Convergent origin and accelerated evolution of vesicle-associated RhoGAP proteins in two unrelated parasitoid wasps

Dominique Colinet, Fanny Cavigliasso, Matthieu Leobold, Apolline Pichon, Serge Urbach, Dominique Cazes, Marine Poullet, Maya Belghazi, Anne-Nathalie Volkoff, Jean-Michel Drezen, et al.

Convergent origin and accelerated evolution of vesicle-associated RhoGAP proteins in two unrelated parasitoid wasps

Dominique Colinet*[1], Fanny Cavigliasso[1], Matthieu Leobold[2], Appoline Pichon[3], Serge Urbach[3], Dominique Cazes[1], Marine Poullet[1], Maya Belghazi[4], Anne-Nathalie Volkoff[3], Jean-Michel Drezen[2], Jean-Luc Gatti[1], and Marylène Poirié[1]

[1] Université Côte d'Azur, INRAE, CNRS, ISA, 06903 Sophia Antipolis, France
[2] Institut de Recherche sur la Biologie de l'Insecte (IRBI), UMR 7261, CNRS - Université de Tours, Tours, France
[3] DGIMI, Univ Montpellier, INRAE, 34095 Montpellier, France
[4] Aix Marseille Université, CNRS, Plateforme Protéomique, IMM FR3479, Marseille Protéomique (MaP), 13402 Marseille, France

*Corresponding author
Correspondence: dominique.colinet@inrae.fr

ABSTRACT

Animal venoms and other protein-based secretions that perform a variety of functions, from predation to defense, are highly complex cocktails of bioactive compounds. Gene duplication, accompanied by modification of the expression and/or function of one of the duplicates under the action of positive selection, followed by further duplication to produce multigene families of toxins is a well-documented process in venomous animals. This evolutionary model has been less described in parasitoid wasps, which use maternal fluids, including venom, to protect their eggs from encapsulation by the host immune system. Here, we evidence the convergent recruitment and accelerated evolution of two multigene families of RhoGAPs presumably involved in virulence in two unrelated parasitoid wasp species, *Leptopilina boulardi* (Figitidae) and *Venturia canescens* (Icheumonidae). In both species, these RhoGAPs are associated with vesicles that act as transport systems to deliver virulence factors, but are produced in different tissues: the venom gland in *Leptopilina* sp. and the ovarian calyx in *V. canescens*. We show that the gene encoding the cellular RacGAP1 is at the origin of the virulent RhoGAP families found in *Leptopilina* sp. and *V. canescens*. We also show that both RhoGAP families have undergone evolution under positive selection and that almost all of these RhoGAPs lost their GAP activity and GTPase binding ability due to mutations in key amino acids. These results suggest an accelerated evolution and functional diversification of these vesicle-associated RhoGAPs in the two phylogenetically distant parasitoid species. The potential new function(s) and the exact mechanism of action of these proteins in host cells remain to be elucidated.

Keywords: parasitoids, venom and other secretions, RhoGAP multigene family, evolution, gene duplication, positive selection

Introduction

Gene duplication is recognized as an important evolutionary process because it drives functional novelty (Chen et al., 2013; Long et al., 2013). In particular, the formation of multigene families by repeated gene duplication is a widely studied evolutionary process in several groups of venomous animals (Fry et al., 2009; Wong & Belov 2012; Casewell et al., 2013), but also in some parasites (Akhter et al., 2012; Arisue et al., 2020). Such repeated duplication events of venomous or virulence protein-coding genes are often accompanied by significant copy divergence through positive selection, allowing the acquisition of new functions. In this work, we studied the process of accelerated evolution by duplication and divergence of two multigene families encoding RhoGAPs presumably involved in virulence in two phylogenetically distant parasitoid wasp species, *Leptopilina* species of the family Figitidae and *Venturia canescens* of the family Icheumonidae.

The development of parasitoid wasps occurs at the expense of another arthropod, whose tissues are consumed by the parasitoid larvae, usually resulting in the host death (Godfray 1994). For endoparasitoids that lay eggs inside the host body, the host immune defense against parasitoids is usually based on the formation of a multicellular melanized capsule around the parasitoid egg, resulting in its death (Carton et al., 2008). To escape encapsulation, parasitoids have evolved several strategies, the main one being the injection of maternal fluids together with the egg into the host at the time of oviposition (Pennacchio & Strand 2006; Poirié et al., 2009). These maternal fluids contain (i) proteins synthesized in the venom glands (Asgari & Rivers 2010; Poirié et al., 2014), some of which, as in *Leptopilina,* can be associated with extracellular vesicles called venosomes and allow their transport to the targeted immune cells (Gatti et al., 2012; Wan et al., 2019), (ii) proteins synthesized in the ovarian calyx cells and associated with particles of viral origin devoid of nucleic acid and named virus-like particles (VLPs) as in *V. canescens* (Reineke et al., 2006; Gatti et al., 2012; Pichon et al., 2015), and (iii) polydnaviruses (PDVs) integrated into the parasitoid genome and found in some groups of parasitoids belonging to the families Braconidae and Ichneumonidae (Drezen et al., 2014). Like VLPs, PDVs are synthesized in the ovarian calyx cells and are present in the ovarian fluid along with the egg. These particles are unique in that they are formed by the integrated viral machinery, but carry circular double-stranded DNA molecules that contain virulence genes that will be expressed in the host cells (Drezen et al., 2014). The process of accelerated evolution by duplication and divergence has been described for some of these virulence genes carried by PDVs, suggesting functional diversification of the proteins produced (Serbielle et al., 2008; Serbielle et al., 2012; Jancek et al., 2013).

Leptopilina boulardi is a parasitoid wasp of *Drosophila* for which two lines have been well characterized (Dubuffet et al., 2009). The ISm line of *L. boulardi* is highly virulent against *Drosophila melanogaster* but is unable to develop in *Drosophila yakuba* whereas the ISy line can develop in both *Drosophila* species but its success depends on the resistance/susceptibility of the host (Dubuffet et al., 2009). One of the most abundant proteins in the venom of the *L. boulardi* ISm line belongs to the Rho GTPase activating protein (RhoGAP) family and has been named LbGAP (Labrosse et al., 2005; Colinet et al., 2013). A LbGAP homolog named LbGAPy is also present in the venom of the *L. boulardi* ISy line, but in lower amounts (Colinet et al., 2010; Colinet et al., 2013). While RhoGAPs are usually intracellular proteins composed of several different domains (Tcherkezian & Lamarche-Vane 2007), LbGAP and LGAPy contain only one RhoGAP domain preceded by a signal peptide that allows its secretion. LbGAP has been shown to be associated with and transported by venosomes to target host hemocytes (Wan et al., 2019). LbGAP and LbGAPy specifically interact with and inactivate two *Drosophila* Rho GTPases, Rac1 and Rac2 (Colinet et al., 2007; Colinet et al., 2010), which are required for hemocyte proliferation in response to parasitism, hemocyte adhesion around the parasitoid egg, or the formation of intercellular junctions necessary for the capsule formation (Williams et al., 2005; Williams et al., 2006). Combined transcriptomic and proteomic analyses have identified eight additional RhoGAP domain-containing venom proteins in addition to LbGAP and LbGAPy for both *L. boulardi* lines, suggesting that a multigene family has derived from repeated duplication (Colinet et al., 2013). One of these venom RhoGAPs, LbGAP2, has also been shown to be associated with venosomes in *L. boulardi* ISm and to be released with LbGAP in host hemocytes (Wan et al., 2019). Interestingly, three RhoGAP domain-containing venom proteins have also been identified in the closely related species *Leptopilina heterotoma* (Colinet et al., 2013), suggesting that the recruitment of RhoGAPs in the venom arsenal may have occurred before the separation of the two species.

Leptopilina venom RhoGAPs are not the only example of the possible use of a protein from this family in parasitoid virulence (Reineke et al., 2006; Du et al., 2020). In the Ichneumonidae *V. canescens*, a RhoGAP domain-containing protein known as VLP2 (named VcVLP2 in this paper) was found to be associated with VLPs formed in the nucleus of ovarian calyx cells (Reineke et al., 2006; Pichon et al., 2015). VLPs, which package proteic virulence factors wrapped into viral envelopes, are then released into the ovarian lumen to associate with eggs and protect them from the host immune response by as yet unknown mechanisms (Feddersen et al., 1986). These observations suggest convergent recruitment of proteins belonging to the same family and injected into the host with the egg in two phylogenetically distant parasitoid species.

In this work, we showed that VcVLP2 in *V. canescens*, like LbGAP in *Leptopilina*, is not unique but is part of a multigene family of virulent RhoGAPs associated with extracellular vesicles. RhoGAPs from different organisms are grouped into distinct subfamilies based on similarities in RhoGAP domain sequence and overall multi-domain organization (Tcherkezian & Lamarche-Vane 2007). Our analyses indicate that an independent duplication of the RacGAP1 gene, a member of the large RhoGAP family, was the origin of the virulent RhoGAP multigene families found in *Leptopilina* and *V. canescens*. We then performed comparative analyses to understand the duplication events at the origin of these two multigene families. Finally, we demonstrated evolution under positive selection for both RhoGAP multigene families in *L. boulardi* and *V. canescens,* suggesting accelerated evolution and functional diversification.

Methods

Biological material

The origin of the *L. boulardi* isofemale lines ISm (Lbm; Gif stock number 431) and ISy (Lby; Gif stock number 486) has been described previously (Dupas et al., 1998). Briefly, the ISy and ISm founding females were collected in Brazzaville (Congo) and Nasrallah (Tunisia), respectively. The *L. heterotoma* strain (Lh; GIF stock number 548) was collected in southern France (Gotheron). The Japanese strain of *Leptopilina victoriae* (Lv), described in Novković et al., 2011, was provided by Pr. M. T. Kimura (Hokkaido University, Japan). All parasitoid lines were reared on a susceptible *D. melanogaster* Nasrallah strain (Gif stock number 1333) at 25 °C. After emergence, the wasps were maintained at 20 °C on agar medium supplemented with honey. All experiments were performed on 5- to 10-day-old parasitoid females.

Phylogeny of members of the Cynipoidea superfamily

The phylogeny of selected members of the Cynipoidea superfamily was constructed using internal transcribed spacer 2 (ITS2) sequences available at NCBI (Supplementary Table S1). Multiple sequence alignment was performed using MAFFT with the --auto option (Katoh & Standley 2013). Poorly aligned regions were removed using trimal with the -automated1 option (Capella-Gutierrez et al., 2009). Phylogenetic analysis was performed using maximum likelihood (ML) with IQ-TREE (Minh et al., 2020).

Search for candidate venom RhoGAPs in Leptopilina transcriptomes

L. victoriae venom apparatus, corresponding to venom glands and associated reservoirs, were dissected in Ringer's saline (KCl 182 mM; NaCl 46 mM; $CaCl_2$ 3 mM; Tris-HCl 10 mM). Total RNA was extracted from 100 venom apparatus using TRIzol reagent (Invitrogen) followed by RNeasy Plus Micro Kit (QIAGEN) according to the manufacturers' instructions. The quality of total RNA was checked using an Agilent BioAnalyzer. Illumina RNASeq sequencing (HiSeq 2000, 2 × 75 pb) and trimming were performed by Beckman Coulter Genomics. The raw data are available at NCBI under the BioProject ID PRJNA974978. Sequence assembly was performed using the "RNASeq de novo assembly and abundance estimation with Trinity and cdhit" workflow available on Galaxy at the BIPAA platform (https://bipaa.genouest.org).

Leptopilina clavipes (whole body), *Ganaspis* sp. G1 (female abdomen), *Andricus quercuscalicis* (whole body) and *Synergus umbraculus* (whole body) transcriptomes were obtained from NCBI under accession numbers GAXY00000000, GAIW00000000, GBNY00000000 and GBWA00000000, respectively. Coding regions in the transcripts were identified and translated using TransDecoder (Haas et al., 2013), which is available on the BIPAA Galaxy platform. Searches for RhoGAP domain-containing sequences in translated coding sequences were performed using hmmsearch from the HMMER package (Eddy 2009) with the RhoGAP (PF00620) HMM profile.

Completion of *L. boulardi* and *L. heterotoma* venom RhoGAP sequences

Venom apparatus, corresponding to venom glands and associated reservoirs, were dissected in Ringer's saline. Total RNA was extracted from venom apparatus using the RNeasy Plus Micro Kit (QIAGEN) according to the manufacturers' instructions. To obtain the full-length coding sequence of LbGAP1.3, LbGAP5, LbGAPy6, LhGAP1 and LhGAP2, rapid amplification of cDNA ends (RACE) was performed using the SMART RACE cDNA Amplification Kit (Clontech). For LhGAP3, 3' RACE could not be applied due to the presence of poly(A) stretches in the sequence. LhGAP3-matching sequences were searched for in the transcriptome assembly obtained by Goecks et al., 2013) from the abdomen of *L. heterotoma* females (GenBank accession number GAJC00000000) using the command line NCBI-BLAST package (version 2.2.24). The resulting complete LhGAP3 sequence was then verified by RT-PCR using the iScript cDNA Synthesis Kit (BioRad) and the GoTaq DNA Polymerase (Promega), followed by direct sequencing of the amplified fragment. Geneious software (Biomatters) was used for sequence editing and assembly.

Obtention of Leptopilina RacGAP1 coding sequences

A combination of RT-PCR and RACE using conserved and specific primers was used to clone the full-length ORF sequence of *L. boulardi* and *L. heterotoma* RacGAP1. The amino acid sequence of *Nasonia vitripennis* RacGAP1 (Supplementary Table S2) was used to search for similar sequences among Hymenoptera species using BLASTP at NCBI (http://www.ncbi.nlm.nih.gov/blast/). A multiple sequence alignment of the RacGAP1 amino acid sequences found in Hymenoptera was then performed using MUSCLE (Edgar 2004). Two pairs of RacGAP1-specific degenerate primers were designed from the identified conserved regions (Supplementary Figure S1). Total RNA was extracted from Lbm and Lh individuals using the TRIzol reagent (Invitrogen), and RT-PCR experiments were performed using the RacGAP1-specific degenerate primers. After direct sequencing of the amplified fragments, Lbm- and Lh-specific primers were designed to complete the sequences obtained by RT-PCR and RACE (Supplementary Figure S1). The coding sequences of *L. victoriae and L. clavipes* RacGAP1 were obtained by BLAST searches in the corresponding transcriptomes using *L. boulardi* and *L. heterotoma* RacGAP1 sequences as queries.

Obtention of the genomic sequences of Leptopilina venom RhoGAPs and RacGAP1

The genomic sequences of *Leptopilina* venom RhoGAPs and RacGAP1 were obtained by BLAST and Exonerate searches (Slater & Birney 2005) using the corresponding coding sequences as queries. *L. boulardi* ISm, *L. boulardi* ISy, *L. heterotoma* and *L. clavipes* genome assemblies were obtained from NCBI under accession numbers GCA_011634795.1, GCA_019393585.1, GCF_015476425.1 and GCA_001855655.1 respectively.

Search for candidate *V. canescens* RhoGAP calyx sequences

Coding regions were identified from the *V. canescens* calyx transcriptome published by Pichon et al., *(*2015) and translated into protein sequences using TransDecoder (Haas et al., 2013). The search for RhoGAP domain-containing sequences was performed using hmmsearch from the HMMER package (Eddy 2009) with the RhoGAP (PF00620) HMM profile. The identified RhoGAP calyx sequences were used as a database for Mascot to explore MS/MS data obtained from purified *V. canescens* VLPs (Pichon et al., 2015) with a false discovery rate of 1%. Only proteins identified by two or more peptides were considered.

The *V. canescens* RacGAP1 coding sequence was obtained by BLAST searches in the *V. canescens* calyx transcriptome using the *N. vitripennis* RacGAP1 sequence as query.

The *V. canescens* calyx RhoGAP and RacGAP1 genome sequences were obtained by BLAST and Exonerate searches (Slater and Birney 2005) using the corresponding coding sequences as queries. The *V. canescens* genome assembly was obtained from NCBI under accession number GCA_019457755.1.

Identification of domains and motifs and prediction of subcellular localization

The presence and position of signal peptide cleavage sites in the identified RhoGAP domain-containing sequences were predicted using the SignalP server (https://services.healthtech.dtu.dk/service.php?SignalP). The searches for domains and motifs were performed using InterProScan 5 (Jones et al., 2014). Coiled-coil regions were predicted using COILS (Lucas et al., 1991).

Prediction of protein subcellular localization and sorting signals for *Venturia* RacGAP1 and calyx RhoGAPs was performed using the DeepLoc2 server (https://services.healthtech.dtu.dk/services/DeepLoc-2.0/).

Nuclear localization for VcRacGAP1 and calyx RhoGAPs was also predicted using the following three tools: NucPred (Brameier et al., 2007) was used to indicate whether a protein spends part of its time in the nucleus (https://nucpred.bioinfo.se/nucpred/), LocTree3 (Goldberg et al., 2014) to provide subcellular localization and Gene Ontology terms (https://rostlab.org/services/loctree3/), and PSORT II (Nakai & Horton 1999) to detect sorting signals and subcellular localization (https://psort.hgc.jp/form2.html). PSORT II results corresponding to ER membrane retention signals and cleavage sites for mitochondrial presequences were included as they may be associated with nuclear localization. Mitochondrial presequences have no sequence homology but possess physical characteristics that allow them to interact with the outer membranes of mitochondria and thus allow the targeting of proteins to the mitochondrial matrix. Regarding mitochondrial targeting, it has been shown from plants to humans that some proteins with mitochondrial presequences are dually targeted to mitochondria and the nucleus (Millar et al., 2006; Mueller et al., 2004). Indeed, anchoring signals to the ER membrane are known to allow baculoviral proteins to migrate to the inner nuclear membrane (Braunagel et al., 1996). Proteins forming VLPs in *V. canescens* could thus follow the same pathways as their homologous proteins in baculoviruses, nudiviruses, hytrosaviruses, and bracoviruses due to the conservation of this mechanism among these viruses belonging to the order Lefavirales (Braunagel et al., 1996; Hong et al., 1997; Braunagel et al., 2004; Abd-Alla et al., 2008; Bézier et al., 2009; Braunagel et al., 2009).

Phylogeny of Leptopilina venom and Venturia calyx RhoGAPs

Searches for *N. vitripennis* RhoGAP sequences were performed using BLASTP at NCBI (http://www.ncbi.nlm.nih.gov/blast/) and hmmsearch from the HMMER package (Eddy 2009) with the RhoGAP (PF00620) HMM profile on the *N. vitripennis* v1.2 proteome database (Rago et al., 2016).

Multiple amino acid sequence alignments were performed using MAFFT with the --auto option (Katoh and Standley 2013). For codon-based analysis of selection (see below), codon-based alignments of complete coding sequences were performed using RevTrans (https://services.healthtech.dtu.dk/service.php?RevTrans) with the amino acid alignments as templates. Poorly aligned regions were removed using trimAl with the -automated1 option (Capella-Gutierrez et al., 2009). Phylogenetic analyses were performed using maximum likelihood (ML) with IQ-TREE (Minh et al., 2020). ModelFinder was used to select the best model selection for phylogeny (Kalyaanamoorthy et al., 2017).

Codon-based analysis of selection

Codon-based alignments of complete coding sequences were performed with RevTrans using the amino acid alignments as templates (see above). Phylogenetic analyses were performed using maximum likelihood (ML) with PhyML (Guindon et al., 2010). Smart Model Selection was used to choose the best model selection for the phylogeny (Lefort et al., 2017).

To detect evolutionary selective pressures acting on RhoGAP sequences, the ratios of non-synonymous substitutions (dN) to synonymous substitutions (dS) were compared using different ML frameworks: the CODEML program in the PAML (Phylogenetic Analysis by Maximum Likelihood) package (Yang 1997; Yang 2007), the HyPhy software implemented at http://www.datamonkey.org/ (Delport et al., 2010), and the Selecton server (Stern et al., 2007) available at http://selecton.tau.ac.il/index.html.

Five different methods were used to detect codons under positive selection. In the first, codon substitution models implemented in CodeML were applied to the codon-based alignment using the F3x4 frequency model. Two pairs of site models were used to determine whether some codons were under positive selection: M1a (neutral) versus M2a (selection) and M7 (ß) versus M8 (ß & ω). The models were compared using a likelihood ratio test (LRT) with 2 degrees of freedom to assess the significance of detection of selection (Yang et al., 2000). Bayes Empirical Bayes (BEB) inference was then used to identify amino acid sites with a posterior probability >95% of being under positive selection (Yang et al., 2005). We then applied four other methods of detection of selection available in the HyPhy package: Single Likelihood Ancestor Counting (SLAC), Fixed Effect Likelihood (FEL), Mixed Effects Model of Evolution (MEME) and Fast Unbiased Bayesian AppRoximation (FUBAR) methods (Kosakovsky Pond & Frost 2005; Murrell et al., 2012;

Murrell et al., 2013). To eliminate false-positive detection, only codons identified by CodeML M2a and M8 and at least one of the other methods were considered under positive selection. Radical or conservative replacements were then determined based on whether they involved a change in the physicochemical properties of a given amino acid, such as charge or polarity (Zhang 2000).

Four different methods were used to detect codons under negative selection: the Selecton server using the pair of site models M8a (ß& ω=1) versus M8 (ß& ω), and the FEL, SLAC and FUBAR methods mentioned above. Only codons identified by Selecton and at least one other method were considered under purifying selection.

To identify specific lineages with a proportion of sites evolving under positive or purifying selection, we performed branch-site REL analyses using the HyPhy package (Kosakovsky Pond et al., 2011). Unlike the branch and branch-site lineage-specific models available in CodeML, branch-site REL does not require *a priori* identification of foreground and background branches.

Structural analysis

Molecular modeling of VcVLP2 RhoGAP was performed using the Phyre server with default parameters (http://www.sbg.bio.ic.ac.uk/phyre2/) (Kelley et al., 2015). The model with the highest confidence (100%) and coverage (46%) was obtained using the crystal structure with PDB code 2OVJ as a template. Model quality was evaluated using the QMEAN server (http://swissmodel.expasy.org/qmean/) (Benkert et al., 2011). Briefly, the QMEAN score is a global reliability score with values ranging between 0 (lower accuracy) and 1 (higher accuracy). The associated Z-score relates this QMEAN score to the scores of a non-redundant set of high-resolution X-ray structures of similar size, with ideal values close to 0. Visualization of LbGAP (Colinet et al., 2007) and VcVLP2 structures and mapping of sites under selection were performed using PyMol (http://sourceforge.net/projects/pymol/). Secondary structure assignment was performed with the DSSP program (Kabsch & Sander 1983). Accessible surface area (ASA) or solvent accessibility of amino acids was predicted using the *ASAView* algorithm (Ahmad et al., 2004).

***In vitro* mutagenesis**

The S76K, V124K, L137D, T143L, S150L, I185K, E200G, and V203R mutations were introduced into the LbGAP cDNA using the QuickChange XL Site-Directed Mutagenesis Kit (Stratagene). The results of *in vitro* mutagenesis were verified by sequencing.

Yeast two-hybrid analysis

Interactions between LbGAP and LbGAP mutants and mutated forms of *Drosophila* RhoA, Rac1, Rac2 and Cdc42 GTPases were individually examined by mating as previously described (Colinet et al., 2007). The plasmids expressing the GTPase proteins were tested against the pGADT7-T control vector, which encodes a fusion between the GAL4 activation domain and the SV40 large T-antigen. Reciprocally, plasmids producing LbGAP and LbGAP mutants were tested against the pLex-Lamin control vector. Interactions between LbGAP and the Rac1 and Rac2 GTPases were used as positive controls (Colinet et al., 2007). Interactions were initially tested by spotting five-fold serial dilutions of cells on minimal medium lacking histidine and supplemented with 3-amino-triazole at 0.5 mM to reduce the number of false positives. ß-galactosidase activity was then detected on plates (Fromont-Racine et al., 1997).

Purification of Leptopilina venosomes and mass spectrometry analysis

Twenty-five Lbm, Lby and Lh female venom apparatus were dissected, and the reservoirs were separated from the gland and opened to release the venom content in 25 µl of Ringer's solution supplemented with a protease inhibitor cocktail (Sigma). The venom suspension was centrifuged at 500xg for 5 min to remove the residual tissues, then centrifuged at 15,000xg for 10 min to separate the vesicular fraction from the soluble venom proteins (supernatant fraction) (Wan et al., 2019). The vesicular fraction was then washed twice by resuspension in 25 µl of Ringer's solution followed by centrifugation at 15,000xg for 10 min. The two samples were mixed with 4x Laemmli buffer containing ß-mercaptoethanol (v/v) and boiled for 5 min. Proteins were then separated on a 6–16% linear gradient SDS-PAGE and the gel was silver stained as previously described (Colinet et al., 2013). Identification of proteins by nano-LC-tandem mass spectrometry (MS/MS) was performed on bands excised from the gels as previously described (Colinet et al., 2013). MS/MS data analysis was performed with Mascot software (http://www.matrixscience.com),

licensed in house using the full-length coding sequences of the *Leptopilina* venom RhoGAP sequences. Data validation criteria were (i) one peptide with individual ion score greater than 50 or (ii) at least two peptides of individual ion score greater than 20. The mascot score is calculated as -10Log (p). The calculated FDR (based on an automated decoy database search) was less than 1%. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2022) partner repository with the dataset identifier PXD041695.

Results

Leptopilina venom RhoGAPs probably emerged in the ancestor of the genus

In the Figitidae, venom RhoGAPs have only been described for *L. boulardi* and *L. heterotoma* (Labrosse et al., 2005; Colinet et al., 2013). A search for candidate homologs was performed in the transcriptomes of *L. victoriae* and *L. clavipes*, two other *Leptopilina* species, *Gasnaspis* sp. G1, another member of the family Figitidae, and *Andricus quercuscalicis* and *Synergus umbraculus,* two gall wasp species representatives of the superfamily Cynipoidea outside of Figitidae (separated by approximately 50 Mya (Peters et al., 2017)). The putative venom RhoGAPs were identified based on the presence of a signal peptide at the N-terminus of the protein, followed by a RhoGAP domain (Table 1). Two transcript sequences encoding RhoGAP domain-containing proteins predicted to be secreted were found in *L. victoriae* (LvGAP1 and LvGAP2) and one in *L. clavipes* (LcGAP1), whereas none were found for *Gasnaspis* sp. G1, *A. quercuscalicis* and *S. umbraculus* (Table 1 and Figure 1A). Our results therefore suggest that venom RhoGAPs are specific to the genus *Leptopilina* although further taxon sampling would be required to fully support this hypothesis.

Leptopilina venom RhoGAPs likely evolved from a single imperfect duplication of RacGAP1

A phylogenetic analysis was performed to learn more about the evolutionary history of *Leptopilina* venom RhoGAPs. The phylogeny was constructed based on the amino acid sequence of the RhoGAP domain of these venom proteins and that of the 19 classical RhoGAPs we identified from the predicted proteome of the jewel wasp *Nasonia vitripennis* (Supplementary Table S2). Prior to analysis, the coding sequence was completed for two Lbm (LbGAP1.3 and LbGAP5), one Lby (LbGAPy6) and all three Lh venom RhoGAPs. Two of the *Leptopilina* venom RhoGAPs, namely LhGAP3 and LcGAP1, were predicted to contain two RhoGAP domains instead of one. Therefore, these two domains were separated for the analysis. The resulting phylogeny identified NvRacGAP1 as the closest RhoGAP from *N. vitripennis* to all *Leptopilina* venom RhoGAPs with confident support values (Figure 1B). Accordingly, the RhoGAP domain found in all *Leptopilina* venom RhoGAPs was predicted to belong to the MgcRacGAP subfamily typically found in RacGAP1-related proteins (Table 1).

The RacGAP1 coding sequence of *L. boulardi, L. heterotoma, L. victoriae and L. clavipes* was obtained either by cloning and sequencing or by searching genomic data, and its domain organization was compared with that of venom RhoGAPs (Figure 1C, Table 1 and Supplementary Figure S2). *Leptopilina* RacGAP1 has a typical domain organization consisting of a coiled-coil region and a C1 motif (protein kinase C-like zinc finger motif, a ~ 50 amino-acid cysteine-rich domain involved in phorbol ester/diacylglycerol and Zn^{2+} binding) followed by a RhoGAP domain (Tcherkezian & Lamarche-Vane 2007). A large part of the RacGAP1 sequence spanning the coiled-coil region and the C1 motif is replaced by the signal peptide in all *Leptopilina* venom RhoGAPs (Figure 1C, Table 1 and Supplementary Figure S2). This suggests that a unique incomplete duplication of the RacGAP1 gene occurred in the ancestor of the *Leptopilina* genus, resulting in a truncated duplicate encoding a RhoGAP with an altered N-terminus.

To further investigate this duplication event and the possible origin of the signal peptide, the genomic sequences of *L. boulardi*, *L. heterotoma* and *L. clavipes* venom RhoGAPs and RACGAP1 were obtained, either by cloning and sequencing or by searching in genomic data, and compared (Supplementary Table S2). The genomic sequence of *Leptopilina* RacGAP1 consists of 10 exons with a “CC region”, “C1 motif” and “RhoGAP domain” encoded by exons 2 and 3, exons 5 and 6 and exons 7 and 8, respectively. The RhoGAP domain is also encoded by two exons for venom RhoGAPs. These two exons are preceded by one to three exons, with the signal peptide encoded by the first exon for most of them. Sequence comparisons revealed similarities between venom RhoGAPs and *Leptopilina* RacGAP1 only for the sequence spanning the two exons encoding the RhoGAP domain and part of the following one. No significant similarities were found for the preceding exons at either the nucleotide or amino acid level (Supplementary Figure S3). This

supports the hypothesis of a partial duplication of the RacGAP1 gene in the ancestor of the *Leptopilina* genus, resulting in the loss of the sequence spanning exons 1 to 6, followed by further duplication of the ancestrally duplicated gene during the diversification of the genus *Leptopilina*.

The two consecutive RhoGAP domains found in LhGAP3 were grouped together with confident support values in the phylogenies (Figure 1B and Supplementary Figure S2). This suggests that a partial tandem duplication spanning exons 2 and 3 has occurred for this *L. heterotoma* venom RhoGAP. In contrast, the two RhoGAP domains found in LcGAP1 did not group together (Figure 1B and Supplementary Figure S2). However, the genomic sequence of LcGAP1 was not complete and we could not find the region corresponding to the second RhoGAP domain in the *L. clavipes* genome (Supplementary Table S3). Therefore, it is unclear whether the LcGAP1 coding sequence found in the transcriptome assembly is true or artifactual.

Table 1. Motif and domain organization of *Leptopilina* venom RhoGAPs. The signal peptide was predicted using SignalP at CBS. The RhoGAP domain (PF00620) and the RacGAP1 domain (PTHR46199) were identified using InterProScan. Two successive RhoGAP domains were found in LhGAP3 and LcGAP1. Lbm: *L. boulardi* ISm ; Lby: *L. boulardi* ISy ; Lh: *L. heterotoma* ; Lc: *L. clavipes* ; Lv: *L. victoriae.*

			Signal peptide		RhoGAP domain (PF00620)				RacGAP1 domain (PTHR46199)			
		Total length	From	To	From	To	Length	E-value	From	To	Length	E-value
Lbm	LbGAP	282	1	20	52	193	142	2.9e-32	38	267	230	7.0e-37
	LbGAP1.1	292	1	23	58	205	148	1.0e-19	42	255	214	7.0e-30
	LbGAP1.2	280	1	23	58	205	148	1.1e-19	42	255	214	1.3e-29
	LbGAP1.3	287	1	23	58	205	148	7.2e-17	42	253	212	1.3e-27
	LbGAP2.1	251	1	23	44	183	140	9.9e-26	26	234	209	5.5e-37
	LbGAP2.2	263	1	23	56	196	141	6.6e-26	49	246	198	1.3e-34
	LbGAP3	295	1	23	56	175	120	3.2e-12	42	266	225	1.2e-27
	LbGAP4	246	1	20	42	178	137	9.1e-11	38	218	181	5.6e-18
	LbGAP5	275	1	20	60	207	148	6.3e-23	41	251	211	2.4e-33
Lby	LbGAPy	286	1	20	52	193	142	1.6e-32	38	267	230	3.4e-37
	LbGAPy1.1	287	1	23	58	205	148	5.6e-17	42	255	214	9.9e-28
	LbGAPy1.2	325	1	23	56	203	148	2.9e-17	42	293	252	1.7e-26
	LbGAPy2	256	1	23	44	185	142	8.3e-27	26	234	209	1.5e-37
	LbGAPy3.1	281	1	23	56	175	120	1.8e-11	42	237	196	1.5e-25
	LbGAPy3.2	266	1	23	41	160	120	5.8e-12	28	222	195	6.4e-25
	LbGAPy4	274	1	20	42	177	136	7.4e-09	30	218	189	2.1e-17
	LbGAPy5	275	1	23	60	207	148	4.4e-24	41	251	211	1.2e-35
	LbGAPy6	239	1	20	53	194	142	8.2e-25	49	231	183	9.4e-34
Lh	LhGAP1	293	1	20	43	180	138	5.5e-36	41	289	249	1.0e-38
	LhGAP2	336	1	20	43	181	139	1.5e-29	45	275	231	8.2e-31
	LhGAP3	467	1	20	35	174	140	1.4e-22	34	221	188	1.3e-38
					257	396	140	2.2e-17				
Lc	LcGAP1	497	1	20	43	184	142	3.2e-30				
					249	384	136	3.5e-28	247	478	232	8.46e-71
Lv	LvGAP1	370	1	20	38	179	142	2.4e-30				
	LvGAP2	266	1	20	38	181	144	8.5e-23	33	224	192	3.7e-28

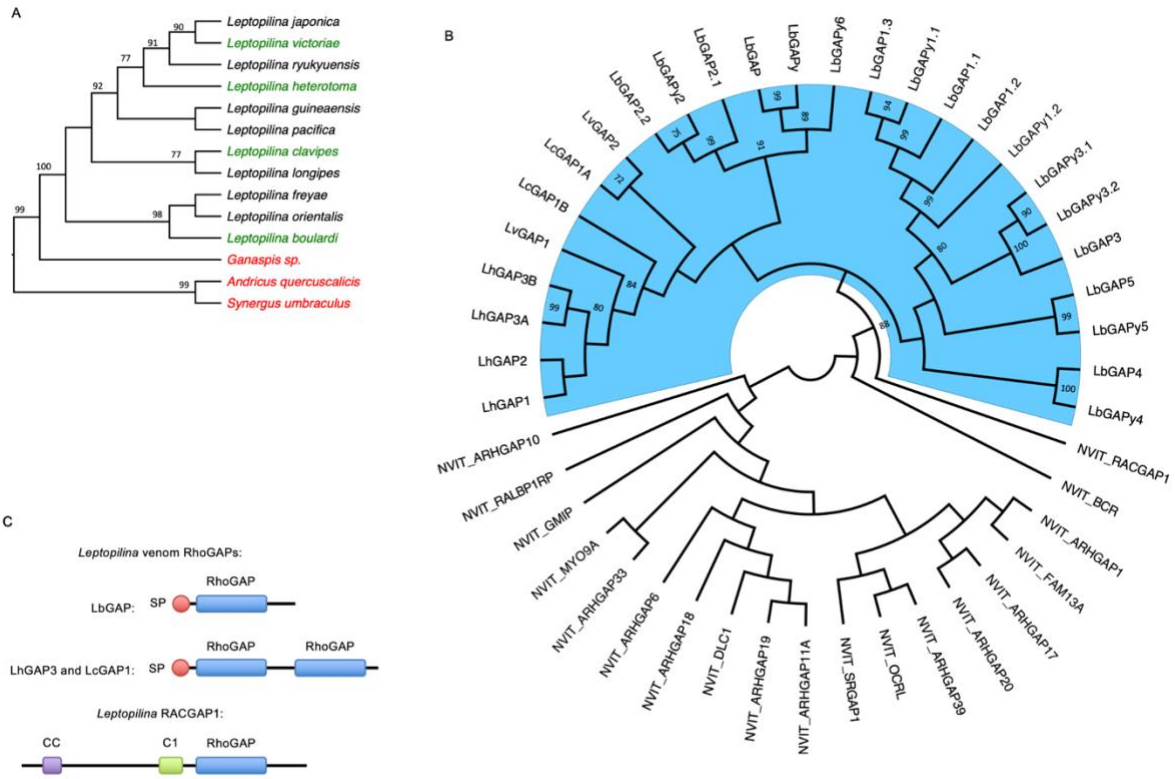


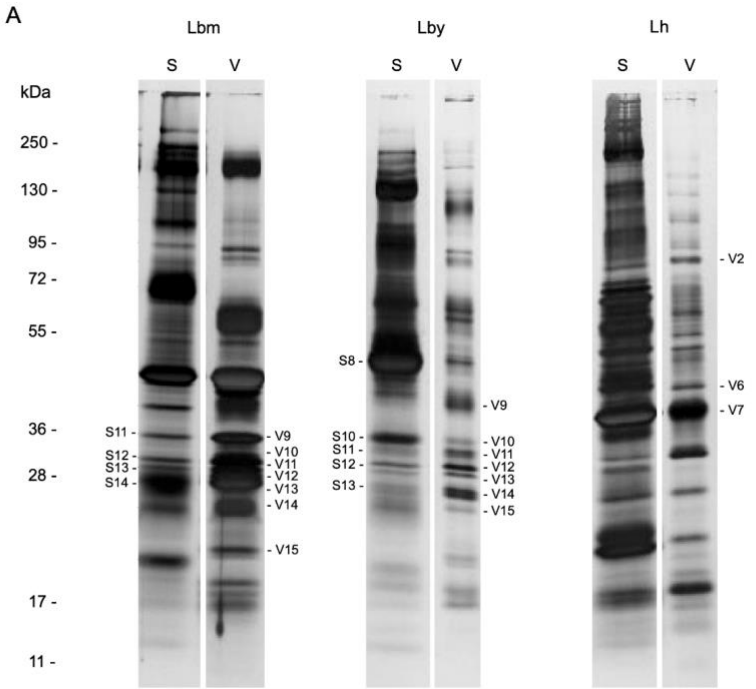
Figure 1. Origin of *Leptopilina* venom RhoGAPs (A) Maximum-likelihood phylogenetic tree of selected members of the Cynipoidea superfamily. The phylogenetic tree was obtained with IQ-TREE using internal transcribed spacer 2 (ITS2) sequence and displayed as a cladogram. Species in red are those for which venom RhoGAPs were not found. No data were available for species in black. (B) Maximum-likelihood phylogenetic tree of *Leptopilina* venom RhoGAPs along with *N. vitripennis* classical RhoGAPs. The phylogenetic tree was obtained with IQ-TREE using the RhoGAP domain amino acid sequence and displayed as a cladogram. For A and B, numbers at corresponding nodes are bootstrap support values in percent (500 bootstrap replicates). Only bootstrap values greater than 70% are shown. *Leptopilina* venom RhoGAPs are highlighted in blue. (C) Comparison of protein domain organization between LbGAP, representative of most *Leptopilina* venom RhoGAPs, LhGAP3 and LcGAP1, the two venom RhoGAPs that contain two RhoGAP domains, and *Leptopilina* RacGAP1. C1: protein kinase C-like zinc finger motif, CC: coiled-coil region, MgcRacGAP: RhoGAP domain found in RacGAP1-related proteins, SP: signal peptide.

L. boulardi and *L. heterotoma* venom RhoGAPs are associated with venosomes

LbGAP and LbGAP2 have been shown to be associated with vesicles named venosomes produced in *L. boulardi* venom that transport them to *Drosophila* lamellocytes (Wan et al., 2019). Our next goal was to investigate whether *Leptopilina* venom RhoGAPs other than LbGAP and LbGAP2 could be associated with venosomes. Proteomic analysis was performed on the supernatant and vesicular fractions separated from the venom by centrifugation. Comparison of the electrophoretic profiles obtained on a 6-16% SDS-PAGE for *L. boulardi* ISm (Lbm), *L. boulardi* ISy (Lby) and *L. heterotoma* (Lh) revealed an important variation between both fractions for all three wasps (Figure 2). All the major bands on the electrophoretic patterns of Lbm, Lby and Lh supernatant and vesicular fractions, as well as several minor bands (35 bands in total for Lbm, 34 for Lby and 37 for Lh), were excised and tryptic peptides were analyzed by mass spectrometry. The coding sequences of *Leptopilina* venom RhoGAPs were used to perform Mascot searches on the mass spectrometry data obtained from both venom fractions. All *L. boulardi* and *L. heterotoma* venom RhoGAPs were detected in the vesicular fraction, where most of them were found to be enriched (Figure 2). We

could then hypothesize that they are associated with and transported by venosomes to target *Drosophila* cells.

A



B

				Supernatant		Vesicular	
		Length	Predicted Mw	Protein band	Mascot Matches	Protein band	Mascot Matches
Lbm	LbGAP	282	32.13	**S11**/S12	11	**V9**/V10/V11/V12/ V13/V14/V15	57
	LbGAP1.1	292	33.02	S11/**S12**/S14	14	V9/**V10**/V11/V12/ V14/V15	85
	LbGAP1.2	280	31.99	S13	3	**V10**/**V11**/V12/V13	36
	LbGAP1.3	287	32.80	S12/**S13**/S14	12	V9/V10/**V11**/V12/ V13/V14/V15	85
	LbGAP2.1	251	28.81	S11/S12/S13/**S14**	14	V10/V11/V12/ **V13**/V14/V15	60
	LbGAP2.2	263	30.18	S12/S13/**S14**	9	V9/V10/V11/**V12**/ V13/V14/V15	85
	LbGAP3	295	34.01	**S12**/S13	3	**V10**/V11/V12/V13	15
	LbGAP4	246	28.57			V13/**V14**	3
	LbGAP5	275	31.72			**V10**/V11	10
Lby	LbGAPy	286	32.56	**S8**	1	**V10**/V11	4
	LbGAPy1.1	287	32.81	S10/S11/**S12**/S13	19	V11/**V12**/V13	19
	LbGAPy1.2	325	37.47			**V13**	1
	LbGAPy2	256	29.37	**S8**/S13	8	V12/V13/**V14**	8
	LbGAPy3.1	281	32.58	**S11**	1	V10/**V11**/V12/V13	5
	LbGAPy3.2	266	30.74	**S11**	1	V10/**V11**/V12/V13	5
	LbGAPy4	274	31.92	S10/**S11**/S12/S13	9	V9/V10/**V11**/V12/ V13/V14/V15	36
	LbGAPy5	275	32.04			**V12**	1
	LbGAPy6	239	27.83			**V15**	3
Lh	LhGAP1	293	33.16			**V7**	13
	LhGAP2	336	38.23			**V6**	4
	LhGAP3	467	53.92			**V2**	7

Figure 2. Proteomic analysis of *Leptopilina* venom supernatant and vesicular fractions. (A) The supernatant (S) and vesicular (V) fractions obtained from 25 Lbm, Lby and Lh venom reservoirs were separated on a 6-16% SDS-PAGE under reducing conditions and visualized by silver staining. Protein bands in which *Leptopilina* venom RhoGAPs were identified by mass spectrometry are numbered on the gel. Molecular weight standard positions are indicated on the left (kDa). (B) For each *Leptopilina* venom RhoGAP, length (number of amino acids), predicted Mw (kDa), bands in which specific peptides were found by mass spectrometry and number of total Mascot matches according to the S and V venom fractions are given. Numbers in bold correspond to the bands in which each RhoGAP was identified as the most abundant according to the number of Mascot matches.

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

Analysis of the *V. canescens* calyx transcriptome allowed the identification of a total of 13 RhoGAP domain-containing protein coding sequences, including VcVLP2 (Table 2). Matches with peptides obtained from a proteomic analysis of *V. canescens* VLPs (Pichon et al., 2015) were detected for all 13 calyx RhoGAPs, indicating that they are associated with VLPs (Table 2).

Similar to the *Leptopilina* venom RhoGAPs, the *V. canescens* calyx RhoGAPs contain a MgcRacGAP1 domain (Table 2) and are closely related to NvRacGAP1 (Figure 3A). In contrast to the *Leptopilina* venom RhoGAPs, none of the *V. canescens* calyx proteins were predicted to be secreted. Most were predicted to be localized in the nucleus and/or to contain a nuclear localization signal (Supplementary Tables S4 and S5). The absence of a predicted nuclear localization signal for VcGAP7 may be explained by the incomplete coding sequence at the 5' end (Table 2). There is greater variation in domain organization for *V. canescens* calyx RhoGAPs compared to *Leptopilina* venom RhoGAPs (Figure 3B and Supplementary Figure S4). Eight *V. canescens* calyx RhoGAPs have retained the C1 motif. In addition, VcGAP12 still contains a CC region in the N-terminal part of the sequence although the total sequence length is shorter than RacGAP1 (Figure 3B and Supplementary Figure S4).

Table 2. Motif and domain organization of *Venturia* calyx RhoGAP proteins associated with VLPs. The coiled-coil (CC) region was predicted using COILS (Lucas et al., 1991). The protein kinase C-like zinc finger (C1) motif (PS50081), the RhoGAP domain (PF00620) and the RacGAP1 domain (PTHR46199) were identified using InterProScan. Mascot 5: number of matches with peptides from purified *V. canescens* VLPs. [a]Coding sequence not complete at 5' end. [b]Coding sequence not complete at 3' end.

			CC		C1 (PS50081)		RhoGAP domain (PF00620)				RacGAP1 domain (PTHR46199)			
	Mascot	Total length	From	To	From	To	From	To	Length	E-value	From	To	Length	E-value
VcVLP2	26	485			56	106	134	277	144	7.4e-36	28	366	339	3.1e-98
VcGAP1	9	207					30	168	139	5.7e-25	12	199	188	5.1e-49
VcGAP2	19	280					23	161	139	6.9e-29	8	195	188	1.5e-50
VcGAP3	4	270			22	69	92	231	140	2.6e-25	14	265	252	2.9e-63
VcGAP4	21	382			37	86	116	240	125	2.2e-16	20	298	279	2.7e-68
VcGAP5	19	307			15	63	91	236	146	4.0e-24	7	281	275	2.6e-60
VcGAP6	16	284			17	67	88	215	128	3.3e-19	14	270	257	8.0e-55
VcGAP7	14	225[a]					15	156	142	5.3e-29	4	212	209	9.2e-57
VcGAP8	5	369			64	113	141	281	141	4.6e-28	20	334	315	3.6e-84
VcGAP9	9	319[b]			54	104	123	266	144	1.1e-28	38	306	269	5.5e-73
VcGAP10	5	233					30	166	137	1.6e-22	7	219	213	7.1e-45
VcGAP11	5	199					27	166	140	4.3e-19	8	198	191	5.7e-42
VcGAP12	8	402	21	48	79	131	159	296	138	3.3e-35	54	372	319	9.7e-91

The genomic coding sequence of VcRacGAP1 was obtained by genomic data mining. It comprised 9 exons with the CC region, C1 motif and RhoGAP domain encoded by exons 2 and 3, exons 5 and 6 and exon 7, respectively (Supplementary Table S6). In contrast to *Leptopilina*, the RhoGAP domain is encoded by a single exon in *V. canescens* RacGAP1 as well as in all calyx RhoGAPs. Interestingly, the VcGAP12 gene is located close to the VcRacGAP1 gene on chromosome 3, but is three exons shorter, suggesting a recent incomplete duplication (Supplementary Table S6). In contrast, VcVLP2 and the other VcGAP genes are more widely distributed in the *V. canescens* genome (Supplementary Table S6). It is therefore possible that incomplete duplication events of the VcRacGAP1 gene, followed by further tandem and dispersed

duplication of the ancestrally duplicated genes, led to the current number of calyx RhoGAPs in *V. canescens* organized in several clusters comprising 2 or more genes in tandem arrays (Figure 3C).

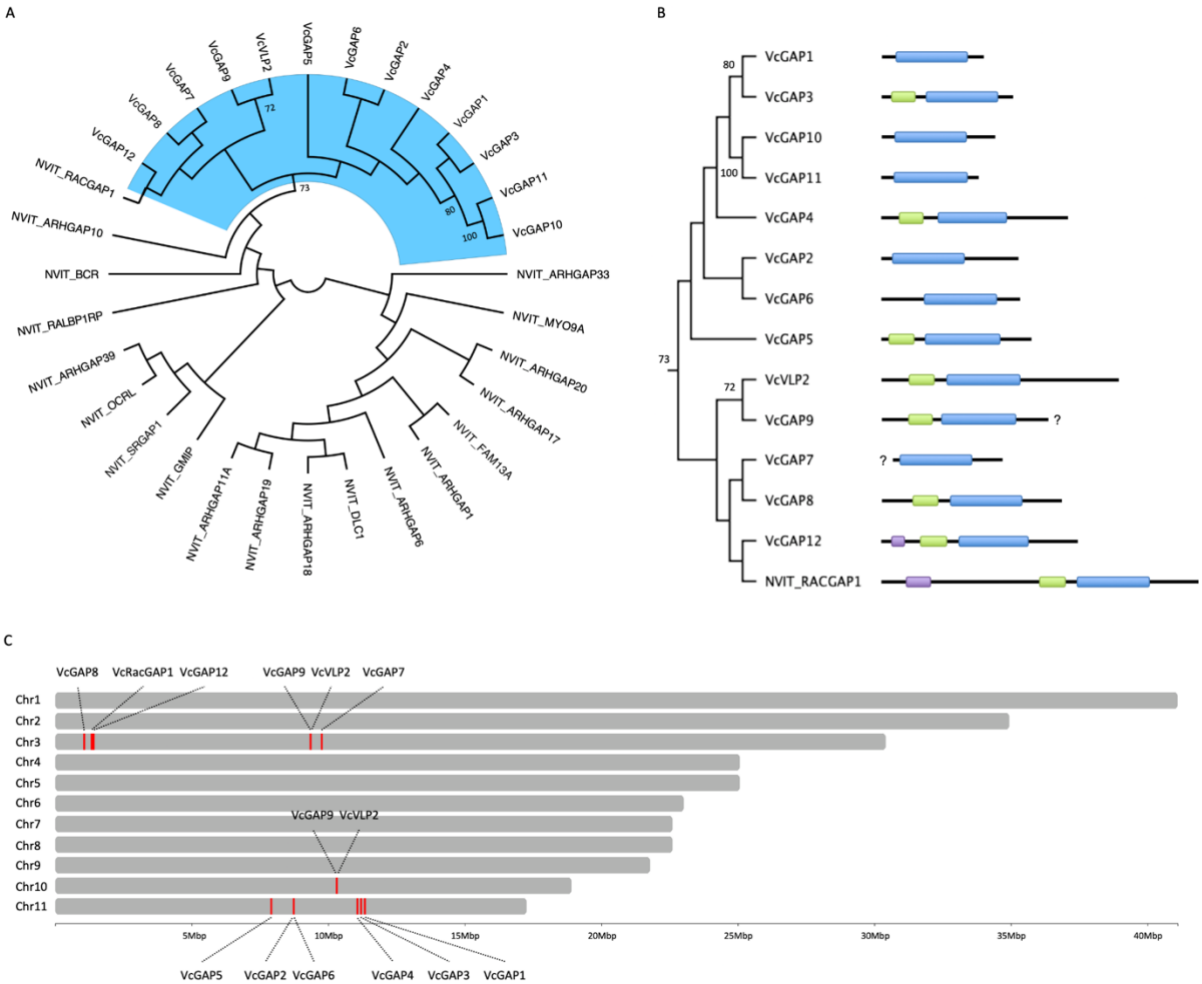
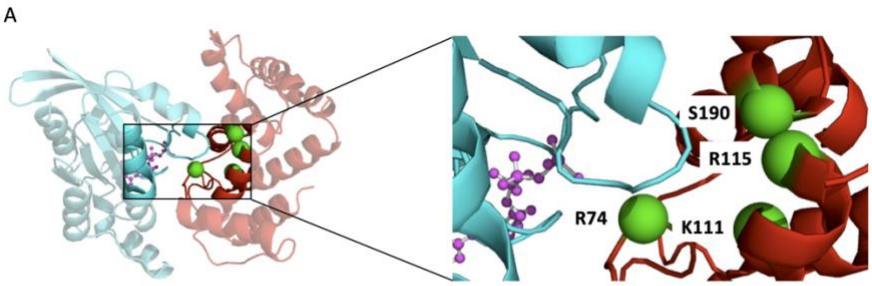


Figure 3. Origin of *V. canescens* calyx RhoGAPs (A) Maximum-likelihood phylogenetic tree of *V. canescens* calyx RhoGAPs along with *N. vitripennis* classical RhoGAPs. The phylogenetic tree was obtained with IQ-TREE using the RhoGAP domain amino acid sequence and displayed as a cladogram. Numbers at corresponding nodes are bootstrap support values in percent (500 bootstrap replicates). Only bootstrap values greater than 70% are shown. *V. canescens* calyx RhoGAPs are highlighted in blue. (B) Comparison of protein domain organization between *V. canescens* calyx RhoGAPs and *N. vitripennis* RacGAP1 using the RhoGAP domain as reference. The color code is as follows: purple for the coiled-coil region (CC), green for the protein kinase C-like zinc finger motif (C1) and blue for the RhoGAP domain. A (?) indicates that either the N- or C-terminal part of the sequence is unknown. (C) *V. canescens* chromosome map with the position of gene loci corresponding to RacGAP1 and calyx RhoGAPs visualized using chromoMap R package (Anand & Rodriguez Lopez 2022).

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

In a previous work, we identified four amino acid residues involved in the interaction of LbGAP with Rho GTPases (Colinet et al., 2007), including the key arginine residue (R74 in LbGAP) required for the GAP catalytic activity (Figure 4A). The other 8 venom RhoGAP sequences found in *L. boulardi* Ism and ISy were all mutated at this arginine residue (Figure 4B). Most also contained mutations in one or more of the other three amino acids involved in Rho GTPase interaction (Figure 4B). Similarly, all *V. canescens* calyx RhoGAP sequences, except VcVLP2, are mutated in one or more of the sites essential for GAP activity and/or involved in the interaction with Rho GTPases (Figure 4C). These observations suggest that only LbGAP and its homolog LbGAPy in *L. boulardi* and VcVLP2 in *V. canescens* are functional as RhoGAPs.



B

		74	111	115	190
	RacGAP1	R	K	R	T
Lbm	LbGAP	R	K	R	S
	LbGAP1.1	L	L	S	T
	LbGAP1.2	L	L	S	I
	LbGAP1.3	L	L	S	T
	LbGAP2.1	M	K	R	T
	LbGAP2.2	M	K	R	T
	LbGAP3	T	T	R	V
	LbGAP4	L	V	K	N
	LbGAP5	L	Q	Y	I
Lby	LbGAPy	R	K	R	S
	LbGAPy1.1	L	L	S	T
	LbGAPy1.2	S	I	S	T
	LbGAPy2	M	K	R	T
	LbGAPy3.1	T	T	R	V
	LbGAPy3.2	T	T	R	V
	LbGAPy4	L	V	K	N
	LbGAPy5	S	Q	Y	I
	LbGAPy6	F	K	R	T

C

		74	111	115	190
	RacGAP1	R	K	R	T
Vc	VLP2	R	K	R	T
	VcGAP1	G	K	A	I
	VcGAP2	Q	R	R	A
	VcGAP3	R	K	T	I
	VcGAP4	T	M	K	I
	VcGAP5	S	K	L	V
	VcGAP6	L	K	S	I
	VcGAP7	A	L	R	T
	VcGAP8	Q	K	R	V
	VcGAP9	K	K	R	V
	VcGAP10	R	K	S	I
	VcGAP11	T	K	F	I
	VcGAP12	R	K	A	T

Figure 4. Mutations in the essential sites for GAP activity and/or involved in interaction with Rho GTPases (A) Tertiary structure of LbGAP (red) in complex with Rac1 (blue) and the transition-state analogue GDP.AlF3 (modeled by homology for sequence spanning amino acid residues 51 to 216 in Colinet et al., 2007). The four sites essential for GAP activity and/or involved in interaction with RhoGTPases are colored green. (B) Amino acids found at the four sites essential for GAP activity and/or involved in the interaction with RhoGTPases for *Leptopilina* RacGAP1 and *L. boulardi* venom RhoGAP sequences. The numbering corresponds to the positions in the LbGAP sequence. (C) Amino acids at the four sites essential for GAP activity and/or involved in interaction with RhoGTPases for *V. canescens* RacGAP1 and calyx RhoGAP sequences. The numbering corresponds to the positions in the LbGAP sequence. In B and C, amino acids are colored according to their properties following the RasMol amino acid color scheme.

A search for positive selection was performed to detect possible functional divergence of *L. boulardi* venom and *V. canescens* calyx RhoGAP sequences. PAML codon-based models with (M2a and M8) and without (M1a and M7) selection were compared using likelihood ratio tests (LRTs). Both M1a/M2a and

M7/M8 comparisons resulted in the rejection of the null hypothesis suggesting that a fraction of codons in RhoGAP sequences are under positive selection (Table 3).

Table 3. Positive selection analysis among sites using CodeML for *L. boulardi* venom and *V. canescens* calyx RhoGAP sequences. lnL is the log likelihood of the model. p-value is the result of likelihood ratio tests (LRTs). Global ω is the estimate of the dN/dS ratio under the model (given as a weighted average). Parameters (ω > 1) are parameters estimates for a dN/dS ratio greater than 1.

L. boulardi

Model	lnL	p-value	Global ω	Parameters (ω > 1)
M1a: neutral	-5735.55		0.71	
M2a: selection	-5688.59	< 0.001	1.65	ω = 4.70, p = 0.21
M7: ß	-5741.07		0.70	
M8: ß&ω	-5688.19	< 0.001	1.56	ω = 4.4, p = 0.22

V. canescens

Model	lnL	p-value	Global ω	Parameters (ω > 1)
M1a: neutral	-12616.14		0.78	
M2a: selection	-12574.17	< 0.001	1.31	ω = 2.72, p = 0.28
M7: ß	-12612.64		0.73	
M8: ß&ω	-12564.70	< 0.001	1.21	ω = 2.30, p = 0.34

Consistently, a total of 7 and 11 branches of the phylogenetic tree constructed with the *L. boulardi* venom and *V. canescens* calyx RhoGAP sequences, respectively, were detected by the REL branch-site test as corresponding to lineages on which a subset of codons has evolved under positive selection (Figure 5A and 6A). For *L. boulardi*, 6 of the 7 lineages under positive selection were internal branches, indicating that selection occurred primarily before the separation of the *L. boulardi* Ism and ISy strains (Figure 5A).

The combined use of five different methods identified a total of 19 codons as candidates under positive selection for the *L. boulardi* venom RhoGAP sequences, most of which were found in the region corresponding to the RhoGAP domain (Figure 5B). In contrast, only 4 amino acids were detected as evolving under negative selection in the *L. boulardi* venom RhoGAPs, respectively (Figure 5C). The leucine at position 9 in LbGAP is located in the signal peptide, demonstrating the importance of this region. The other three amino acids in *L. boulardi* and most of those in *V. canescens* under negative selection are buried in the protein and are probably important for the structural stability (Figures 5D and 5E).

Interestingly, at least one radical change in charge and/or polarity of the corresponding amino acids was found for all selected codons (Figure 5B). For example, the polar uncharged serine residue at position 76 in LbGAP was replaced by a negatively charged lysine residue in LbGAP4 (Figure 5B). Since radical changes are more likely to modify protein function than conservative changes, this suggests that the identified non-synonymous substitutions may be adaptive. Interestingly, the corresponding amino acids are mostly exposed on the surface of the protein and therefore likely to interact with partners (Figures 5C and 5D).

We therefore generated site-specific mutants of LbGAP for 8 of the 19 amino acids under positive selection and compared their binding capabilities to Rho GTPases. Two-hybrid analysis revealed a lack of interaction for five of the mutants, indicating that the corresponding amino acids are essential for interaction with Rac GTPases (Table 4). On the other hand, the remaining three mutants were still able to interact strongly with Rac1 and Rac2, suggesting that the corresponding amino acids are not involved in the interaction with Rac GTPases (Table 4).

For the *V. canescens* calyx RhoGAP sequences, six amino acids were identified as candidates under positive selection and 19 under negative selection (Figures 6B and 6C). Four of the amino acids under positive selection are predicted to be exposed on the surface of the protein, indicating functional diversification in relation to interaction with partners as in *L. boulardi* (Figures 6D and 6E). Most of the amino acids under negative selection are buried within the protein and are likely to be important for structural stability (Figures 6D and 6E).

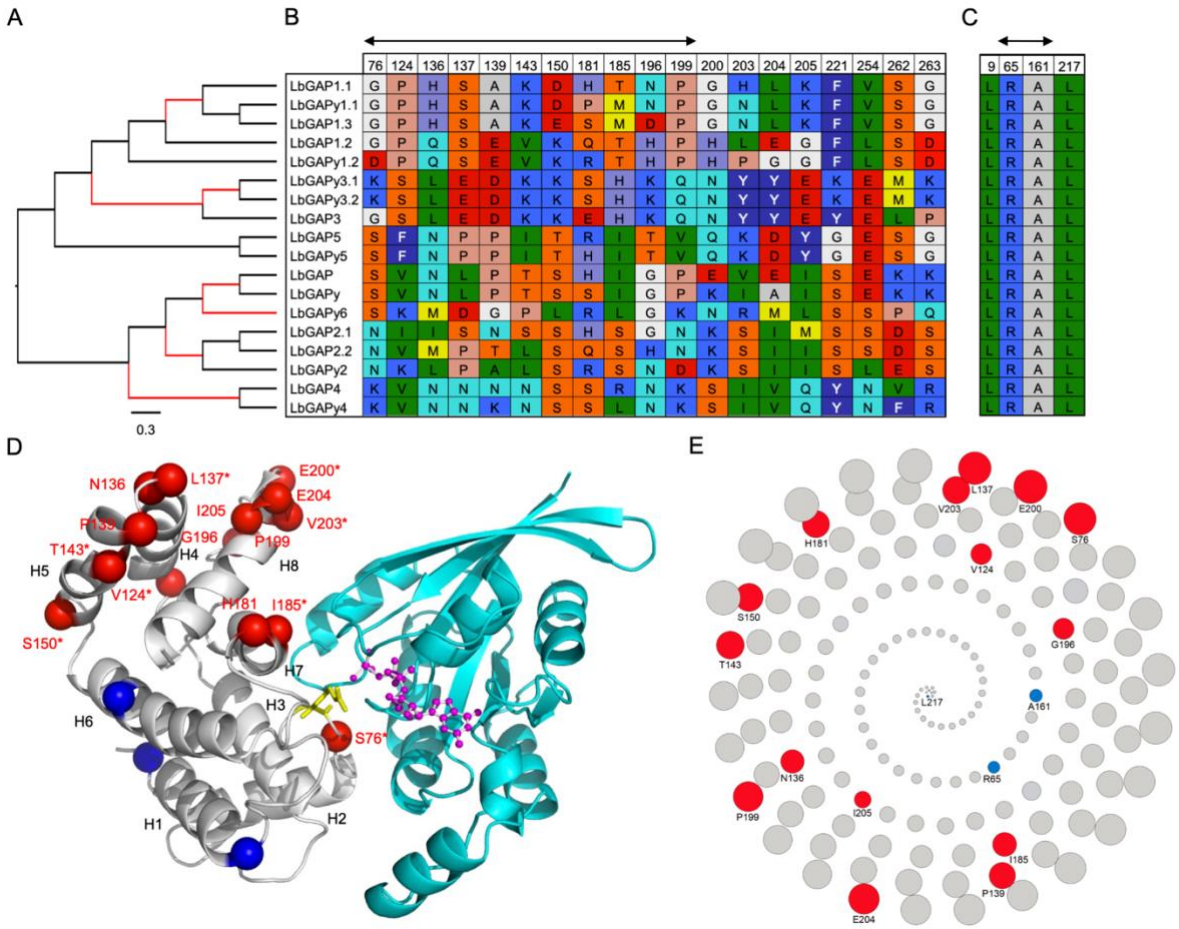


Figure 5. Identification of positively selected branches and codons in *L. boulardi* venom RhoGAPs. (A) Cladogram of *L. boulardi* venom RhoGAPs. Branches identified under positive selection by the HyPhy BSR method are colored in red. (B) Identified positively selected codons numbered according to LbGAP amino acid sequence. Amino acids are colored according to their properties following the RasMol amino acid color scheme. The two-sided arrow indicates sites located in the RhoGAP domain. (C) Identified negatively selected codons numbered according to LbGAP amino acid sequence. (D) Positively- and negatively selected sites displayed on the tertiary structure of LbGAP (grey) in complex with Rac1 (blue) and the transition-state analogue GDP.AlF3 (modeled by homology for sequence spanning amino acid residues 51 to 216 in Colinet et al., 2007). Helices are shown as ribbons and numbered according to their location in the amino-acid sequence of LbGAP. Positively selected sites are displayed as red-colored spheres and numbered according to the LbGAP sequence. A star indicates amino acid residues tested in mutagenesis and two-hybrid interaction assays. Negatively selected sites are displayed as spheres and colored in blue. Protruding LbGAP Arg74 is shown as sticks and colored in yellow. GDP.AlF3 is shown as ball-and-sticks models and colored in magenta. (E) Spiral view of LbGAP amino acids in the order of their solvent accessibility. Most accessible amino acids come on the outermost ring of the spiral whereas the buried amino acids are occurring in the innermost ring. Radius of the circles corresponds to the relative solvent accessibility. Red and blue colors are used for positively- and negatively selected amino acids respectively. All other amino acids are in grey.

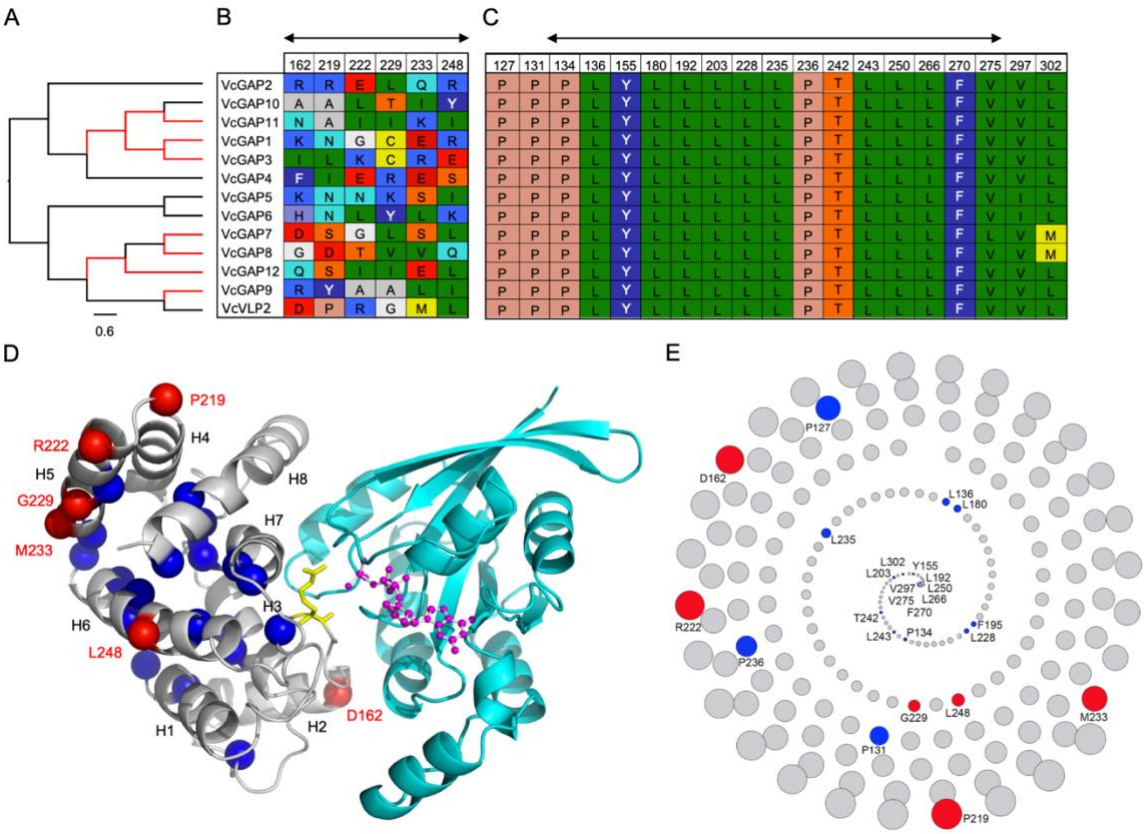


Figure 6. Identification of positively selected branches and codons in *V. canescens* calyx RhoGAPs. (A) Cladogram of *V. canescens* calyx RhoGAPs. Branches identified under positive selection by the HyPhy BSR method are colored in red. (B) Identified positively selected codons numbered according to VcVLP2 amino acid sequence. Amino acids are colored according to their properties following the RasMol amino acid color scheme. The two-sided arrow indicates sites located in the RhoGAP domain. (C) Identified negatively selected codons numbered according to VcvLP2 amino acid sequence. (D) Positively- and negatively selected sites displayed on the tertiary structure of VcVLP2 (grey) in complex with Rac1 (blue) and the transition-state analogue GDP.AlF3 (modeled by homology for sequence spanning amino acid residues 117 to 316). Helices are shown as ribbons and numbered according to their location in the amino-acid sequence of VcVLP2. Positively selected sites are displayed as red and negatively selected sites blue spheres, respectively, and numbered according to the VcVLP2 sequence. Protruding VcVLP2 Arg156 is shown as sticks and colored in yellow. GDP.AlF3 is shown as ball-and-sticks models and colored in magenta. (E) Spiral view of VcVLP2 amino acids in the order of their solvent accessibility. Most accessible amino acids come on the outermost ring of the spiral whereas the buried amino acids are occurring in the innermost ring. Radius of the circles corresponds to the relative solvent accessibility. Red and blue colors are used for positively- and negatively selected amino acids respectively. All other amino acids are in grey.

Table 4. Summary of the results of interaction assays for the LbGAP mutants. Results are based on growth on selective medium lacking histidine and qualitative ß-galactosidase overlay assays. -: no interaction; (+): very weak interaction; +++: strong interaction.

	LbGAP	LbGAP S76K	LbGAP V124K	LbGAP L137D	LbGAP T143L	LbGAP S150L	LbGAP I185K	LbGAP E200G	LbGAP V203R	T
Rac1	+++	-	-	+++	+++	+++	-	-	-	-
Rac2	+++	-	-	+++	+++	+++	-	-	-	-
CDC42	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
RhoA	-	-	-	-	-	-	-	-	-	-
Lamin	-	-	-	-	-	-	-	-	-	-

Discussion

Independent convergent recruitment of vesicle-associated RhoGAP proteins in distantly related parasitoid species

The process of duplication of a gene encoding a protein usually involved in a key physiological process whose function will be hijacked is a major mechanism of toxin recruitment (Fry et al., 2009; Wong & Belov 2012; Casewell et al., 2013). The results of phylogeny and domain analyses indicate that the *Leptopilina* venom RhoGAPs and the calyx RhoGAPs found in *V. canescens* VLPs originated from duplication of the gene encoding the cellular RacGAP1 in these species. This independent convergent recruitment may suggest a similar function in virulence in these two evolutionarily distant parasitoid species. Consistently, our results indicate that venom RhoGAPs in *L. boulardi* and *L. heterotoma* and calyx RhoGAPs in *V. canescens* are associated with extracellular vesicles, named venosomes and VLPs, respectively. While these two types of vesicles have different tissue origins and formation mechanisms, *Leptopilina* venosomes and *V. canescens* VLPs would act as transport systems to deliver virulent proteins, including RhoGAPs, into host hemocytes to alter their function (Pichon et al., 2015; Wan et al., 2019). However, the role of RhoGAPs on host hemocytes is still unclear or even unknown, although we have previously shown that LbGAP inactivates Rac-like GTPases (Colinet et al., 2007), which are required for successful encapsulation of *Leptopilina* eggs (Williams et al., 2005; Williams et al., 2006).

RacGAP1 is an intracellular multidomain protein consisting of a coiled-coil region followed by a C1 motif (protein kinase C-like zinc finger motif) and a RhoGAP domain (Tcherkezian & Lamane-Vane 2007). *Leptopilina* venom RhoGAPs, on the other hand, possess only a RhoGAP domain preceded by a secretory signal peptide. Sequence comparisons at the genomic level support the hypothesis of a single partial duplication of the RacGAP1 gene in the ancestor of the *Leptopilina* genus, resulting in the loss of a 5' portion of the sequence encoding the coiled-coil region and the C1 motif, followed by the acquisition of the signal peptide sequence. This initial evolutionary event would then have been followed by further duplication of the ancestral duplicate gene throughout the diversification of the genus *Leptopilina,* leading to the current number of venom RhoGAPs. A first hypothesis for the origin of the signal peptide would be that the loss of part of the 5' coding sequence in the duplicate resulted in a new N-terminal end for the encoded protein with signal peptide properties. Modification of the N-terminal end of the protein into a fully functional signal peptide may have required some point mutations after partial duplication, as in the case of the antifreeze protein of an Antarctic zoarcid fish (Deng et al., 2010). Another possibility is that the signal peptide originated from a process of partial duplication with recruitment (Katju & Lynch 2006), in which a "proto-peptide signal" coding sequence was acquired from the new genomic environment into which the partial copy was integrated. Finally, it is also possible that the new 5' region of the duplicated copy results from a chimeric duplication (Katju & Lynch 2006), e.g. from the partial duplication of another gene or by exon shuffling (Vibranovski et al., 2006). The acquisition of the signal peptide probably facilitated the neofunctionalization of the venom RhoGAPs according to the evolutionary model of protein subcellular relocalization (PSR) proposed by Byun McKay & Geeta (2007), according to which a modification of the N- or C-terminal region of a protein can change its subcellular localization, by the loss or acquisition of a specific localization signal, and enable it to acquire a new function.

In *V. canescens,* the N-terminal region of the calyx RhoGAPs is more variable in length, with some proteins retaining the C1 motif upstream of the RhoGAP domain and one retaining a coiled-coil region upstream of the C1 motif. Most of the *V. canescens* calyx RhoGAPs were predicted to contain a nuclear localization signal, consistent with (i) the accumulation of VcVLP2 in the nucleus of the calyx cells prior to its association with the virus-derived VLPs (Pichon et al., 2015) and (ii) our results from a proteomic analysis indicating that all 13 calyx RhoGAPs are associated with VLPs. Human RacGAP1 has been described to localize to the cytoplasm and nucleus (Mishima et al., 2002), suggesting that *V. canescens* calyx RhoGAPs have retained the RacGAP1 nuclear localization signal, unlike *Leptopilina* venom RhoGAPs. One of the *V. canescens* calyx RhoGAP genes is located close to the RacGAP1 gene in the genome, suggesting a recent partial tandem duplication. The other genes, on the other hand, are scattered throughout the genome suggesting that they originate from one or more older duplication events of the RacGAP1 gene. In contrast to *Leptopilina*, it cannot be excluded that two or more partial duplication events of the RacGAP1-encoding gene occurred in *V. canescens*, followed by further duplication of ancestral duplicates.

Accelerated evolution through duplication and divergence: pseudogenization and/or neofunctionalization?

Our work revealed a significant divergence for the two multigene families of RhoGAPs in *L. boulardi* and *V. canescens* compared to RacGAP1, illustrated by the presence of mutations on the arginine essential for GAP activity and/or on one or more of the amino acids shown to be important, or even necessary, for interaction with the Rac GTPases in all RhoGAPs except LbGAP (and its homolog LbGAPy) and VcVLP2. The presence of mutations at key sites in the majority of *L. boulardi* venom and *V. canescens* calyx RhoGAPs suggests a loss of function by pseudogenization. However, peptide matches were found in proteomic analyses for each of these RhoGAPs, indicating that the corresponding genes are successfully transcribed and translated. Furthermore, our results indicate that the ancestrally acquired secretory signal peptide, in which we identified an amino acid under negative selection, is conserved in all *L. boulardi* RhoGAPs and that most *V. canescens* RhoGAPs conserved the nuclear localization signal of RacGAP1. In addition, all RhoGAPs retained the ability to associate with vesicles, namely venosomes in *L. boulardi* and VLPs in *V. canescens*. Finally, we showed that several amino acids embedded in the protein structure evolved under negative selection, suggesting that they are important for protein stability. Taken together, these observations are not consistent with the hypothesis that any of the mutated RhoGAPs is undergoing a pseudogenization process, although it cannot be excluded that this process is quite recent and could have been initiated by the loss of the GAP active site.

An alternative hypothesis to pseudogenization would be functional diversification of the mutated RhoGAPs independent of the Rho pathway. Indeed, we have evidenced an evolution under positive selection for the majority of RhoGAPs in the two multigenic families. Most of the sites under positive selection are located on the surface of the protein and are therefore likely to interact with partners. In addition, directed mutagenesis and two-hybrid experiments in *L. boulardi* have shown that some of these amino acids are not involved in the interaction with Rac GTPases. Although the majority of studies on the RhoGAP domain concern the interaction with Rho GTPases, interactions with other proteins have also been described (Ban et al., 2004; Xu et al., 2013), supporting the hypothesis of a neofunctionalization of the mutated RhoGAPs. Further studies will be needed to determine which functions the mutated RhoGAPs have acquired in relation to parasitism. Recently, a possible role has been proposed for one of the RhoGAPs found in the venom of the ISy line of *L. boulardi* in the induction of reactive oxygen species (ROS) in the central nervous system of *D. melanogaster* in the context of superparasitism avoidance (Chen et al., 2021). However, ROS production is notably regulated by the GTPases Rac1 and Rac2 (Hobbs et al., 2014), whereas this venom protein (named EsGAP1 in Chen et al., 2021 and corresponding to LbGAPy6 in our study) has likely lost its RhoGAP activity due to a mutation on the arginine at position 74. Therefore, the mechanism by which EsGAP1 would be involved in ROS induction is unclear.

Other potential viral factors involved in parasitic success have been described previously that are encoded by large gene families and similarly correspond to truncated forms of cellular proteins that have retained only a single conserved domain, such as protein tyrosine phosphatases (PTPs) or viral ankyrins (V-ANKs) of bracoviruses. In the case of bracovirus PTPs, some are active as phosphatases, while others are mutated in their catalytic site and have been shown to be inactive as PTPs (Provost et al., 2004). In the latter case, as well as for V-ANK, it has been suggested that by binding to their target, they may act as constitutive inhibitors of the function of the corresponding cellular protein (Provost et al., 2004; Thoetkiattikul et al., 2005). Moreover, the model of adaptive evolution by competitive evolution of duplicated gene copies (Francino 2005) predicts that after a first step in which different copies explore the mutation space, once a protein with an optimal function is obtained, the other copies will begin to decay and undergo pseudogenization. This can lead to intermediate situations where both functional and non-functional proteins are produced, as observed for different virulence protein families of parasitoid wasps. Some of the described features of *L. boulardi* and *V. canescens* RhoGAPs may suggest that a process of competitive evolution is underway in *V. canescens* and *L. boulardi*, although our results are also consistent with the hypothesis of a possible neofunctionalization of these vesicle-associated proteins as discussed above.

In conclusion, we evidenced the independent convergent origin and accelerated evolution of a multigene family of vesicle-associated RhoGAP proteins in two unrelated parasitic wasps. Strikingly, these vesicles, which are similarly involved in parasitism, are produced in distinct organs: the venom apparatus in *Leptopilina* and the ovarian calyx in *V. canescens*. In the case of *Leptopilina* RhoGAPs, the acquisition of

a secretory signal peptide after incomplete duplication of the RacGAP1 gene allowed secretion into the venom where the vesicles are formed. *V. canescens* RhoGAPs, on the other hand, probably retained the RacGAP1 nuclear localization signal, because it is in the nucleus of calyx cells that VLPs are formed. Another striking point is that our results suggest a possible functional diversification of vesicle-associated RhoGAPs in both species, with the exception of LbGAP in *L. boulardi* and VcVLP2 in *V. canescens*, which are probably the only ones with RhoGAP activity. An open question would be whether all RhoGAPs are important for parasite success, but by different mechanisms, independent of the Rho pathway or not, depending on whether RhoGAPs are mutated or not.

Acknowledgements

We are highly grateful to Christian Rebuf for help in insects rearing. We also thank Pr. M. T. Kimura (Hokkaido University, Japan) for providing biological material. We would also like to thank Anthony Bretaudeau (INRIA, Rennes, France) for help in the use of the workflows available at the BIPAA platform, the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12) for providing computing resources and the BIG bioinformatics platform from the PlantBios infrastructure for providing facilities and technical support.

Data and supplementary information availability

Raw data from the Illumina RNASeq sequencing of the *L. victoriae* venom apparatus are available at NCBI under the BioProject ID PRJNA974978.

The mass spectrometry proteomics data from *L. boulardi* and *L. heterotoma* venosomes have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2022) partner repository with the dataset identifier PXD041695.

Supplementary information is available online at Data INRAE: https://doi.org/10.57745/K82IWQ

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Jean-Luc Gatti and Marylène Poirié are recommenders of PCI Zoology.

Funding

This work was supported by the Department of Plant Health and Environment from INRAE and the French Government (National Research Agency, ANR) through the “Investments for the Future” programs LABEX SIGNALIFE ANR-11-LABX-0028-01 and IDEX UCAJedi ANR-15-IDEX-01.

References

Abd-Alla AM, Cousserans F, Parker AG, Jehle JA, Parker NJ, Vlak JM, Robinson AS, Bergoin M (2008) Genome analysis of a *Glossina pallidipes* salivary gland hypertrophy virus reveals a novel, large, double-stranded circular DNA virus. *Journal of Virology,* **82**, 4595-4611. https://doi.org/10.1128/jvi.02588-07

Anand L, Rodriguez Lopez CM (2022) ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. *BMC Bioinformatics,* **23**, 33. https://doi.org/10.1186/s12859-021-04556-z

Ban R, Irino Y, Fukami K, Tanaka H (2004) Human mitotic spindle-associated protein PRC1 inhibits MgcRacGAP activity toward Cdc42 during the metaphase. *Journal of Biological Chemistry,* **279**, 16394-16402. https://doi.org/10.1074/jbc.m313257200

Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics,* **27**, 343-350. https://doi.org/10.1093/bioinformatics/btq662

Bézier A, Herbinière J, Lanzrein B, Drezen JM (2009) Polydnavirus hidden face: the genes producing virus particles of parasitic wasps. *Journal of Invertebrate Pathology,* **101**, 194-203. https://doi.org/10.1016/j.jip.2009.04.006

Brameier M, Krings A, MacCallum RM (2007) NucPred--predicting nuclear localization of proteins. *Bioinformatics,* **23**, 1159-1160. https://doi.org/10.1093/bioinformatics/btm066

Braunagel SC, Elton DM, Ma H, Summers MD (1996) Identification and analysis of an *Autographa californica* nuclear polyhedrosis virus structural protein of the occlusion-derived virus envelope: ODV-E56. *Virology,* **217**, 97-110. https://doi.org/10.1006/viro.1996.0097

Braunagel SC, Williamson ST, Saksena S, Zhong Z, Russell WK, Russell DH, Summers MD (2004) Trafficking of ODV-E66 is mediated via a sorting motif and other viral proteins: facilitated trafficking to the inner nuclear membrane. *Proceedings of the National Academy of Sciences,* **101**, 8372-8377. https://doi.org/10.1073/pnas.0402727101

Braunagel SC, Cox V, Summers MD (2009) Baculovirus data suggest a common but multifaceted pathway for sorting proteins to the inner nuclear membrane. *Journal of Virology,* **83**, 1280-1288. https://doi.org/10.1128/jvi.01661-08

Byun McKay SA, Geeta R (2007) Protein subcellular relocalization: A new perspective on the origin of novel genes. *Trends in Ecology & Evolution*, **22**, 338–344. https://doi.org/10.1016/j.tree.2007.05.002

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics,* **25**, 1972-1973. https://doi.org/10.1093/bioinformatics/btp348

Carton Y, Poirié M, Nappi AJ (2008) Insect immune resistance to parasitoids. *Insect Science,* **15**, 67-87. https://doi.org/10.1111/j.1744-7917.2008.00188.x

Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG (2013) Complex cocktails: the evolutionary novelty of venoms. *Trends in Ecology & Evolution,* **28**, 219-229. https://doi.org/10.1016/j.tree.2012.10.020

Chen S, Krinsky BH, Long M (2013) New genes as drivers of phenotypic evolution. *Nature Reviews Genetics,* **14**, 645-660. https://doi.org/10.1038/nrg3521

Chen J, Fang G, Pang L, Sheng Y, Zhang Q, Zhou Y, Zhou S, Lu Y, Liu Z, Zhang Y, Li G, Shi M, Chen X, Zhan S, Huang J (2021) Neofunctionalization of an ancient domain allows parasites to avoid intraspecific competition by manipulating host behaviour. *Nature Communications,* **12**, 5489. https://doi.org/10.1038/s41467-021-25727-9

Colinet D, Schmitz A, Depoix D, Crochard D, Poirié M (2007) Convergent use of RhoGAP toxins by eukaryotic parasites and bacterial pathogens. PloS Pathog **3**: e203. https://doi.org/10.1371/journal.ppat.0030203

Colinet D, Schmitz A, Cazes D, Gatti J-L, and Poirié M (2010) The origin of intraspecific variation of virulence in an eukaryotic immune suppressive parasite. *PLoS Pathogens,* **6**, e1001206. https://doi.org/10.1371/journal.ppat.1001206

Colinet D, Deleury E, Anselme C, Cazes D, Poulain J, Azema-Dossat C, Belghazi M, Gatti J-L, Poirié M (2013) Extensive inter- and intraspecific venom variation in closely related parasites targeting the same host: the case of *Leptopilina* parasitoids of *Drosophila*. *Insect Biochemistry and Molecular Biology,* **43**, 601–611. https://doi.org/10.1016/j.ibmb.2013.03.010

Deng C, Cheng C-HC, Ye H, He X, Chen L (2010) Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proceedings of the National Academy of Sciences,* **107**, 21593-21598. https://doi.org/10.1073/pnas.1007883107

Delport W, Poon AF, Frost SD, Kosakovsky Pond SL (2010) Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics,* **26**, 2455-2477. https://doi.org/10.1093/bioinformatics/btq429

Drezen J-M, Chevignon G, Louis F, Huguet E (2014) Origin and evolution of symbiotic viruses associated with parasitoid wasps. *Current Opinion in Insect Science*, **6**, 35-43. https://doi.org/10.1016/j.cois.2014.09.008

Du J, Lin Z, Volovych O, Lu Z, Zou Z (2020) A RhoGAP venom protein from *Microplitis mediator* suppresses the cellular response of its host *Helicoverpa armigera*. *Developmental & Comparative Immunology,* **108**, 103675. https://doi.org/10.1016/j.dci.2020.103675

Dupas S, Frey F, Carton Y (1998) A single parasitoid segregating factor controls immune suppression in *Drosophila*. *Journal of Heredity,* **89**, 306-311. https://doi.org/10.1093/jhered/89.4.306

Edgar RC (2004) Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* **32**, 1792–1797. https://doi.org/10.1093/nar/gkh340

Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. *Genome Informatics,* **23**, 205–211. https://doi.org/10.1142/9781848165632_0019

Feddersen I, Sander K, Schmidt O (1986) Virus-like particles with host protein-like antigenic determinants protect an insect parasitoid from encapsulation. *Experientia,* **42**, 1278-1281. https://doi.org/10.1007/BF01946422

Francino MP (2005) An adaptive radiation model for the origin of new gene functions. *Nature Genetics,* **37**, 573-577. https://doi.org/10.1038/ng1579

Fromont-Racine M, Rain JC, Legrain P (1997) Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genetics,* **16**, 277–282. https://doi.org/10.1038/ng0797-277

Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, Nevalainen TJ, Normann JA, Lewis RJ, Norton RS, Renjifo D, Rodriguez de la Vega RC (2009) The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annual Review of Genomics and Human Genetics,* **10**, 483–511. https://doi.org/10.1146/annurev.genom.9.081307.164356

Gatti J-L, Schmitz A, Colinet D, Poirié M (2012) Diversity of virus-like particles in parasitoids' venom: viral or cellular origin? In: Beckage NE, Drezen J-M, editors, Parasitoid viruses, pp. 181–192, London: Academic Press. https://doi.org/10.1016/B978-0-12-384858-1.00015-1

Godfray HCJ (1994) Parasitoids: behavioral and evolutionary ecology (Vol. 67): Princeton University Press. https://doi.org/10.2307/j.ctvs32rmp

Goecks J, Mortimer NT, Mobley JA, Bowersock GJ, Taylor J, Schlenke TA (2013) Integrative approach reveals composition of endoparasitoid wasp venoms. *PLoS One,* **8**, e64125. https://doi.org/10.1371/journal.pone.0064125

Goldberg T, Hecht M, Hamp T, Karl T, Yachdav G, Ahmed N, Altermann U, Angerer P, Ansorge S, Balasz K, Bernhofer M, Betz A, Cizmadija L, Do KT, Gerke J, Greil R, Joerdens V, Hastreiter M, Hembach K, Herzog M, Kalemanov M, Kluge M, Meier A, Nasir H, Neumaier U, Prade V, Reeb J, Sorokoumov A, Troshani I, Vorberg S, Waldraff S, Zierer J, Nielsen H, Rost B (2014) LocTree3 prediction of localization. *Nucleic Acids Research,* **42**, W350-355. https://doi.org/10.1093/nar/gku396

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology,* **59**, 307–321. https://doi.org/10.1093/sysbio/syq010

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols,* **8**, 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hobbs GA, Zhou B, Cox AD, Campbell SL (2014) Rho GTPases, oxidation, and cell redox control. *Small GTPases,* **5**, e28579. https://doi.org/10.4161/sgtp.28579

Hong T, Summers MD, Braunagel SC (1997) N-terminal sequences from *Autographa californica* nuclear polyhedrosis virus envelope proteins ODV-E66 and ODV-E25 are sufficient to direct reporter proteins to the nuclear envelope, intranuclear microvesicles and the envelope of occlusion derived virus. *Proceedings of the National Academy of Sciences,* **94**, 4050-4055. https://doi.org/10.1073/pnas.94.8.4050

Jancek S, Bézier A, Gayral P, Paillusson C, Kaiser L, Dupas S, Le Ru BP, Barbe V, Periquet G, Drezen J-M, Herniou E A (2013) Adaptive selection on bracovirus genomes drives the specialization of *Cotesia* parasitoid wasps. *PLoS One*, **8**, e64432. https://doi.org/10.1371/journal.pone.0064432

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577-2637. https://doi.org/10.1002/bip.360221211

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods,* **14**, 587–589. https://doi.org/10.1038/nmeth.4285

Katju V, Lynch M (2006) On the formation of novel gene by duplication in the *Caenorhabditis elegans* genome. *Molecular Biology and Evolution,* **23**, 1056–1067. https://doi.org/10.1093/molbev/msj114

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution,* **30**, 772–780. https://doi.org/10.1093/molbev/mst010

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols,* **10**, 845-858. https://doi.org/10.1038/nprot.2015.053

Kosakovsky Pond SL, Frost SD (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution,* **22**, 1208-1222. https://doi.org/10.1093/molbev/msi105

Kosakovsky Pond SL, Murrell B, Fourment M, Frost SD, Delport W, Scheffler K (2011) A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution,* **28**, 3033-3043. https://doi.org/10.1093/molbev/msr125

Labrosse C, Stasiak K, Lesobre J, Grangeia A, Huguet E, Drezen J-M, Poirié M (2005) A RhoGAP protein as a main immune suppressive factor in the *Leptopilina boulardi* (Hymenoptera, Figitidae)-*Drosophila melanogaster* interaction. *Insect Biochemistry and Molecular Biology,* **35**, 93–103. https://doi.org/10.1016/j.ibmb.2004.10.004

Lefort V, Longueville J-E, Gascuel O (2017) SMS : Smart model selection in PhyML. *Molecular Biology and Evolution,* **34**, 2422-2424. https://doi.org/10.1093/molbev/msx149

Long M, VanKuren NW, Chen S, Vibranovski MD (2013) New gene evolution: little did we know. *Annual Review of Genetics,* **47**, 307-333. https://doi.org/10.1146/annurev-genet-111212-133301

Lucas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162-1164. https://doi.org/10.1126/science.252.5009.1162

Millar AH, Whelan J, Small I (2006) Recent surprises in protein targeting to mitochondria and plastids. *Current Opinion in Plant Biology,* **9**, 610-615. https://doi.org/10.1016/j.pbi.2006.09.002

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution,* **37**, 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mishima M, Kaitna S, Glotzer M (2002) Central spindle assembly and cytokinesis require a kinesin-like protein/RhoGAP complex with microtubule bundling activity. *Developmental Cell,* **2**, 41–54. https://doi.org/10.1016/s1534-5807(01)00110-1

Mueller JC, Andreoli C, Prokisch H, Meitinger T (2004) Mechanisms for multiple intracellular localization of human mitochondrial proteins. *Mitochondrion,* **3**, 315-325. https://doi.org/10.1016/j.mito.2004.02.002

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics,* **8**, e1002764. https://doi.org/10.1371/journal.pgen.1002764

Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K (2013) FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution,* **30**, 1196-1205. https://doi.org/10.1093/molbev/mst030

Nakai K, Horton P (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends in Biochemical Sciences,* **24**, 34-36. https://doi.org/10.1016/s0968-0004(98)01336-x

Novković B, Mitsui H, Suwito A, Kimura MT (2011) Taxonomy and phylogeny of *Leptopilina* species (Hymenoptera: Cynipoidea: Figitidae) attacking frugivorous drosophilid flies in Japan, with description of three new species. *Entomological Science,* **14**, 333–346. https://doi.org/10.1111/j.1479-8298.2011.00459.x

Pennacchio F, Strand MR (2006) Evolution of developmental strategies in parasitic Hymenoptera. *Annual Review of Entomology,* **51**, 233–258. https://doi.org/10.1146/annurev.ento.51.110104.151029

Perez-Riverol Y, Bai J, Bandla C, Hewapathirana S, García-Seisdedos D, Kamatchinathan S, Kundu D, Prakash A, Frericks-Zipper A, Eisenacher M, Walzer M, Wang S, Brazma A, Vizcaíno JA (2022). The PRIDE database resources in 2022: A Hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research,* **50**, D543-D552. https://doi.org/10.1093/nar/gkab1038

Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A, Podsiadlowski L, Petersen M, Lanfear R, Diez PA, Heraty J, Kjer KM, Klopfstein S, Meier R, Polidori C, Schmitt T, Liu S, Zhou X, Wappler T, Rust J, Misof B, Niehuis O (2017) Evolutionary history of the Hymenoptera. *Current Biology,* **27**, 1013-1018. https://doi.org/10.1016/j.cub.2017.01.027

Pichon A, Bézier A, Urbach S, Aury J-M, Jouan V, Ravallec M, Guy J, Cousserans F, Thézé J, Gauthier J, Demettre E, Schmieder S, Wurmser F, Sibut V, Poirié M, Colinet D, da Silva C, Couloux A, Barbe V, Drezen J-M, Volkoff A-N (2015) Recurrent DNA virus domestication leading to different parasite virulence strategies. *Science Advances,* 1, e1501150. https://doi.org/10.1126/sciadv.1501150

Poirié M, Carton Y, Dubuffet A (2009) Virulence strategies in parasitoid Hymenoptera as an example of adaptive diversity. *Comptes Rendus Biologies,* **332**, 311–320. https://doi.org/10.1016/j.crvi.2008.09.004

Poirié M, Colinet D, Gatti J-L (2014) Insights into function and evolution of parasitoid wasp venoms. *Current Opinion in Insect Science,* **6**, 52–60. https://doi.org/10.1016/j.cois.2014.10.004

Provost B, Varricchio P, Arana E, Espagne E, Falabella P, Huguet E, La Scaleia R, Cattolico L, Poirié M, Malva C, Olszewski JA, Pennacchio F, Drezen JM (2004) Bracoviruses contain a large multigene family coding for protein tyrosine phosphatases. *Journal of Virology,* **78**, 13090-13103. https://doi.org/10.1128/jvi.78.23.13090-13103.2004

Rago, A, Gilbert DG, Choi J-H, Sackton TB, Wang X, Kelkar YD, Werren JH, Colbourne JK (2016) OGS2: genome re-annotation of the jewel wasp *Nasonia vitripennis*. *BMC Genomics,* **17**, 678. https://doi.org/10.1186/s12864-016-2886-9

Reineke A, Asgari S, Schmidt O (2006) Evolutionary origin of *Venturia canescens* virus-like particles. *Archives of Insect Biochemistry and Physiology,* **61**, 123–133. https://doi.org/10.1002/arch.20113

Serbielle C, Chowdhury S, Pichon S, Dupas S, Lesobre J, Purisima E, Drezen J-M, Huguet E (2008) Viral cystatin evolution and three-dimensional structure modelling: a case of directional selection acting on a viral protein involved in a host-parasitoid interaction. *BMC Biology,* **6**, 38. https://doi.org/10.1186/1741-7007-6-38

Serbielle C, Dupas S, Perdereau E, Héricourt F, Dupuy C, Huguet E, Drezen, J-M (2012) Evolutionary mechanisms driving the evolution of a large polydnavirus gene family coding for protein tyrosine phosphatases. *BMC Ecology and Evolution,* **12**, 253. https://doi.org/10.1186/1471-2148-12-253

Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics,* **6**, 31. https://doi.org/10.1186/1471-2105-6-31

Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T (2007) Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Research,* **35**(Web Server issue), W506-511. https://doi.org/10.1093/nar/gkm382

Tcherkezian J, Lamarche-Vane N (2007). Current knowledge of the large RhoGAP family of proteins. *Biology of the Cell,* **99**, 67–86. https://doi.org/10.1042/bc20060086

Thoetkiattikul H, Beck MH, Strand MR (2005) Inhibitor kappaB-like proteins from a polydnavirus inhibit NF-kappaB activation and suppress the insect immune response. *Proceedings of the National Academy of Sciences,* **102**, 11426-11431. https://doi.org/10.1073/pnas.0505240102

Vibranovski MD, Sakabe NJ, de Souza SJ (2006) A possible role of exon-shuffling in the evolution of signal peptides of human proteins. *FEBS Letters,* **580**, 1621–1624. https://doi.org/10.1016/j.febslet.2006.01.094

Wan B, Goguet E, Ravallec M, Pierre O, Lemauf S, Volkoff A-N, Gatti J-L, Poirié M (2019) Venom atypical extracellular vesicles as interspecies vehicles of virulence factors involved in host specificity: the case of a drosophila parasitoid wasp. *Frontiers in Immunology,* **10**, 1688. https://doi.org/10.3389/fimmu.2019.01688

Williams MJ, Ando I, Hultmark D (2005) *Drosophila melanogaster* Rac2 is necessary for a proper cellular immune response. *Genes to Cells*, **10**, 813–823. https://doi.org/10.1111/j.1365-2443.2005.00883.x

913 Williams MJ, Wiklund ML, Wikman S, Hultmark D (2006) Rac1 signalling in the *Drosophila* larval cellular
914 immune response. *Journal of Cell Science,* **119**, 2015–2024. https://doi.org/10.1242/jcs.02920
915 Wong ESW, Belov K (2012) Venom evolution through gene duplications. *Gene,* **13**, 59–69.
916 https://doi.org/10.1016/j.gene.2012.01.009
917 Xu J, Zhou X, Wang J, Li Z, Kong X, Qian J, Hu Y, Fang J-Y (2013) RhoGAPs attenuate cell proliferation by
918 direct interaction with p53 tetramerization domain. *Cell Reports,* **3**, 1526-1538.
919 https://doi.org/10.1016/j.celrep.2013.04.017
920 Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics,*
921 **13**, 555-556. https://doi.org/10.1093/bioinformatics/13.5.555
922 Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution,* **24**,
923 1586-1591. https://doi.org/10.1093/molbev/msm088
924 Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous
925 selection pressure at amino acid sites. *Genetics,* **155**, 431-449.
926 https://doi.org/10.1093/genetics/155.1.431
927 Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive
928 selection. *Molecular Biology and Evolution,* **22**, 1107–1118. https://doi.org/10.1093/molbev/msi097
929 Zhang J (2000) Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian
930 nuclear genes. *Journal of Molecular Evolution,* **50**, 56–68. https://doi.org/10.1007/s002399910007
931