Authors' response:

Unfortunately, the question of correspondence between the species tree and the gene tree cannot be answered for the majority of venom RhoGAPs. The reason is that some of the nodes, especially those that group certain *L. bouleardi* venom RhoGAPs with venom RhoGAPs from other Leptopilina species, are not robust. From the ancestral node, a total of 6 robust nodes are found, 4 of which correspond to different groups of *L. bouleardi* venom RhoGAPs, but no venom RhoGAPs from other Leptopilina species are found in these 4 groups. In the only group containing the three other Leptopilina species (*L. heterotoma*, *L. victoriae* and *L. clavipes*), there is a correspondence between the species tree and the gene tree. The question of the correspondence between the species tree and the gene tree remains largely unresolved, most likely due to the high divergence between venom RhoGAPs, and we do not think that a discussion of this question in the manuscript is relevant enough.

—*V. canescens* calyx RhoGAPs probably evolved from two or more imperfect RacGAP1 duplication events

- Comment 17:

The position in the tree of *Nasonia* RacGAP1 in Fig3A and S4 as being very closely related to *Venturia* RacGAP12 is unexpected and needs to be discussed (unless this is a typo for *VcRacGAP1*)

Authors' response:

The grouping of *V. canescens* VcGAP12 and *N. vitripennis* RacGAP1 in the phylogeny was not supported by bootstrap analysis. Therefore, we cannot infer a close relationship between the two. In contrast, the grouping of all *V. canescens* calyx RhoGAPs with *N. vitripennis* RacGAP1 in Figure 1D (Figure 3A in the previous version) was robust, allowing us to infer a close relationship between the calyx RhoGAPs and the ortholog of RacGAP1 in *V. canescens* (the *N. vitripennis* and *V. canescens* RacGAP1 share 85.6% identity in the RhoGAP domain, whereas the identity between each of the calyx RhoGAPs with RacGAP1 in *V. canescens* ranged from 37.4% to 56.1%).

- Comment 18:

The topologies in Fig3A and S4 are discordant and need to be discussed.

Authors' response:

The two phylogenetic trees differ in that all *N. vitripennis* classical RhoGAPs were included in the tree corresponding to Figure 1D (Figure 3A in the previous version), whereas only *N. vitripennis* RacGAP1 together with *V. canescens* RacGAP1 was included in the tree corresponding to Supplementary Figure S4. Regarding the topology of *V. canescens* calyx RhoGAPs, a few differences are observed between the two phylogenetic trees, most likely due to a difference in the alignment as well as in the number of aligned sites (155 aligned sites in the alignment corresponding to the phylogeny in Figure 1D compared to 189 aligned sites in the alignment corresponding to the phylogeny in Supplementary Figure S4). However, the corresponding nodes were not supported by bootstrap analysis in either tree, meaning that relationships cannot be inferred from these nodes.

In contrast, the topology was exactly the same between the two trees for each of the groups considered robust by bootstrap analysis.

Please note that a correction has been made to the tree in Figure S4, as the version of the tree was not the final one, but a preliminary version by mistake.

- Comment 19:

The chromosomic locations of the paralogs in the *Venturia* genome and their concordance distribution in the tree allow for a finer description of their evolutionary history that may merit a deeper description and discussion.

Authors' response:

L539-557: The concordance between the chromosomal locations of some of the *V. canescens* calyx RhoGAPs and the relationships that could be inferred (or not) from the phylogenetic analyses has been developed.

—Evidence of positive selection in *L. bouleari* venom and *V. canescens* calyx RhoGAP sequences

- Comment 20:

In L452 and thereafter, I think the use of substitution instead of mutation would be more appropriate. Similarly, I am not sure that "negatively selected codons" is better than "codons under negative selection"

Authors' response:

The term "mutation" has been replaced with "substitution" where appropriate.

The terms "positively selected" and "negatively selected" were replaced by "under positive selection" and "under negative selection" " where appropriate.

- Comment 21:

Also regarding language, I would suggest to use throughout the manuscript the appropriate naming of orthologs and paralogs, as it may correspond (see for instance the use of "LbGAP and its homolog LbGAPy in *L. bouleari*" in L456.

Authors' response:

The term orthologs have been used throughout the manuscript instead of homologs where appropriate.

- Comment 22:

It is unclear whether the positive selection analyses have been performed using one single sequence per in-paralog gene, or instead the multiple instances of allelic sequences (again, this point may make no sense depending of the meaning of the dot numbering scheme).

Authors' response:

As explained above, the dot numbering (e.g., LbmGAP1.1, LbmGAP1.2, and LbmGAP1.3) was used for *Lepidopilina* venom RhoGAPs that share high amino acid sequence identity in the RhoGAP domain (greater than 70% identity), but were considered genes rather than alleles due to their locations in the genomic scaffolds. This numbering was chosen to facilitate the comparison of the *L. bouleari* venom RhoGAPs and their orthologs in *L. bouleari* ISy in the phylogenetic analysis. Thus, the positive selection analyses were performed using a single sequence for each of the paralogs.

- Comment 23:

The polarity of the cladogram depicted in fig5A does not match the tree in figS2. The inference of branches under selection is most likely sensitive to the polarity in the tree.

Authors' response:

We agree that the detection of branches (but also sites) under selection is sensitive to the topology of the phylogenetic tree. The phylogenetic trees in Figure 5A and in Supplementary Figure S2 differ in that only *L. bouleardi* venom RhoGAPs were included in the latter. Regarding the topology of the *L. bouleardi* venom RhoGAPs, only one difference was found between the two trees. Again, it concerns the position of the group consisting of LbmGAP4 and LbyGAP4. As explained above, the node grouping LbmGAP4 and LbyGAP4 with other *L. bouleardi* venom RhoGAPs was not robust according to the bootstrap analysis, so no relationship can be inferred from that node. Apart from the position of the group consisting of LbmGAP4 and LbyGAP4, no difference was observed in the topology of *L. bouleardi* venom RhoGAPs compared to Supplementary Figure S2. The observation that only the grouping of LbmGAP4 and LbyGAP4 with other *L. bouleardi* venom RhoGAPs differs between phylogenetic trees is not so surprising, since LbmGAP4 and LbyGAP4 are the most divergent among *L. bouleardi* venom RhoGAPs (according to Supplementary Figure S2 and pairwise sequence identity).

- Comment 24:

The possible functionality of VcVLP2 GAP may merit discussion, specially given its close evolutionary relationship with VcGAP9.

Authors' response:

L589-594: The possible functionality of VcVLP2 has been previously developed, but it has been further explored through a comparison with VcGAP9, considering their close relationship inferred from phylogenetic analyses.

[Download tracked changes file](#)

Decision by [Ignacio Bravo](#) , posted 17 October 2023, validated 17 October 2023

Two reviewers have evaluated the manuscript and provided valuable feedback. The two of them agree on the interest of the subject, the pertinence of the model and on the quality of the analyses, and I largely agree with them. Both of them, have notwithstanding raised a number of comments that may require to be revised before the text can be recommended in PCI Genomics, and I think these comments are relevant.

I have further identified a number of instances that may require to be revised, and that are listed below. Beyond these comments, I think the manuscript may benefit from a more thorough evolutionary analysis, proposing a timeline (or at least an order) for the duplication events described.

I also consider that a more global approach to the subject of gene duplication as a source of novelty may help conceptualise the findings. A reference to Ohno's works may also be needed.

I think also that performing some analyses at the nucleotide level may help interpret the mechanisms promoting functional diversity, identify possible instances of gene replacement, or provide keys for understanding the possible loss of function of the enzymes. In this regard, the literature regarding a role of nucleotide composition divergence of in-paralogs that retain a high amino acid identity may be relevant for discussion and interpretation.

—Phylogeny of members of the Cynipoidea superfamily

please include the fraction of the alignment that was removed with trimal, and specify the length and pattern content of the final alignment used for reconstruction. Please include a proper phylogenetic tree with branch lengths. For reproducibility, I suggest to deposit or to include in SM the final alignment used for phylogenetic inference.

—Phylogeny of *Leptopilina* venom and *Venturia* calyx RhoGAPs

Please provide accession numbers for proteins listed in Table S2.

It is unclear whether phylogenetic inference shown in Fig 1b was performed at the amino acid or at the codon-based nucleotide level.

For the sake of clarity, the abbreviation Lbm and Lby are used for the two *Lb* isogenic lines, but the sequences and branches do not include systematically the m or y indication, or it is included misleading, as in “LbGAPy4”, instead of “LbyGAP4”. The reader needs to resort to Table 1 for decoding.

It is never explained to what integration level the different sequences named with dot numbering (as in LbGAP1.1 and LbGAP1.2) refer to. I understand that they may be different alleles found among the females sequenced, but this needs to be explained. The authors need to explain/substantiate whether their analyses are sensitive enough to simultaneously address inter and intraspecific diversity, as the inclusion of deeper branches for interspecific analyses comes along with a loss of the exploitable alignment length for the intraspecific analyses.

The razor algorithm underlying the results in Figure 2B needs to be explained, as several protein bands are matched to different sequences, but still a differential identification for a higher abundance for some proteins in certain bands is provided.

The functional and/or evolutionary connection between the RhoGAP domain (PF00620) and the RacGAP1 domain (PTHR46199) needs to be explained and discussed. The latter is not described in the methods and appears directly on Tables 1 and 2.

In the version I have evaluated, Fig S2 is not accompanied by any legend. The tree inference is not described. It is unclear whether only the blue, shared domain, was used for inference, and at what level.

The topology in Fig 1b and in Fig S2 are different. Please discuss.

The correspondence (or the lack of) between the species and the genes trees for the *Leptopilina* RhoGAP sequences needs to be analysed and discussed. The authors may need to address possible instances of allele replacement or gene conversion to explain the topology described in Fig S2.

—*V. canescens* calyx RhoGAPs probably evolved from two or more imperfect RacGAP1 duplication events

The position in the tree of *Nasonia* RacGAP1 in Fig 3A and S4 as being very closely related to *Venturia* RacGAP12 is unexpected and needs to be discussed (unless this is a typo for VcRacGAP1)

The topologies in Fig 3A and S4 are discordant and need to be discussed.

The chromosomal locations of the paralogs in the *Venturia* genome and their concordance distribution in the tree allow for a finer description of their evolutionary history that may merit a deeper description and discussion.

—Evidence of positive selection in *L. bouvardi* venom and *V. canescens* calyx RhoGAP sequences

In L452 and thereafter, I think the use of substitution instead of mutation would be more appropriate. Similarly, I am not sure that “negatively selected codons” is better than “codons under negative selection”

Also regarding language, I would suggest to use throughout the manuscript the appropriate naming of orthologs and paralogs, as it may correspond (see for instance the use of “LbGAP and its homolog LbGAPy in *L. bouvardi*” in L456).

It is unclear whether the positive selection analyses have been performed using one single sequence per in-paralog gene, or instead the multiple instances of allelic sequences (again, this point may make no sense depending on the meaning of the dot numbering scheme).

The polarity of the cladogram depicted in Fig 5A does not match the tree in Fig S2. The inference of branches under selection is most likely sensitive to the polarity in the tree.

The possible functionality of VcVLP2 GAP may merit discussion, specially given its close evolutionary relationship with VcGAP9.

Reviewed by anonymous reviewer 1, 19 September 2023

The article explores the fact that two distantly related families of parasitoid wasps harbor similar RhoGAP proteins in the venom used to subjugate their egg hosts.

Authors demonstrated that a specific endophysiological gene was recruited through gene duplication in the ancestral of the genus *Leptopilina*, suffering a reassignment of the cell compartment through the incorporation of a novel signal peptide, and then interspecifically expanded into more paralogues. For the genus *Venturia*, the venom genes expanded to 13 copies, but the absence of further structural modification in the venom genes makes it less evident the difference between venom and classical RhoGAPs.

I am convinced about the convergence of having RhoGAP recruited to venom independently in both groups, since they are very distant. However, I missed more information about the situation in the sister taxa in both cases, what would be important to understand when the recruitment occurred and what the functional meaning. For example, in the sequence phylogenies of figures 1 and 3, a NvRacGAP from *N. vitripennis* is shown as the sister sequence to both *Leptopilina* and *V. canescens* RhoGAPs. As far as I understand, *Nasonia vitripennis* is distantly related to both families, so what happens in the several genera phylogenetically placed between them? Are there other genus close to *Leptopilina* and *Venturia* that could also share the venom protein, thus representing an earlier event of recruitment? Moreover, it is not clear for me, as a non-specialist in the field, if the other genera of these families are also parasitoid. If yes, do they produce venoms that act in the process? And if so, do these venoms have any sort of RhoGAP?

Perhaps, adding more background about the group would help to understand if the reported RhoGAP recruitments were conditions to develop a major phenotype (venom mediated parasitism) or were selected to further improve it.

Beyond the evolutionary questions, the authors did a terrific job in finding and validating the mutations of these proteins, further detailing biochemical aspects of the venom proteins.

The methodology used is overall appropriated and the article brings novelty to its field.

Reviewed by anonymous reviewer 2, 29 September 2023

In this paper, Colinet et al. studied the venom convergence of parasitoid wasps, focusing on the distinct origins of their venom proteins and the evolutionary trends individually within *Leptopilina* and *Venturia*. Several improvements could enhance the clarity, cohesion, and overall impact of the manuscript.

For better comprehension through side-by-side comparison, RhoGAPs in the *Leptopilina* and *Venturia* lineages should be compared in the same figure. This might necessitate a substantial reorganization of figures. Combining Fig. 1 and Fig. 3 would clearly show the distinct origin of their venom proteins and present their phylogenetic positions on a more sensible scale. A gene family-wide, unified phylogeny should be presented. Fig. 5 and Fig. 6 should also be combined to facilitate an easy comparison of how different/similar selected sites are in the two lineages.

Also, the description of the protein domain and exon structures for individual genes is somewhat lengthy. While detailed descriptions are appreciated, more concise explanations would facilitate understanding overall evolutionary trends. It is recommended that these detailed texts be moved to Supplemental Information and be made more concise in the main text.

L110: Clarification is needed on the "two multigene families" mentioned, as only one (RhoGAP) seems to be described.

L239: Please explain what β & ω symbolize.

L285: Clarify whether the 25 samples were processed separately or partly pooled.

L320: The term "completed" needs clarification. Does it indicate that the complete coding sequences were recovered using partial protein sequences?

L345: It would be enlightening to explore whether the exons encoding the signal peptide originated through de-novo gene birth or by exon shuffling with other genes. Is there any evidence of sequence similarity with other genes in the genome? Such information would be pertinent to the discussion on their origin, which is addressed later in the Discussion section.

L421: "Mascot 5" does not appear in the table.

L485: Providing the number of intersection, as well as union, of sites detected by different methods is suggested.

Fig. 1B, Fig. 3A, B: Rooting method should be described. This is applicable to other figures as well.

Table 3: Clarify what the parameter p provided along with ω is.

Fig. 5E, Fig. 6E: Since the site-wise relative solvent accessible surface can be computed, the current visualization method loses its quantitative information. I recommend exploring alternative forms of visualization to convey this data more effectively. This suggestion is also applicable to Fig. 6E.