Genomic signatures of the evolution of loss of egg yolk in parasitoid wasps

Xianxin Zhao[1], Yuanyuan Liu[1], Yi Yang[1], Chun He[1], Kevin C. Chan[2], Haiwei Lin[1], Qi Fang[1], Gongyin Ye[1, *], Xinhai Ye[3, *]

[1]State Key Laboratory of Rice Biology and Breeding & Ministry of Agricultural and Rural Affairs Key Laboratory of Molecular Biology of Crop Pathogens and Insects, Institute of Insect Sciences, Zhejiang University, Hangzhou 310058, China.

[2]Shanghai Institute for Advanced Study, Zhejiang University, Shanghai 201203, China.

[3]College of Advanced Agriculture Science, Zhejiang A&F University, Hangzhou 311300, China.

[*]Corresponding authors: yexinhai@zafu.edu.cn; chu@zju.edu.cn.

Abstract

Losing or regressing traits has occurred repeatedly throughout evolutionary history in response to changing living environments. In parasitoid wasps, a mega-diverse group of insects in Hymenoptera, the loss or reduction of yolk in eggs has been reported in numerous species, and this phenotypic change is likely to have evolved as a response to the shift from ectoparasitism to endoparasitism. However, the genetic basis of this trait loss and the impact of trait loss on genome evolution are still poorly understood. Here, we perform comparative genomic analysis of 64 hymenopteran insects, and observe that a conserved insect yolk protein gene *vitellogenin* (*Vg*) was repeatedly lost in 23 endoparasitoids, representing five independent loss events in five families. Whole genome alignments suggest the loss of *Vgs* was probably associated with genome rearrangements. We also discover the gene losses, relaxed selection, and

protein structural variations of *Vg* receptor genes (*VgRs*) in the *Vg* loss lineages, showing the obviously functional biased patterns of gene loss. Our hypothesis test further reveals that the ecto-to-endo transition is not the main reason for the *Vg* losses and the subsequent evolution of *VgRs*. In addition, we find a number of parallel and convergent genomic changes including gene family evolution and gene selection related to transport, development, and metabolism in the independent *Vg* loss lineages. And these changes may have facilitated the embryonic development in these lineages. Together, our findings uncover the genomic basis of a unique trait loss in parasitoid wasps. More broadly, our study enhances our understanding of yolk loss evolution outside mammals, highlighting a potential evolutionary trend when an alternative nutrient source is available for the embryonic development.

Key words: Trait loss evolution, Egg yolk loss, Gene loss, Vitellogenin, Vitellogenin receptor, Parasitoid wasp, Parasitic strategy evolution

Introduction

Understanding the genetic basis of phenotype shifts or evolutionary novelties throughout the tree of life is a fundamental task in evolutionary biology. Trait loss or regression (i.e., complete loss or reduction of an ancestral trait), as a special phenotype shift, has occurred repeatedly across evolutionary history, likely in response to changes in living environment, habitat and species interaction network[1–3]. Multiple genome sequencing and comparative genomic works have uncovered the genetic basis of evolutionary trait losses in animals, such as loss of eye and pigmentation of animals living in darkness[4–8], teeth loss in birds[9–11], limb loss in snakes[12], tail loss in humans and apes[13]. These studies highlighted

the importance of gene loss and inactivating mutation in the evolution of trait loss or regression. And these stories also underlined the power of comparative genomics in finding the genetic mechanisms behind the evolution of complex traits.

Parasitic species are ubiquitous in nature, and trait loss and the associated genetic basis has been a hot topic in their evolution, as long-term host dependence may lead to the loss of some traits, especially nutrient-related pathways[14,15]. For example, in parasitic plants, prevalent gene loss involved in photosynthesis, nutrient uptake, flowering regulation, defense response, and root development were found[16–19]. In parasitic worms, a large-scaled genome analysis of parasitic nematodes and platyhelminths revealed the widespread losses of metabolic pathways and differential pathway coverage across different evolutionary clades[20]. In addition, some gene losses of homeobox gene families[21] and cilium-related genes[22] may be related to their simplified morphology. In insects, the largest and most diverse group of organisms on Earth, parasitoids (i.e., parasitic species that eventually kill their hosts) have evolved repeatedly and independently in several orders[23]. The most common and representative species are parasitoid wasps of the Hymenoptera, which are extremely diverse and are rapidly becoming a model for evolutionary biology with the accumulation of genomic data[24–31]. In addition, comparative genomics of parasitoid wasps and their free-living relatives may reveal novel evolutionary trait losses or regressions and provide new insights into the evolution of parasitism.

In parasitoid wasps, an intriguing trait loss with visible phenotypic change is the loss of yolk in the eggs of some species. The eggs with little or no yolk was first recorded in endoparasitoid *Chalcis sispes*. In 1942, Flanders termed such yolkless eggs “hydropic” due to the way that they

acquire the nutrients from host hemolymph, which distinguishes them from anhydropic eggs[32]. Moreover, the phenomenon of radiolabeled amino acids absorption in host hemolymph was found in *Microplitis croceipes*, but not the protein taken[33], implying that such yolkless eggs may have evolved to obtain the nutrients they need for development from other sources. More systematic research, mainly from a morphological point of view, then showed that this yolkless egg is widespread in parasitoid wasps from different taxa[33–37](Supplementary Table 1). And this trait loss was often linked to the evolution of parasitic strategies (shifts from ectoparasitism to endoparasitism), and was likely evolved in response to the development of embryos inside the hosts[23,38,39]. To uncover the genetic basis associated with the loss of egg yolk in parasitoid wasps, we carried out a comparative genomic analysis with vitellogenin gene (*Vg*, a conserved gene encodes a precursor of the egg yolk protein in insects[38,40]) as a breakthrough.

Results

Recurrent loss of the *Vg* gene in Hymenoptera is confined to endoparasitoid wasps

To identify *Vgs* and other homologous genes in Hymenoptera, we investigated the genomes of 64 representative species, covering the major lineages of Hymenoptera, such as sawflies, parasitoid wasps, ants and bees (Fig. 1b). Using a comprehensive homology-based search approach (Bitacora tool), followed by protein domain screening and manual curation, we obtained totally 237 genes, which included classical *Vgs*, and other homologous members (see Methods). According to the presence or absence of functional domains, we categorized these genes into three types, including 64 classical complete *Vgs* (*Vgs*), 169 *Partial*

Vgs (*PVgs*), and 4 *Vg likes* (*Vgls*) (Fig. 1a; Supplementary Table 2). The classical *Vg* encodes a protein with three major domains, namely Vitellogenin_N or LPD_N (PF01347), Vit_openb-sht (PF09172) and vWD (von Willebrand factor type D, PF00094), which plays a well-documented role as an egg yolk precursor in many insects[38,41]. And the *PVg* encodes other proteins containing Vitellogenin_N domains, while the *Vgl* encodes the remaining proteins without Vitellogenin_N domains. The molecular functions of these proteins in insects are still largely unknown. Our identification also corrected 7 *Vgs* that were incorrectly annotated in the previous genome annotation in 7 species (Supplementary Fig.1).

To place our analyses in an evolutionary context, we constructed a phylogenetic tree of the species examined using 1,116 single-copy orthologous genes, and then mapped the *Vg* annotation results onto it. Overall, we find variation in *Vg* copy number within Hymenoptera, and notably, we discovered the *Vg* gene was absent from the genomes of 24 parasitoid wasps from three superfamilies (Ichneumonoidea, Chalcidoidea and Cynipoidea) with diverse parasitoid species (Fig. 1b; Supplementary Table 2). In contrast, this gene was conserved as one to four copies in all other non-parasitoid hymenopterans, including a sawfly (*Athalia*), vespid wasps (*Vespa* and *Polistes*), ants (*Ooceraea*, *Atta*, and *Solenopsis*) and bees (*Colletes*, *Megalopta*, *Apis*, and *Bombus*). We noted that the *Vg* loss in the *Drosophila* parasitoid *Asobara japonica* (Braconidae) might be an artifact, because we found a *PVg* gene was located at the end of the scaffold (Supplementary Fig.2; pink arrow with question mark in Fig. 1b). In total, if we exclude the uncertain status in *A. japonica*, the loss of *Vg* has independently occurred in five different lineages of parasitoid wasps (totally 23 species involved). Specifically,

this gene was lost in the common ancestor of two *Leptopilina* wasps in superfamily Cynipoidea, the common ancestor of two *Trichogramma* wasps in superfamily Chalcidoidea, the common ancestor of *Diadegma semiclausum* and *Hyposoter didymator* (Ichneumonidae), the common ancestor of three aphid parasitoids (*Aphidius* and *Lysiphlebus*) in Braconidae, and the common ancestor of 14 braconid parasitoids from the genuses *Cotesia*, *Microplitis*, *Chelonus*, *Macrocentrus*, and *Eumacrocentrus* (Braconidae), respectively (pink arrows in Fig. 1b). In addition, we observed extensive gene copy number changes and gene losses in *PVg* and *Vgl* during Hymenoptera evolution, suggesting their complex evolutionary histories.

To further bridge the gap between *Vg* losses and parasitic strategies in parasitoid wasps, we divided the species into two categories according to where the wasp eggs are laid and where the embryos develop, namely ectoparasitoid and endoparasitoid, and then mapped this information to the phylogeny together with the *Vg* loss events. Ecto- and endoparasitoids are two well defined parasitic strategies in parasitoid wasps: ectoparasitoids usually lay eggs on the external surface of their hosts and their offspring feed on hosts from outside, whereas endoparasitoids lay eggs into internal bodies of their hosts and the embryos and larvae live inside the hosts[23]. One of the most fascinating differences between these two parasitic strategies is that the available nutrient status of the environment in which their embryos grow is completely different, i.e., the nutrients for embryonic development of ectoparasitoids are strictly in the eggs (mainly yolk proteins), while the access to the nutrients required for embryo development in endoparasitoids may be diverse, as the embryo is directly exposed to a nutrient-rich environment (insect hemolymph)[42,43]. Thus, considering the key role of Vg proteins in yolk formation (i.e.,

vitellogenesis)[40], we hypothesized that the losses of *Vg* should be restricted to endoparasitoids. Our results showed that all the parasitoid wasps with *Vg* loss are endoparasitoids, consistent with our expectation. We also noted that not all endoparasitoids (35% endoparasitoids retained *Vgs*) in our study have lost the *Vg*, suggesting endoparasitism may not be the main driving force for *Vg* loss in parasitoid wasps.

PVg does not result from _Vg_ degradation in the _Vg_ loss species and is unlikely to be involved in vitellogenesis

We argued that the loss of *Vg* may not directly result in vitellogenesis disruption in some cases, because the recently degraded *Vg* that retains the N-terminal domain (Vitellogenin_N) may still be functional due to the important contribution of the N-domain in Vg receptor binding for oocyte uptake[44,45]. To investigate the evolutionary origins of *PVgs* and to test whether some *PVgs* were arose by recent *Vg* degradation in the species without *Vg*, we extracted the protein sequences of all *Vgs* and *PVgs* identified in this study and performed a phylogenetic analysis. If the *PVgs* mainly originated from *Vg* degradation, the *PVg* branches should be largely mixed with *Vg* branches. However, our results showed that all *Vg* sequences have a different evolutionary history from most of *PVg* sequences (Fig. 2a). We also found that the N-terminal domain sequences of eight *PVgs* from seven species were located in the *Vg* clade, suggesting the these *PVgs* may result from *Vg* degradation. However, none of these species were involved in *Vg* loss events, with the exception of *A. japonica*, whose *Vg* loss was ambiguous as noted above.

The functions of *PVgs* remain uncharacterized. We examined their expression in different sexes of adults in 10 selected hymenopterans with available RNA-seq data from both sexes. Compared to *Vg*, which was

generally higher in females than in males, the expression of *PVg* was more male-biased (Fig. 2b; Supplementary Table 3). Therefore, our results indicate the distinct expression patterns between *Vgs* and *PVgs* in adult hymenopterans, further implying that the function of *PVgs* is likely to be different from that of classical *Vgs*.

Overall, these results reveal the dramatic differences between *Vgs* and *PVgs* from the point of view of evolution and gene expression. This further supports the hypothesis that vitellogenesis is largely arrested in the endoparasitoid wasps with *Vg* loss, consistent with previous reports about yolkless eggs[32,35] (see Supplementary Table 1).

Loss of *Vg* likely caused by genome rearrangement

To explore the potential causes of *Vg* loss in these endoparasitoid wasps, we first surveyed the genomes of *Vg* loss species and their *Vg* retaining relatives for regions syntenic with *Vgs*. Because *Vg* loss events are scattered across the Hymenoptera phylogeny over large evolutionary distances, we examined the syntenic regions of *Vg* in each family or superfamily respectively. In Ichneumonidae, we identified the genes coding chondroitin sulfate N-acetylgalactosaminyltransferase and alanine-glyoxylate aminotransferase, as the best marker genes for the *Vg* region. These two genes were flanked the *Vg* genomic region in the *Vg* retaining *Ophion* and *Venturia* wasps. However, when we searched for remnants of *Vg* (e.g., pseudogene, or inactivate gene with in-frame stop codons or frameshifting mutations) in the corresponding region in *Vg* loss *Hyposoter didymator*, we found no homologous sequences (Fig. 3a; see Methods). Similarly, in other three independent searches, we did not find any remnants of *Vgs* in other *Vg* loss species in Braconidae, Cynipoidea, and Chalcidoidea (Fig. 3b-c; Supplementary Fig.3). These results suggest

that the loss of *Vgs* may not be due to gene-inactivating mutations. An alternative explanation is that rapid sequence evolution has made the detection of homologous sequence challenging.

Next, we asked whether genome rearrangements had contributed to these *Vg* losses. To this end, we detected genome rearrangement breakpoints between *Vg* loss species and their *Vg* retaining relatives in three independent lineages (Cynipoidea, Chalcidoidea, and Braconidae) by whole genome alignments (Fig. 3d; see Methods). We excluded ichneumonid species from this analysis because their genome assemblies are fragmented (over 13,000 scaffolds for *Vg* loss species), which could affect the accuracy of genome-wide alignments. Totally, in Cynipoidea (Group 1), we identified 2,151 genome rearrangement breakpoints between *Leptopilina boulardi* (*Vg* loss) and *Belonocnema kinseyi* (*Vg* retaining), and we noticed that the breakpoints were significantly overlapped with *Vg* locus in the *B. kinseyi* genome (Fig. 4e; Observed breakpoints vs. a set of control regions by 10,000 times randomly shuffling breakpoints on the same chromosome ($p = 0$, one-tailed permutation test). Similar results were also obtained for Chalcidoidea and Braconidae (Fig. 4e). Taken together, our findings support the contribution of genome rearrangement to *Vg* loss events in the endoparasitoid wasps.

Gene loss and relaxed selection of *Vg* receptor gene after *Vg* loss

The evolutionary patterns of gene loss are not stochastic but often display a clear bias towards gene function or genomic position[46–48]. This pattern prompted us to investigate how the genes associated with *Vg* function evolve in the *Vg* loss lineages. Here, we first decided to focus on the *Vg* receptor gene *(VgR*), which is crucial for incorporation of Vg proteins

into oocytes in insects[38,40,41,49,50]. We hypothesized that, in the *Vg* loss species, their *VgRs* may also be absent, or if they are still present, they may be under different selective constraints. To test this hypothesis, we identified *VgR* genes in all 64 hymenopteran genomes examined in this study. In total, our analysis showed that *VgR* gene was absent in 19 hymenopteran genomes, and in other 45 genomes, this gene is kept as a single copy (Fig. 4a). Based on the presence or absence of *VgR*, we summarized four independent *VgR* loss events throughout the Hymenoptera evolution, which were highly correlated with *Vg* loss events (Fig. 4a). A high proportion (78%, 18/23) of *Vg* loss species (*A. japonica* excluded) also had lost *VgR*, and only five of them (including two *Leptopilina* species, two *Trichogramma* species, and *H. didymator*) still retained *VgR*. Interestingly, we found only one species (*Telenomus remus*) from superfamily Platygastroidea with a different pattern, who retained *Vg* but lost *VgR*. Next, we focused on the *Vg* loss lineages that retained *VgR* to test whether these *VgRs* evolved under relaxed constraints, since *Vg* losses are thought to weaken the selection acting on *VgRs* that was previously important for the maintenance of Vg-related functions. First, we screened for signatures of relaxed selection on all test branches (General test), under the RELAX model[51] in HyPhy[52]. As expected, the result revealed a significant relaxed selection pattern on these *VgRs* (Fig. 4b; K = 0, p = 0; K is a selection intensity parameter, K < 1 indicates relax selection while K > 1 indicates intensified selection). In addition, we also performed independent tests for each of the *Vg* loss lineage. The results also supported significant relaxed selection in the *Leptopilina* wasps (Fig. 4c; K = 0, p = 0) and *H. didymator* (Fig. 4e; K = 0.72, p = 0.004). However, we did not detect the signal of relaxed selection on the *Trichogramma* branches (Fig. 4d; K = 1.06, p = 0.916).

We also extended the analysis to *A. japonica*, which retained *VgR* but had ambiguous *Vg* loss, and we found no relaxed selection acting on the *VgR* (Fig. 4f; K = 1.21, p = 0.094). Overall, we describe the evolutionary trajectory for *Vg* and *VgR* in Hymenoptera. Our findings suggest that the loss of *Vg* genes in endoparasitoid wasps had subsequently resulted in gene loss or relaxed selection of *VgRs*, thus providing a novel example of co-elimination of functionally linked genes[46].

Structural variations in conserved domain of *Leptopilina* VgRs may lead to loss of the original function of the proteins

We next asked whether amino acid sequence changes due to relaxed selection would alter the three-dimensional structure of the VgRs, which in turn could change or eliminate the function of these proteins. Here, we predicted the structures of 45 VgRs identified in this study by AlphaFold2, and we only focused three well-structured regions (Part A, B, and C) according to the confidence scores (pLDDT) (Fig. 5b-d). Each region contains one YWTD repeated region (six YWTD motifs) and two epidermal growth factor (EGF)-like repeats (Fig. 5a). Our comparative structural analysis suggested that the structures of these three regions were generally conserved (Fig. 5b-d). Of note, we identified a portion of the Part A region of the two *Leptopilina* VgRs that differed substantially from all other VgRs (Fig. 5f), and this region overlapped with the second EGF-like repeat in the Part A region. Previously, studies have shown that mutation of the EGF-like repeat of VgR severely disrupts the process of yolk formation[53] and this functional domain is essential for Vg dissociation under acidic conditions[54]. Therefore, we hypothesized that structural variants in this EGF-like repeat region might lead to the loss of

the original function of the VgR protein in *Leptopilina* species, and this hypothesis has been demonstrated in a recent study[55].

Ecto-to-endo transition does not lead directly to the gene loss or relaxed constraint acting on *Vg* and *VgR*

The transition from ectoparasitism to endoparasitism is considered to be a key innovation in the evolutionary history of parasitoid wasps, and it is also a major trend in this process[23,56]. As noted above, although all the identified loss events of *Vgs* and *VgRs* were restrict to endoparasitoids, we show that there are still a number of endoparasitoids that retained these genes. This finding implies that ecto-to-endo transition is unlikely a main driving force for these gene losses. To gain deeper insights, we next tested whether the *Vgs* and *VgRs* in endoparasitoids evolve under relaxed purifying selection. If the transition to endoparasitism is indeed driving the *Vg* evolution, we expect to detect clear signals of relaxed selection on *Vgs* in other *Vg* retaining endoparasitoids compared to ectoparasitoids and free-living species. Following the similar logic of the *Vg* analysis, if there is a causal relationship between the ecto-to-endo transition and gene loss/relaxed selection of *VgRs*, we expect to find pervasive relaxed selection signals acting on the *VgRs* of endoparasitoids, not just the *VgRs* of *Vg* loss species as we reported above. However, in our tests, we found no evidence for the relaxed selection on the *Vgs* or *VgRs* in the endoparasitoids retaining *Vgs* (*Vg* test: K = 1.04, p = 0.561; *VgR* test: K = 1.21, p = 0), inconsistent with our hypothesis. Therefore, we conclude that the ecto-to-endo transition is not a main reason for the loss of *Vgs* and the subsequent loss and relaxed selection of *VgRs* in parasitoid wasps.

Shared genomic changes in the *Vg* loss lineages

To understand how parasitoid wasps evolve after *Vg* loss in different lineages and whether parallel or convergent genomic changes have occurred, we performed two genome-wide screens for gene changes that occurred in the *Vg* loss species. First, we investigated the changes in the size of gene families (OGs, Orthogroups). We identified expanded and contracted OGs in the oldest branches leading to the *Vg* loss lineages, and resulting in 39 to 462 expanded OGs and 120 to 713 contracted OGs in the five independent branches associated with *Vg* loss (Fig. 6a). Our comparative analysis showed that most OGs were branch-specific, with only two expanded and 40 contracted OGs likely to have undergone parallel changes (i.e., shared OGs between three or more lineages; Fig. 6b,c; Supplementary Table 4,5). Among these, we found that the expanded OGs included a multidrug resistance-associated protein, which is a member of ATP-binding cassette (ABC) transporters and is known to transport a number of different substrates across cellular membranes[57]. We inferred this parallel expanded OGs may enhance the ability of endoparasitoids to exploit host nutrients in the early development stage, to compensate for their lack of yolk nutrient availability. The parallel contracted OGs included lipase, protein spaetzle, farnesol dehydrogenase, and trypsin. In addition, we compared the functional annotations of the expanded or contracted OGs between the five lineages, and searched for the signature of functional convergent evolution (see Methods). Our results showed that the convergently expanded or contracted OGs (shared by four or more lineages) were mainly associated with transport (e.g., organic substance transport, monoatomic ion transport), metabolism (e.g., steroid metabolic process, amine metabolic process, and peptide

metabolic process), and development (e.g., embryo development and cuticle development) (Fig. 6d,e; Supplementary Table 6,7).

Second, we screened the 2,652 one-to-one orthologue protein-coding genes for signatures of positive/relaxed selection on the five branches leading to the *Vg* loss lineages (see Methods). We identified one unknown gene under positive selection on all five test branches, five genes under positive selection on four test branches, and 27 genes under positive selection on three test branches (Supplementary Table 8). These genes, which have diverse functions, included double-stranded RNA-specific editase, histone lysine N-methyltransferase, zinc finger protein, etc. We also identified 199 genes showing signals of relaxed selection on the five test branches, and enrichment analysis suggested that these genes were enriched in GO terms related to nervous system development, embryo development, locomotion, stress response ($p < 0.05$, FDR-adjusted; Supplementary Table 9,10, Supplementary Fig.4). Together, we found some parallel or convergent gene family changes and shared gene selection events across five independent *Vg* loss lineages, and these genomic changes may be related to adaptation to the endoparasitoid lifestyle, with no yolk available for the early embryo (Fig. 7).

Discussion

In this study, we utilized comparative genomics to explore the genetic basis of the origin of yolkless eggs in some endoparasitoid wasps, which is a long-standing puzzle in the evolutionary biology of parasitoid wasps. Due to this special phenotype (yolkless egg), we begin to focus on the *Vg* gene, which encodes a protein as an egg yolk precursor in nearly all oviparous animals including fish, birds, and insects[58]. To better understand *Vg* gene evolution in Hymenoptera, we annotated *Vg* genes

and their homologous members in 64 representative hymenopteran genomes. In addition, our phylogenetic and gene expression analyses further revealed the distinct evolutionary history and gene expression pattern between *Vg* and its homolog *PVg*, suggesting their divergent functions. The structure and function of *Vg* in insects have been well studied[38,50,59–62], whereas the roles of *PVgs* in *Apis mellifera* and *Formica* ants were reported differ from the function of conventional *Vg*, and these *PVg* might get involved in inflammation, oxidative stress or other sub-or neofunction[63,64]. Functional studies of *PVgs* still require more deeply investigation in the future.

We discovered notable *Vg* loss events (23 species involved) in the five independent lineages of endoparasitoid wasps from three superfamilies (Ichneumonoidea, Chalcidoidea, and Cynipoidea). Among the species with *Vg* loss, 13 wasps (*Aphidius ervi*, *Aphidius gifuensis*, *Cotesia congregata*, *Cotesia glomerata*, *Cotesia vestalis*, *Diachasma alloeum*, *Diadegma semiclausum*, *Fopius arisanus*, *Leptopilina boulardi*, *Lysiphlebus fabarum*, *Macrocentrus cingulum*, *Microplitis demolitor*, *Venturia canescens*) have been described in previous literature with the phenotype of yolkless eggs (see Supplementary Table 1). For the remaining 10 species there are no relevant egg morphology studies. In addition, we found no species containing the *Vg* gene that have historically been reported to have yolkless eggs. This result implies a high correlation between *Vg* loss and egg yolk disappearance. In addition, different from endoparasitoids, we found all ectoparasitoids and free-living hymenopterans retained *Vg* genes with variable copy number. This finding thus suggests that the *Vg* losses are only restrict to endoparasitoids, and might be associated with the ectoparasitoid habit (discussed below). We also performed genome alignment and synteny

analyses to search for the dead *Vg* genes, which could provide insights into the genetic mechanism of the gene loss. However, we did not see any obvious disablements, such as, insertions, deletions, frameshifts and truncations, even with the alignment e-value cutoff set to 1. We reason that this phenomenon may be due to the relatively old age of the gene losses that has occurred (24.32 to 124.58 million years ago) and the high evolutionary rates of hymenopteran genomes[65,66]. Interestingly, we found that, in three independent lineages, the *Vg* loci were often overlapped with genome rearrangement breakpoints, implying that a specific genome rearrangement at the *Vg* region is likely to be the cause of gene loss in these species (Fig. 7). More extensive collection of genomic data and identification of more recent *Vg* loss events will help to provide more robust evidence to support our findings on the mechanism of the gene loss.

Previous genomic studies have revealed an apparent functional bias in the evolution of gene loss[46]. And they found that this biased pattern was likely due to the co-elimination of functionally related genes, probably caused by the selection change on a specific trait adapted to the new living environment or new biological characteristics[46,67–69]. For example, eight skin-related genes were inactivated in both cetaceans and hippos, which may be associated with aquatic skin adaptations[70]; Losses of three renal transporter genes (*SLC22A12*, *SLC2A9*, and *SLC22A6*) and a renal ammonium secreting transporter gene (*RHBG*) were found in fruit bats, and these gene losses likely contributed to help fruit bats to efficiently excrete excess water in their diet[71]. In this study, we discovered co-elimination of *Vg* and *VgR* in 21 endoparasitoid wasps. Moreover, based on the current knowledge about the evolution of *Vg* and *VgR* in Hymenoptera, we proposed that the gene loss or relaxed selection of *VgR*

is likely a consequence of the *Vg* loss, probably because the disruption of vitellogenesis caused by *Vg* loss renders *VgR* dispensable (Fig. 7). Further genome sequencing works focused on endoparasitoid wasps from different families and superfamilies are needed for exploring the evolutionary history of the co-elimination of *Vg* and *VgR*. We also found that five of the *Vg* loss species retained the *VgR*, and the *VgRs* of the two *Leptopilina* wasps (*L. boulardi* and *L. heterotoma*) and *H. didymator* were found under significantly relaxed selection, and the structure of a functional domain were altered in *Leptopilina* species, suggesting their original functions as Vg receptor may have been lost. There are several reasons for the phenomenon of maintaining of *VgRs* in the genomes of *Vg* loss species. One hypothesis is that the relatively short evolutionary time is not enough to produce inactivating mutations or a complete loss of genes. An alternative hypothesis is that, although the original function has been lost, the *VgRs* may have evolved to gain a novel function. Consistent with the latter hypothesis, recent study of the *VgR* in *L. boulardi* (under relaxed selection) had shown that this gene no longer performed its original function related to ovary development, but was involved in host-searching and mating behavior[55].

It is reasonable to hypothesize that, the transition from ectoparasitism to endoparasitism may relax the selection regimes on the nutrition-related functions (such as yolk formation) at early developmental stages, as the source of the nutrients that parasitoid wasp embryos need for development has changed (i.e., only yolk nutrients are available to ectoparasitoid embryos, whereas nutrients from both yolk and host hemolymph are available to endoparasitoid embryos). Thus, the dispensability of yolk-related functions could lead to gene losses. However, our study found no evidence for direct selection between the

ecto-to-endo transition and the co-elimination of *Vg* and *VgR*. This result is also support by previous reports of functional Vgs in endoparasitoid wasps, such as *Pteromalus puparum*[72–75]. But we did notice the co-elimination of *Vg* and *VgR* was restrict to endoparasitoids. One explanation is that the endoparasitism opens up an alternative pathway for embryo development to obtain nutrients, which is a prerequisite for the co-elimination of *Vg* and *VgR*. In addition, gene loss is likely irreversible[71]. In particular, *Vg* and *VgR* are crucial for the formation of yolk that supports embryonic development, and the loss of these genes could lead to dependence on external nutrients, which in turn could hinder their ability to return to an ectoparasitoid state. Determining the mechanism (neutral or adaptive) underlying trait loss or regressive trait evolution remains challenging[46]. In our case, it is a difficult task to assess whether loss of yolk protein in endoparasitoids has a neutral or positive effect on fitness. The hypothesis for positive selection suggested that yolk loss could save energy costs and increase fecundity[76]; however, but this hypothesis has not been tested directly. Even if we attributed the loss of yolk protein in endoparasitoids to a neutral process, the genomic changes, such as transporter gene duplication we found by comparative genomics could be adaptive.

Finally, our study provides an example of loss of yolk protein genes in invertebrates. Another loss event of egg yolk protein genes (*VITs*) has been reported in mammalian evolution, coinciding with the origin of lactation and placentation[77]. Therefore, we propose that loss of egg yolk is likely a consequence associated with regressive evolution, which could occur repeatedly across animal evolution when an alternative nutrient source is available for the embryonic development. And it is interesting to investigate the loss of yolk in nematode species, especially

endoparasitic nematodes, because a study of the model species *Caenorhabditis elegans* has shown that the embryos can develop in the yolkless treatment[78].

Materials and methods

Data collection

Genomic data (including genomic sequences and genome annotation results) of 64 representative hymenopteran insects were mainly collected from publicly database, including the National Center for Biotechnology Information (NCBI)[79], InsectBase v2[80], and Darwin Tree of Life Project[81]. For RNA-seq analysis, we also collected RNA-seq data of 10 selected species from NCBI Sequence Read Archive (SRA) database. Find details about data collection in Supplementary Table 11.

Vg and *VgR* identification

We used Bitacora v1.4[82] to identify candidate *Vgs* and *VgRs*. Full mode who searches genes in genomic sequences and protein annotations was selected to identify gene candidates. Besides ncbi-blast v2.13.0+ [83]and hmmer v3.3.2[84], the implemented tool, GEMOMA v1.7.1[85], a homology-based gene prediction program that uses amino acid sequence and intron position conservation to reconstruct gene models from TBLASTN alignments was used to identify new genomic regions encoding gene families. We use 13 vitellogenin protein sequences of 11 species from Uniprot[86] or NCBI database as query reference (see Supplementary Table 11). And two refence sequences of vitellogenin receptor of *Drosophila melanogaster* and *Solenopsis invicta* were retrieved from Uniprot database (see Supplementary Table 12). The specific HMM profiles were generated with alignment files of these reference sequences using

hmmbuild command of hmmer v3.3.2[84]. Finally, each candidate gene was manually inspected, and further verified by checking their complete open reading frames. Then, the gene was divided into subgroups according to the following rule: we consider sequences with N-terminal domain (Vitellogenin_N or LPD_N) as classical *Vg* (*Vg*) and partial *Vg* (*PVg*), while the other homologous members without N-terminal domain were categorized as *Vg*-like sequence (*Vgl*). For the potential chimeric sequences (e.g., potential gene model errors), we extracted the upstream and downstream 2000 bp sequences of these genes for re-annotation, using Augusts web interface[87] (http://augustus.gobics.de/) or FGENESH+ service online[88] (http://www.softberry.com/).

Gene expression

We used STAR v2.7.10b[89] and RSEM v1.3.3[90] for transcriptomes assembly and expression calculation, respectively. Fragments Per Kilobase Million (FPKM) values were used to quantify gene expression.

Synteny analysis

Gene synteny of selected species in three family Ichneumonidae, family Braconidae and superfamily Cynipoidea were performed with JCVI toolkit (MCScan Python version 1.3.3) [91] based on the alignments of each pair of species generated by comparisons from LAST[92]. And high-quality blocks of the *Vg* gene and its flanking genes were extracted for visualization using a tool implemented in the JCVI toolkit.

Genomic rearrangement

We used LASTZ v1.04.03[93] to perform the pairwise genome alignments with genomes of *Vg*-loss and *Vg*-retained wasps. The *netSyntenic* files

were identified with axtChain/chainNet/netSyntenic tools based on resulting alignments, and as input to identify the genomic rearrangement breakpoints followed the workflow mentioned by Liao et al[94]. We conducted a shuffling test to determine whether the median distance between the randomly shuffled breakpoints and the *Vg* locus/loci was significantly different from what would be observed. The shuffled breakpoints were produced using shuffle function of bedtools[95] with constraint of breakpoints region on the same chromosome, and shuffling was repeated 10,000 times. The *p* value was calculated according to the method described by Sun et al.[96].

Gene Selection

We used HyPhy package v2.5.33[52] for detecting gene selection signals. Briefly, we used RELAX for testing relaxed selection on *Vgs* and *VgRs*. To test whether the *VgRs* of the *Vg* loss species were under relaxed selection, we first performed a general test in which all branches leading to the *Vg* loss species (excluding *A. japonica*) were set as test branches, while the remaining branches were set as reference. In addition, we also performed a separate test on each branch leading to the *Vg* loss species. To test whether the ecto-to-endo transition is a major driving force to relax the purifying selection on *Vgs* and *VgRs* and then eventually lead to their losses, we compared branches leading to endoparasitoid wasps (test) and non-endoparasitoid wasps (reference) for both *Vgs* and *VgRs*. In RELAX result, K was introduced as a parameter to assess the degree of relaxed selection, specifically $K < 1$ indicates that the selection was relaxed, while $K > 1$ indicates that the selection was strengthened. After the positive selection test, the value of dN/dS greater than 1 indicated that the test branches were positively selected, otherwise they were under

purifying selection. The *p* value was set at 0.05. We also searched for common positively/relaxedly selected genes on the branches leading to the *Vg* loss lineages, which may reflect the parallel or convergent genomic changes in the *Vg* loss lineages. aBSREL[97] was used for detecting positively selection genes, and RELAX was used for identifying genes evolving under relaxed selection ($p < 0.05$). In total, 2,652 one-to-one orthologue protein-coding genes from 17 species were involved in this selection analysis.

Orthology inference and phylogenetic reconstruction

We utilized OrthoFinder v2.5.5[98] to identify the orthologous and paralogous genes, the longest transcripts of genes were kept to analysis. The STAG (Species Tree Inference from All Genes) algorithm[99] that integrated in OrthoFinder to identify Orthogroups (OGs) for phylogenetic reconstruction. Based on the usage of "-M msa" parameter, genes from 1,116 OGs which were developed to leverage not only strictly single-copy orthologous genes but also the data from multi-copy gene families were used for species tree inference. Multiple sequence alignments (MSA) were inferred for each OG using MAFFT v7.520[100] and the species tree was inferred from the concatenated MSAs using IQ-TREE v2.2.2.3[101] with 1,000 replicates for ultrafast bootstrap analysis. The best-fitting model of sequence evolution estimated by ModelFinder that integrated in IQ-TREE v2.2.2.3[101] was used. And we used r8s v1.81[102] to estimate the divergence times between species or clade. The information of divergent time was previously reported[103] and we used two time points as fix age, 281 million years ago (mya) for Hymenoptera generated, and 8 mya for the divergent time of *Nasonia vitripennis* and *Pteromalus puparum*, and other 12 to constrain nodes marked by MCRA to calibrate

the phylogenetic tree: Orussidae-Pteromalinae: 211-289 mya, Apinae-Ampulicidae: 128-182 mya, *Apis mellifera-Bombus terrestris*: 93-132 mya, Apinae-Microgastrinae: 203-276 mya, Apinae-Gonatopodinae: 160-224 mya, Myrmicinae-Dorylinae: 65-127 mya, Vespinae-Polistinae: 34-71 mya, Cynipinae-Telenominae: 181-246 mya, Microgastrinae-Macrocentrinae: 82-147 mya, Microgastrinae-Aphidiinae: 116-177 mya, Microgastrinae-Campopleginae: 151-218 mya, Campopleginae-Cryptinae: 79-135 mya.

Gene family evolution

The gene family expansion and contraction were calculated using the CAFE v5[104]. The content of each OG of 64 species generated from OrthoFinder and the phylogenetic tree with divergence times inferred by r8s in above were used as input files. Based on the results, the expanded and contracted OGs of interested branches/clades that experienced *Vg* loss events were specifically extracted, and the shared OGs and associated function among these five branches/clades were inferred.

Gene ontology enrichment and functional similarity analysis

We annotated the amino acid sequences of each species using eggNOG-mapper v.2.1.12 webserver[105] (http://eggnog-mapper.embl.de.) against eggNOG5 database under default settings. And the biological process category of gene ontology (GO) terms was mainly focused to explore the most relevant functions in the interested OGs. In addition, GO enrichment analyses among candidate genes were performed with GOATOOLS[106] ($p < 0.05$) with the go-basic.obo database (http://geneontology.org/), and p values for each GO terms were corrected for an FDR < 0.05 (Benjamini-Hochberg method of multiple

testing correction). Using a simple clustering algorithm in REVIGO v1.8.1[107], semantic relationships of enriched GO terms were analyzed and representative subset of the terms was selected for visualization.

To further unscramble the functional convergence in five *Vg* loss linages, with the aid of the R package constellatoR (https://github.com/MetazoaPhylogenomicsLab/constellatoR), we firstly assigned the interested OGs (expanded and contracted) onto GO terms based on the functional similarity. And the function will be clustered with an exemplar OG with the distance matrices which were transferred from pairwise OG similarity matrices.

Protein structure analysis

We predicted the three-dimensional structures of VgR proteins from various species using LocalColabFold[108], which implements the AlphaFold algorithm[109]. The initial multiple sequence alignments (MSAs) for each protein were generated with the MMSeqs2 server[110]. We controlled the depth of the input MSAs by setting max_msa_clusters to 512 and max_extra_msa to 1024. All five neural networks were utilized for non-template-based predictions. The max_recycle parameter was set to 3 to balance accuracy and computational efficiency, without performing the post-prediction model refinements. From the predictions generated for each species, we selected the structure with the highest internal rank for subsequent analyses. Structural alignment was performed with MatchMaker[111] in ChimeraX[112]. A detailed list of sequences used for predictions is provided in the Supplementary Table 13. All figures were made with ChimeraX[112].

Acknowledgements

This work was supported by Key Program of National Natural Science Foundation of China (NSFC) (Grant No. 32330085 to G.Y.Y.), Program of NSFC (Grant No. 32202376 to X.H.Y.; Grant No. 32302337 to X.X.Z.), the Young Elite Scientists Sponsorship Program by China Association for Science and Technology (Grant No. 2022QNRC001 to X.H.Y.), the China Postdoctoral Science Foundation (Grant No. 2022M722804 to X.X.Z.), the Scientific Research Startup Fund Project of Zhejiang A&F University (Grant No. 2023LFR119 to X.H.Y.), the Program for Chinese Innovation Team in Key Areas of Science and Technology of Ministry of Science and Technology of the People's Republic of China (Grant No. 2016RA4008 to G.Y.Y.), the Fundamental Research Funds for Central Universities (Grant No. 2021FZZX001-31 to G.Y.Y.).

Author contributions

X.H.Y. and G.Y.Y. conceived and designed the study. X.X.Z., Y.Y.L., C.H., Y.Y., and X.H.Y. participated in comparative genomic analysis, gene selection, genome rearrangement. K.C.C., X.X.Z., and X.H.Y. performed VgR protein structure predictions. H.W.L. and Q.F. participated in data collection and discussion. X.X.Z. and X.H.Y. wrote the manuscript with input from all of the authors. All authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Ellers, J., Toby Kiers, E., Currie, C. R., McDonald, B. R. & Visser, B. Ecological interactions drive evolutionary loss of traits. *Ecol. Lett.* **15**, 1071–1082 (2012).

2. Lahti, D. C. *et al.* Relaxed selection in the wild. *Trends Ecol. Evol.* **24**, 487–496 (2009).

3. Sifuentes-Romero, I. *et al.* Trait loss in evolution: what cavefish have taught us about mechanisms underlying eye regression. *Integr. Comp. Biol.* **63**, 393–406 (2023).

4. Rétaux, S. & Casane, D. Evolution of eye development in the darkness of caves: adaptation, drift, or both? *EvoDevo* **4**, 26 (2013).

5. Leys, R., Cooper, S. J. B., Strecker, U. & Wilkens, H. Regressive evolution of an eye pigment gene in independently evolved eyeless subterranean diving beetles. *Biol. Lett.* **1**, 496–499 (2005).

6. Protas, M. E., Trontelj, P. & Patel, N. H. Genetic basis of eye and pigment loss in the cave crustacean, *Asellus aquaticus*. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5702–5707 (2011).

7. Emerling, C. A. & Springer, M. S. Eyes underground: Regression of visual protein networks in subterranean mammals. *Mol. Phylogenet. Evol.* **78**, 260–270 (2014).

8. Protas, M. E. *et al.* Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat. Genet.* **38**, 107–111 (2006).

9. Louchart, A. & Viriot, L. From snout to beak: the loss of teeth in birds. *Trends Ecol. Evol.* **26**, 663–673 (2011).

10. Sire, J.-Y., Delgado, S. C. & Girondot, M. Hen's teeth with enamel cap: from dream to impossibility. *BMC Evol. Biol.* **8**, 246 (2008).

11. Meredith, R. W., Zhang, G., Gilbert, M. T. P., Jarvis, E. D. & Springer, M. S. Evidence for a single loss of mineralized teeth in the common avian ancestor. *Science* **346**, 1254390 (2014).

12. Peng, C. *et al.* Large-scale snake genome analyses provide insights into vertebrate development. *Cell* **186**, 3519 (2023).

13. Xia, B. *et al.* The genetic basis of tail-loss evolution in humans and apes.

14. Poulin, R. & Randhawa, H. S. Evolution of parasitism along convergent lines: from ecology to genomics. *Parasitology* **142**, S6–S15 (2015).

15. Visser, B. *et al.* Loss of lipid synthesis as an evolutionary consequence of a parasitic lifestyle. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8677–8682 (2010).

16. Cai, L. *et al.* Deeply Altered Genome Architecture in the Endoparasitic Flowering Plant *Sapria himalayana* Griff. (Rafflesiaceae). *Curr. Biol.* **31**, 1002-1011.e9 (2021).

17. Guo, X. *et al.* The *Sapria himalayana* genome provides new insights into the lifestyle of endoparasitic plants. *BMC Biol.* **21**, 134 (2023).

18. Sun, G. *et al.* Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nat. Commun.* **9**, 2683 (2018).

19. Chen, X. *et al.* Balanophora genomes display massively convergent evolution with other extreme holoparasites and provide novel insights into parasite–host interactions. *Nat. Plants* **9**, 1627–1642 (2023).

20. International Helminth Genomes Consortium. Comparative genomics of the major parasitic worms. *Nat. Genet.* **51**, 163–174 (2019).

21. The Taenia solium Genome Consortium *et al.* The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* **496**, 57–63 (2013).

22. Cunha, T. J., De Medeiros, B. A. S., Lord, A., Sørensen, M. V. & Giribet, G. Rampant loss of universal metazoan genes revealed by a chromosome-level genome assembly of the parasitic Nematomorpha. *Curr. Biol.* **33**, 3514-3521.e4 (2023).

23. Pennacchio, F. & Strand, M. R. Evolution of developmental strategies in parasitic Hymenoptera. *Annu. Rev. Entomol.* **51**, 233–258 (2006).

24. Werren, J. H. *et al.* Functional and evolutionary insights from the genomes of three parasitoid *nasonia* species. *Science* **327**, 343–348 (2010).

25. Xiao, J.-H. *et al.* Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. *Genome Biol.* **14**, R141 (2013).

26. Park, D. *et al.* Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing. *BMC Genom.* **16**, 1 (2015).

27. Lindsey, A. R. I. *et al.* Comparative genomics of the miniature wasp and pest control agent *Trichogramma pretiosum*. *BMC Biol.* **16**, 54 (2018).

28. Ye, X. *et al.* A chromosome-level genome assembly of the parasitoid wasp *Pteromalus puparum*. *Mol. Ecol. Resour.* **20**, 1384–1402 (2020).

29. Xu, H. *et al.* Comparative Genomics Sheds Light on the Convergent Evolution of Miniaturized Wasps. *Mol. Biol. Evol.* **38**, 5539–5554 (2021).

30. Ye, X. *et al.* Genomic signatures associated with maintenance of genome stability and venom turnover in two parasitoid wasps. *Nat. Commun.* **13**, 6417 (2022).

31. Yang, Y. *et al.* Genome of the pincer wasp *Gonatopus flavifemur* reveals unique venom evolution and a dual adaptation to parasitism and predation. *BMC Biol.* **19**, 145 (2021).

32. Flanders, S. E. Oösorption and Ovulation in Relation to Oviposition in the Parasitic Hymenoptera. *Ann. Entomol. Soc. Am.* **35**, 251–266 (1942).

33. Ferkovich, S. M. & Dillard, C. R. A study of uptake of radiolabeled host proteins and protein synthesis during development of eggs of the endoparasitoid, *Microplitis croceipes* (Cresson) (Braconidae). *Insect Biochem.* **16**, 337–345 (1986).

34. Ralec, A. Egg contents in relation to host-feeding in some parasitic hymenoptera. *Entomophaga* **40**, 87–93 (1995).

35. Grbic, M. & Strand, M. R. Shifts in the life history of parasitic wasps correlate with pronounced alterations in early development. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1097–1101 (1998).

36. Donnell, D. M. Vitellogenin of the parasitoid wasp, *Encarsia formosa* (Hymenoptera: Aphelinidae): gene organization and differential use by members of the genus. *Insect Biochem. Mol. Biol.* **34**, 951–961 (2004).

37. King, P. E., Ratcliffe, N. A. & Fordy, M. R. Oögenesis in a braconid, *Apanteles glomeratus* (L.) possessing an hydropic type of egg. *Z. Zellforsch.* **119**, 43–57 (1971).

38. Tufail, M. & Takeda, M. Molecular characteristics of insect vitellogenins. *J. Insect Physiol.* **54**, 1447–1458 (2008).

39. Fisher, R. C. Aspects of the physiology of endoparasitic Hymenoptera. *Biol. Rev.* **46**, 243–278 (1971).

40. Raikhel, A. S. & Dhadialla, T. S. Accumulation of yolk proteins in insect oocytes. *Annu. Rev. Entomol.* **37**, 217–51 (1992).

41. Wu, Z., Yang, L., He, Q. & Zhou, S. Regulatory mechanisms of vitellogenesis in insects. *Front. Cell Dev. Biol.* **8**, 593613 (2021).

42. Pennacchio, F., Caccia, S. & Digilio, M. C. Host regulation and nutritional exploitation by parasitic wasps. *Curr. Opin. Insect. Sci.* **6**, 74–79 (2014).

43. Caccia, S. *et al.* Nutrient absorption by *Aphidius ervi* larvae. *J. Insect Physiol.* **51**, 1183–1192 (2005).

44. Anderson, T., Levitt, D. & Banaszak, L. The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. *Structure* **6**, 895–909 (1998).

45. Du, X. *et al.* Functional characterization of Vitellogenin_N domain, domain of unknown function 1943, and von Willebrand factor type D domain in vitellogenin of the non-bilaterian coral *Euphyllia ancora*: Implications for emergence of immune activity of vitellogenin in basal metazoan. *Dev. Comp. Immunol.* **67**, 485–494 (2017).

46. Albalat, R. & Cañestro, C. Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016).

47. Koludarov, I. *et al.* Domain loss enabled evolution of novel functions in the snake three-finger toxin gene superfamily. *Nat. Commun.* **14**, 4861 (2023).

48. Sharma, V., Hecker, N., Walther, F., Stuckas, H. & Hiller, M. Convergent losses of tlr5 suggest altered extracellular flagellin detection in four mammalian lineages. *Mol. Biol. Evol.* **37**, 1847–1854 (2020).

49. Schneider, W. J. Vitellogenin receptors: oocyte-specific members of the low-density lipoprotein receptor supergene family. *Int. Rev. Cytol.* **166**, 103–137 (1996).

50. Sappington, T. W. & S. Raikhel, A. Molecular characteristics of insect vitellogenins and vitellogenin receptors. *Insect Biochem. Mol. Biol.* **28**, 277–300 (1998).

51. Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L. & Scheffler, K. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832 (2015).

52. Kosakovsky Pond, S. L. *et al.* HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol. Biol. Evol.* **37**, 295–299 (2020).

53. Lin, Y. *et al.* Vitellogenin receptor mutation leads to the oogenesis mutant phenotype “scanty vitellin” of the silkworm, *Bombyx mori*. *J. Biol. Chem.* **288**, 13345–13355 (2013).

54. Davis, C. G. *et al.* Acid-dependent ligand dissociation and recycling of LDL receptor mediated by growth factor homology region. *Nature* **326**, 760–765 (1987).

55. Sheng, Y. *et al.* The vitellogenin receptor gene contributes to mating and host-searching behaviors in parasitoid wasps. *iScience* **26**, 106298 (2023).

56. Sharanowski, B. J. *et al.* Phylogenomics of Ichneumonoidea (Hymenoptera) and implications for evolution of mode of parasitism and viral endogenization. *Mol. Phylogenet. Evol* **156**, 107023 (2021).

57. Locher, K. P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat. Struct. Mol. Biol.* **23**, 487–493 (2016).

58. Chen, J.-S., Sappington, T. W. & Raikhel, A. S. Extensive sequence conservation among insect, nematode, and vertebrate vitellogenins reveals ancient common ancestry. *J. Mol. Evol.* **44**, 440–451 (1997).

59. Leipart, V. *et al.* Identification of 121 variants of honey bee Vitellogenin protein sequences with structural differences at functional sites. *Protein Sci.* **31**, e4369 (2022).

60. Leipart, V. *et al.* Resolving the zinc binding capacity of honey bee vitellogenin and locating its putative binding sites. *Insect Mol. Biol.* **31**, 810–820 (2022).

61. Leipart, V. *et al.* Structure prediction of honey bee vitellogenin: a multi-domain protein important for insect immunity. *FEBS Open Bio* **12**, 51–70 (2022).

62. Tufail, M., Nagaba, Y., Elgendy, A. M. & Takeda, M. Regulation of vitellogenin genes in insects. *Entomol. Sci.* **17**, 268–282(2014)

63. Salmela, H. *et al.* Ancient duplications have led to functional divergence of vitellogenin-like genes potentially involved in inflammation and oxidative stress in honey bees. *Genome Biol. Evol.* **8**, 495–506 (2016).

64. Morandin, C. *et al.* Not only for egg yolk—functional and evolutionary insights from expression, selection, and structural analyses of *Formica* ant vitellogenins. *Mol. Biol. Evol.* **31**, 2181–2193 (2014).

65. Dam, M. H. V. *et al.* The Easter Egg Weevil (*Pachyrhynchus*) genome reveals syntenic patterns in Coleoptera across 200 million years of evolution. *PLOS Genet.* **17**, e1009745 (2021).

66. Thomas, G. W. C. *et al.* Gene content evolution in the arthropods. *Genome Biol.* **21**, 15 (2020).

67. Aravind, L., Watanabe, H., Lipman, D. J. & Koonin, E. V. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11319–11324 (2000).

68. Blomme, T. *et al.* The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol.* **7**, R43 (2006).

69. Koonin, E. V. *et al.* A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**, R7 (2004).

70. Springer, M. S. *et al.* Genomic and anatomical comparisons of skin support independent adaptation to life in water by cetaceans and hippos. *Curr. Biol.* **31**, 2124–2139.e3 (2021).

71. Sharma, V. *et al.* A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nat. Commun.* **9**, 1215 (2018).

72. Dong, S. Z. *et al.* Vitellin of *Pteromalus puparum* (Hymenoptera: Pteromalidae), a pupal endoparasitoid of *Pieris rapae* (Lepidoptera: Pieridae): biochemical characterization, temporal patterns of production and degradation. *J. Insect Physiol.* **53**, 468–477 (2007).

73. Guo, J. Y. *et al.* Oosorption in the endoparasitoid, *Pteromalus puparum*. *J. of Insect Sci.* **11**, 90 (2011).

74. Ye, G. Y. *et al.* Molecular cloning and developmental expression of the vitellogenin gene in the endoparasitoid, *Pteromalus puparum*. *Insect Mol. Biol.* **17**, 227–233 (2008).

75. Dong, S. Z., Ye, G. Y., Guo, J. Y., Yu, X. P. & Hu, C. Oogenesis and programmed cell death of nurse cells in the endoparasitoid, *Pteromalus puparum*. *Microsc. Res. Tech.* **73**, 673–680 (2010).

76. Heimpel, G. E. & Collier, T. R. The evolution of host-feeding behaviour in insect parasitoids. *Biol. Rev.* **71**, 373–400 (1996).

77. Brawand, D., Wahli, W. & Kaessmann, H. Loss of egg yolk genes in mammals and the origin of lactation and placentation. *PLoS Biol.* **6**, e63 (2008).

78. Van Rompay, L., Borghgraef, C., Beets, I., Caers, J. & Temmerman, L. New genetic regulators question relevance of abundant yolk protein production in *C. elegans*. *Sci. Rep.* **5**, 16381 (2015).

79. Sayers, E. W. *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **50**, D20–D26 (2021).

80. Mei, Y. *et al.* InsectBase 2.0: a comprehensive gene resource for insects. *Nucleic Acids Res.* **50**, D1040–D1045 (2022).

81. The Darwin Tree of Life Project Consortium. Sequence locally, think globally: The Darwin Tree of Life Project. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2115642118 (2022).

82. Vizueta, J., Sánchez-Gracia, A. & Rozas, J. bitacora: A comprehensive tool for the identification and annotation of gene families in genome assemblies. *Mol. Ecol. Resour.* **20**, 1445–1452 (2020).

83. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).

84. Eddy, S. R. Accelerated profile HMM searches. *PLOS Comput. Biols* **7**, e1002195 (2011).

85. Keilwagen, J., Hartung, F., Grau, J. GeMoMa: Homology-Based Gene Prediction Utilizing Intron Position Conservation and RNA-seq Data. In *Gene Prediction: Methods and Protocols, Methods in Molecular Biology*. vol. 1962, Humana, New York, NY. (ed Kollmar, M.). 161-177 (Springer Science+Business Media, 2019).

86. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).

87. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467 (2005).

88. Salamov, A. A. & Solovyev, V. V. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516–522 (2000).

89. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

90. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011).

91. Tang, H. *et al.* Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).

92. Kiełbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).

93. Harris, R. S. Improved pairwise alignment of genomic DNA. (Pennsylvania State University, 2007).

94. Liao, Y., Zhang, X., Chakraborty, M. & Emerson, J. J. Topologically associating domains and their role in the evolution of genome structure and function in *Drosophila*. *Genome Res.* **31**, 397–410 (2021).

95. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

96. Sun, Y. H. *et al.* Amniotes co-opt intrinsic genetic instability to protect germ-line genome integrity. *Nat. Commun.* **14**, 812 (2023).

97. Smith, M. D. *et al.* Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353 (2015).

98. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

99. Emms, D. M. & Kelly, S. STAG: Species Tree Inference from All Genes. Preprint at https://www.biorxiv.org/content/10.1101/267914v1 (2018).

100. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

101. Minh, B. Q. *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).

102. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).

103. Peters, R. S. *et al.* Evolutionary history of the Hymenoptera. *Curr. Biol.* **27**, 1013–1018 (2017).

104. Mendes, F. K., Vanderpool, D., Fulton, B. & Hahn, M. W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518 (2021).

105. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).

106. Klopfenstein, D. V. *et al.* GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.* **8**, 10872 (2018).

107. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLOS ONE* **6**, e21800 (2011).

108. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

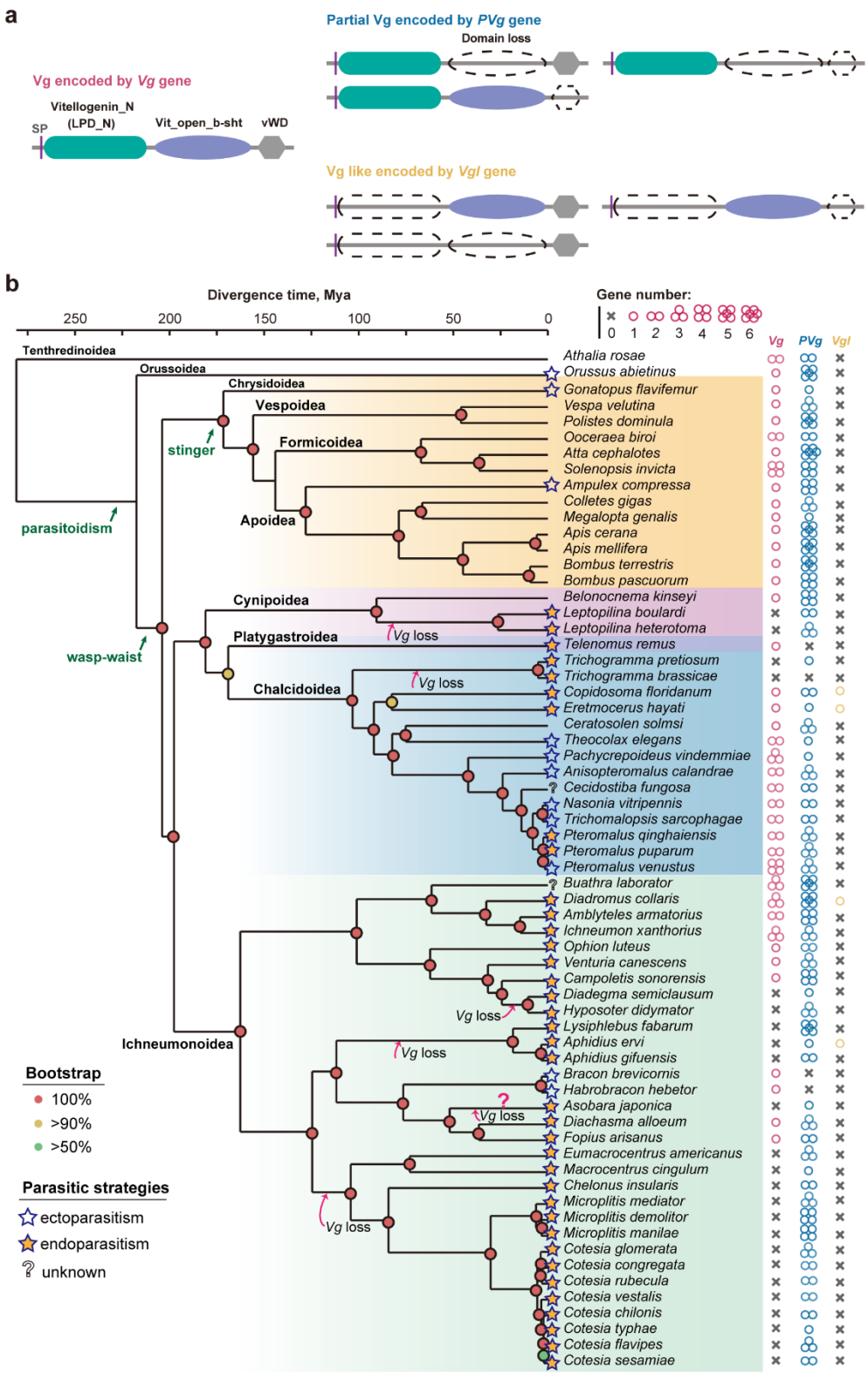
109. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

110. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).

111. Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C. & Ferrin, T. E. Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinform.* **7**, 339 (2006).

112. Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).

968 **Figures**



969

Fig 1 | Evolutionary analyses of vitellogenin (*Vg*) and *Vg*-related genes identify *Vg* loss in at least five independent hymenopteran lineages. (a) Domain schematics of vitellogenin protein (Vg, encoded by *Vg* genes) and other proteins with sequence similarity to Vg (encoded by *Pvg* and *Vgl* genes). The Vg comprises with three major domains named Vitellogenin_N (LPD_N) (green), Vit_open_b-sht (purple) and von Willebrand factor type D (vWD) (gray). The other Vg-related proteins with different types of domain losses are further divided into two categories based on the presence or absence of Vitellogenin_N domain, namely "Partial Vg" and "Vg-like". SP, signal peptide. (b) Phylogeny of hymenopteran insects with gene numbers of *Vg* and *Vg*-related genes, indicating at least five independent *Vg* losses. Phylogenetic tree of 64 hymenopterans was reconstructed by IQTREE2 using the protein sequences from 1,116 orthologous genes. Divergence times were estimated using 12 calibration points. The colored solid circle on each tree node indicates the bootstrap value. The star locates beside the species represented different parasitic strategies of the parasitoid wasps, and the question mark represents the unknown strategies. The number of hollow circles adjacent to the tree indicates the number of *Vg* and *Vg*-related genes while the cross refers to the absence of genes. The loss of *Vg* in *A. japonica* is ambiguous (marked by question mark in pink) because this loss could be an artefact caused by genome assembly problems (Supplementary Fig.2).

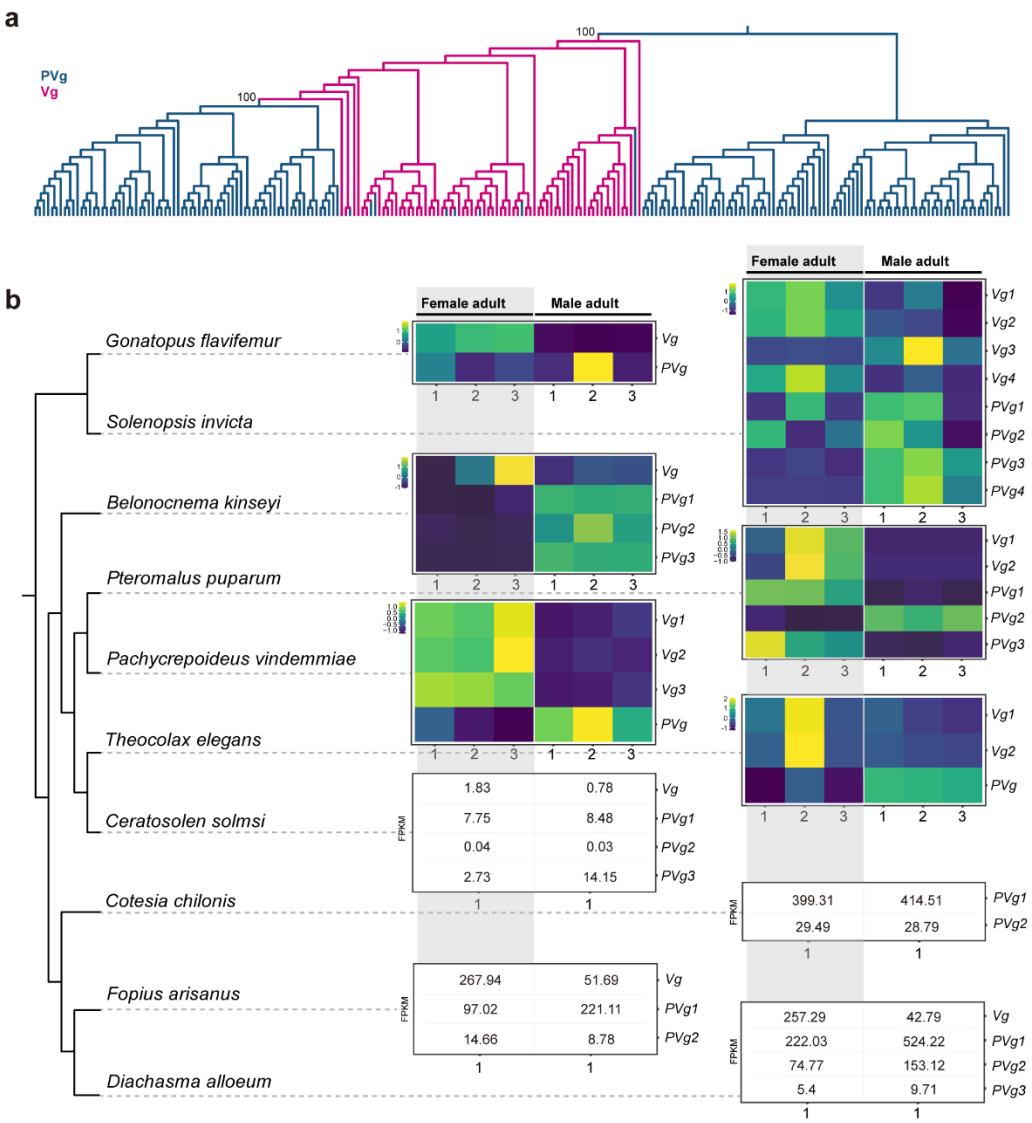


Fig 2 | Phylogenetic and expression analyses suggest distinct function and evolutionary history between *Vg* and *PVg* genes. (a) Phylogenetic tree of *Vgs* and *PVgs* in Hymenoptera inferred using IQTREE2 maximum likelihood algorithm with 1000 ultrafast bootstrap replicates. *Vgs* were indicated with red line and *PVgs* were in blue. This phylogenetic analysis showed that all *Vgs* and a proportion of *PVgs* (~48%) have high bootstrap support for the grouping together (100%) to the exclusion of other *PVgs* which clustered together as a monophyletic group. (b) Expression patterns of *Vgs* and *PVgs* in female adults and male adults in 10 hymenopterans. For each species, each line in heatmap represents the expression of *Vg* or *PVg* genes, and multiple expression values were shown when replicated sequencing libraries were available. The FPKM value of single samples were directly showed on the white frames. Phylogeny of the 10 selected hymenopteran insects was extracted from Fig 1.

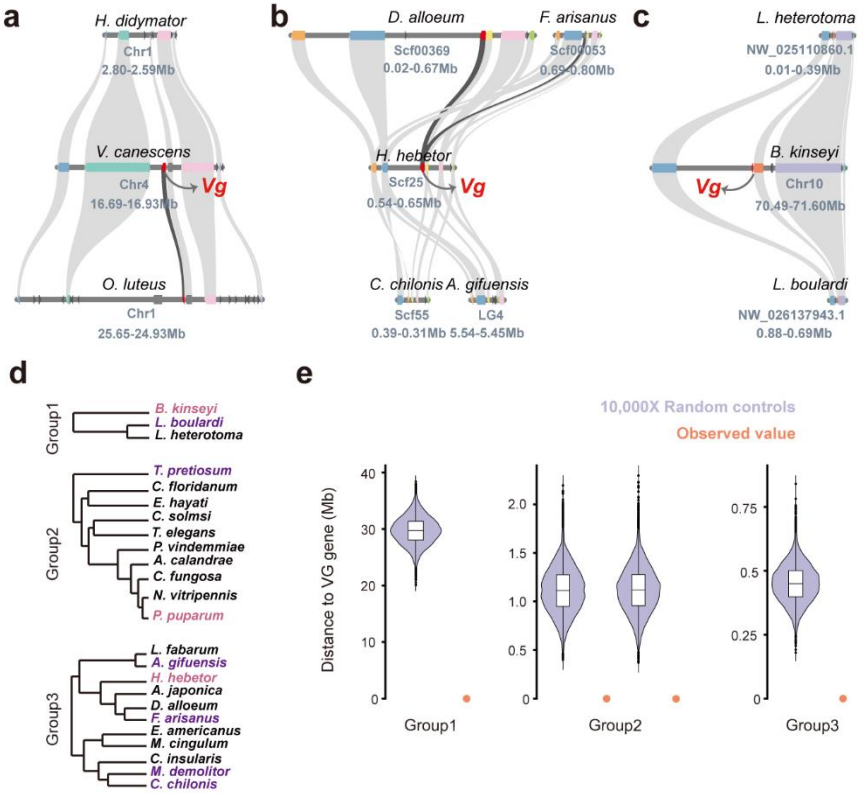
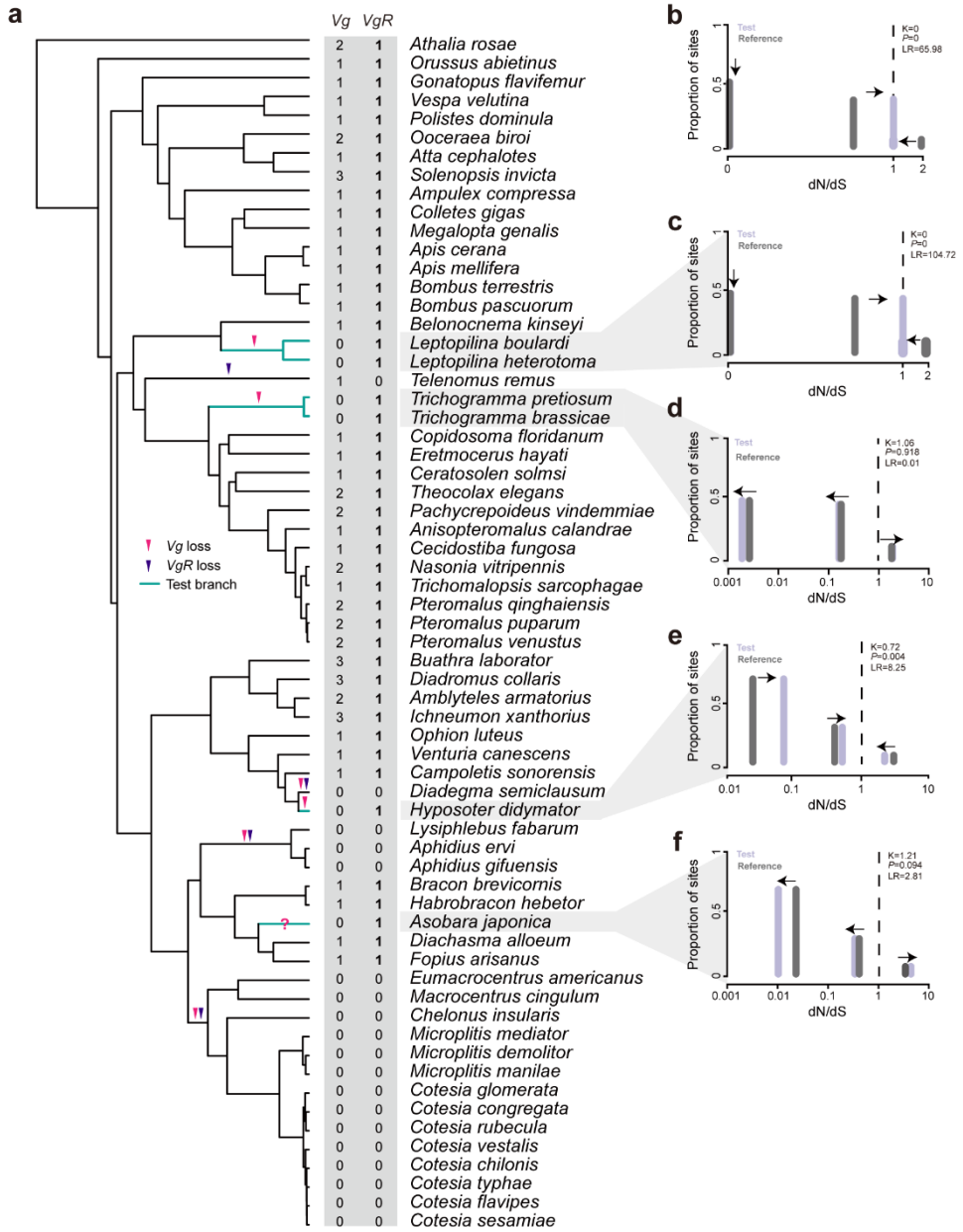


Fig 3 | Loss of *Vg* is likely due to genome rearrangement. Synteny of *Vg* gene and its flanking genes among representative *Vg*-loss species and *Vg*-retained species within family Ichneumonidae (a), family Braconidae (b) and superfamily Cynipoidea (c). *Vg* genes were highlighted in red, and the grey link represents the orthologous relationships. (d) Three selected groups representing three independent lineages with *Vg* loss to investigate the association between *Vg* loss and genome rearrangement. Group1 to 3 represent Cynipoidea, Chalcidoidea and Braconidae, respectively. In each testing group, pairwise genome alignments were performed to identify genomic rearrangement breakpoints, and species in pink who retain *Vg*, were considered as reference while species in purple were lack of *Vg* genes and used as query. (e) Distance of the observed genomic rearrangement breakpoints (orange point) and randomly shuffled control bins (purple) from *Vg* locus/loci in three testing groups. Violin plots represent the medians of the closest distances of randomly shuffled control bins to the *Vg* locus/loci that were computed 10,000 times.


a
Vg VgR
2 1 Athalia rosae
1 1 Orussus abietinus
1 1 Gonatopus flavifemur
1 1 Vespa velutina
1 1 Polistes dominula
2 1 Ooceraea biroi
1 1 Atta cephalotes
3 1 Solenopsis invicta
1 1 Ampulex compressa
1 1 Colletes gigas
1 1 Megalopta genalis
1 1 Apis cerana
1 1 Apis mellifera
1 1 Bombus terrestris
1 1 Bombus pascuorum
1 1 Belonocnema kinseyi
0 1 Leptopilina boulardi
0 1 Leptopilina heterotoma
1 0 Telenomus remus
0 1 Trichogramma pretiosum
0 1 Trichogramma brassicae
1 1 Copidosoma floridanum
1 1 Eretmocerus hayati
1 1 Ceratosolen solmsi
2 1 Theocolax elegans
2 1 Pachycrepoideus vindemmiae
1 1 Anisopteromalus calandrae
1 1 Cecidostiba fungosa
2 1 Nasonia vitripennis
1 1 Trichomalopsis sarcophagae
2 1 Pteromalus qinghaiensis
2 1 Pteromalus puparum
2 1 Pteromalus venustus
3 1 Buathra laborator
3 1 Diadromus collaris
2 1 Amblyteles armatorius
3 1 Ichneumon xanthorius
1 1 Ophion luteus
1 1 Venturia canescens
1 1 Campoletis sonorensis
0 0 Diadegma semiclausum
0 1 Hyposoter didymator
0 0 Lysiphlebus fabarum
0 0 Aphidius ervi
0 0 Aphidius gifuensis
1 1 Bracon brevicornis
1 1 Habrobracon hebetor
0 1 Asobara japonica
1 1 Diachasma alloeum
1 1 Fopius arisanus
0 0 Eumacrocentrus americanus
0 0 Macrocentrus cingulum
0 0 Chelonus insularis
0 0 Microplitis mediator
0 0 Microplitis demolitor
0 0 Microplitis manilae
0 0 Cotesia glomerata
0 0 Cotesia congregata
0 0 Cotesia rubecula
0 0 Cotesia vestalis
0 0 Cotesia chilonis
0 0 Cotesia typhae
0 0 Cotesia flavipes
0 0 Cotesia sesamiae
Vg loss
VgR loss
Test branch
b
Test
Reference
K=0
P=0
LR=65.98
Proportion of sites
dN/dS
c
Test
Reference
K=0
P=0
LR=104.72
Proportion of sites
dN/dS
d
Test
Reference
K=1.06
P=0.918
LR=0.01
Proportion of sites
dN/dS
e
Test
Reference
K=0.72
P=0.004
LR=8.25
Proportion of sites
dN/dS
f
Test
Reference
K=1.21
P=0.094
LR=2.81
Proportion of sites
dN/dS

1020

Fig 4 | Relaxed selection and completely loss of vitellogenin receptor (*VgR*) genes in many lineages with *Vg* loss. (a) The schematic tree on the left shows evolutionary relationships and is obtained from Figure 1b. Two columns of number refer to the *Vgs* and *VgRs* of each species listed on the right. The events of *Vg* loss (in pink) and *VgR* loss (in purple) upon branches are indicated with arrows. Branches in green were chosen as the test to performed the molecular selection analyses. The selection signatures of *VgR* genes in different branches or clades that without *Vg* were detected by RELAX model in Hyphy: (b) all *Vg* loss branches or clades, except *A. japonica*, $P = 0$, $K = 0$; (c) independent test of *Leptopilina* groups, $P = 0$, $K = 0$; (d) independent test of *Trichogramma* groups, $P = 0.981$, $K = 1.06$; (e) independent test of *H. didymator*, $P = 0.004$, $K = 0.72$; (f) independent test of *A. japonica*, $P = 0.094$, $K = 1.21$. *P* value was calculated using Likelihood-ratio test (LRT). The distribution of dN/dS across sites in the *VgRs* is illustrated by three categories of dN/dS for test (light purple) and reference (dark gray) branches. Neutral evolution is represented by the vertical dashed line at dN/dS = 1. Bars at dN/dS > 1 represent sites under positive selection, and bars at dN/dS <1 represent sites under purifying selection. The arrows show the direction of selection pressure in dN/dS between test and reference branches. K < 1 indicates selection was relaxed while K > 1 indicates selection was intensified.

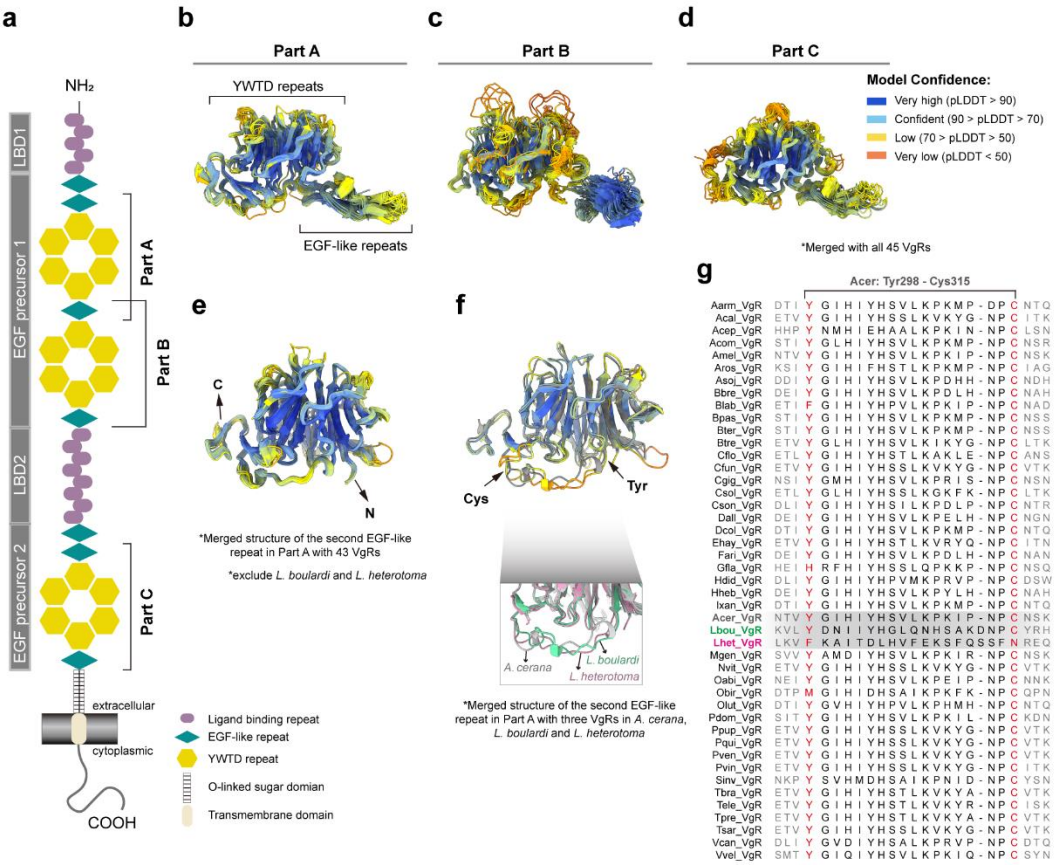


Fig 5 | Protein structure parsing of vitellogenin receptors (VgRs). (a) Domain schematics of the classical VgR protein. Ligand binding domain (LBD1, LBD2) and EGF precursor domain (EGF precursor1, EGF precursor2) were marked on the left. Three well-structured regions according to the confidence scores (pLDDT), Part A, B, C were labeled on the right. (b, c, d) Predicted protein structures of the Part A, B, and C of VgRs by AlphaFold2. Each figure showed the merged structure of the Part A, B, and C from 45 VgR sequences. Confidence scores (pLDDT) (between 0 and 100) of each residue produced by AlphaFold2 was indicated. (e) The merged structure of the Part A from 43 VgRs (exclude the sequences in *L. boulardi* and *L. heterotoma*) showing the highly conserved structure of the second EGF-like repeat of Part A. (f) Structure comparison further highlighted that structural variations of the second EGF-like repeat of Part A in *L. boulardi* and *L. heterotoma*. (g) Multiple sequence alignment of the second EGF-like repeat in Part A from all 45 VgRs, supporting the sequence variations in both *L. boulardi* and *L. heterotoma*.

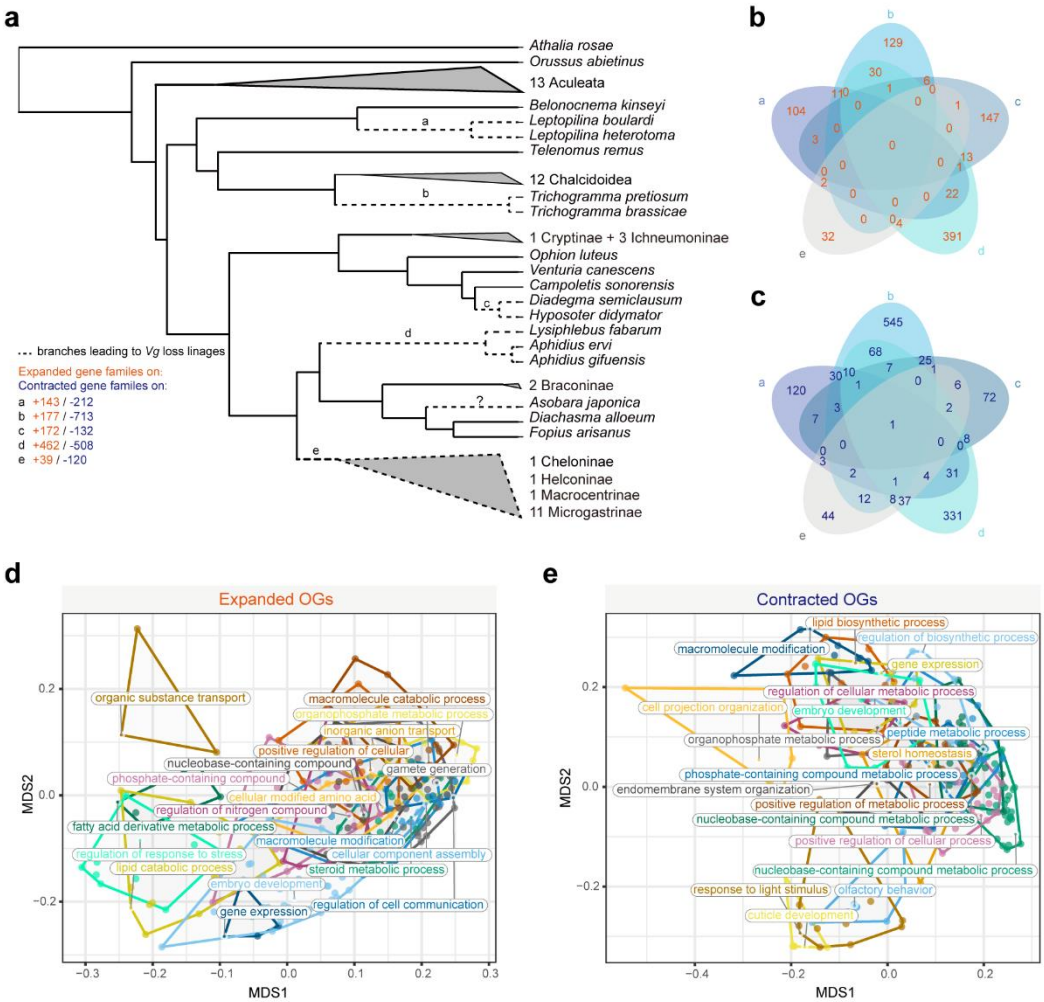


Fig 6 | Parallel or convergent gene family evolution in the *Vg* loss lineages. (a) Expansion and contraction of gene family (also called Orthogroup, OG) on the branches leading to the *Vg* loss lineages. Tree obtained from Figure 1b. Branches suffered *Vg* loss events were showed in dashed line and also marked as alphabets a-e. Venn diagrams showing the overlapping (b) expanded and (c) contracted OGs among the five independent *Vg* loss branches. (d, e) Constellation plots showing the functional convergence of expanded (d) and contracted OGs (e) among the five independent *Vg* loss branches (see Methods). Dots represent the expanded or contracted OGs shared in four or five independent lineages which lost *Vgs* and the centroid OG is pointed in polygons with putative function labeled. GO terms associated with transport, development and metabolism were functional converged.

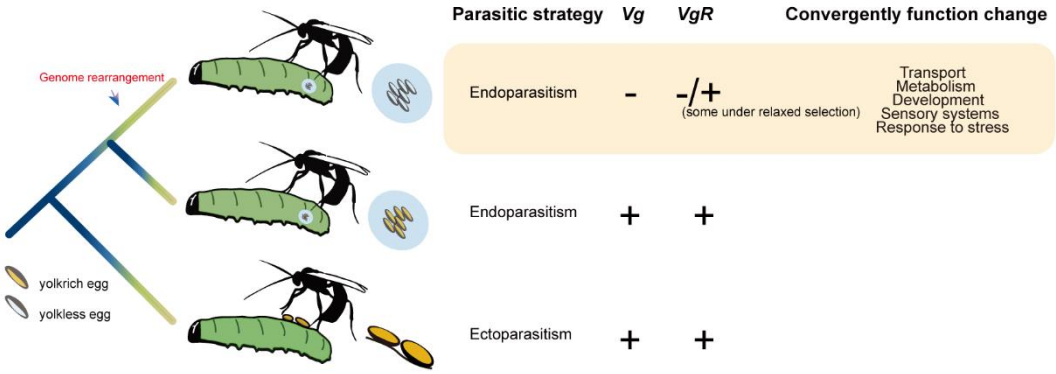


Fig 7 | Schematic of the evolution of *Vg* and *VgR* in parasitoid wasps. The ancestral parasitoid wasp is thought as an ectoparasitoid, retaining both the *Vg* and *VgR* genes and laying yolkrich eggs. During the evolution of Hymenoptera, shifts from ectoparasitism to endoparasitism have occurred many times independently. In endoparasitoid wasps, some species also retain *Vg* and *VgR* genes, and lay yolkrich eggs, such as *Pteromalus puparum*[72–75]. Of note, some endoparasitoid species have lost Vg, probably through genome rearrangements, and subsequently, their *VgRs* are also lost or under relaxed selection, likely due to the co-elimination of vitellogenesis system. And these wasps often produce yolkless eggs, according to previous literatures (see Supplementary Table 1). We also found a number of shared genomic changes related to transport, metabolism, development, sensory systems and stress resistance among five independent *Vg* loss lineages, which may be related to adaptation to the endoparasitoid lifestyle with no yolk available for the early embryo.