Genomic signatures associated with the evolutionary 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

Trait regression and loss have occurred repeatedly in numerous lineages throughout evolutionary history in response to changing environments. In parasitoid wasps, a mega-diverse group of hymenopteran insects, loss or reduction of yolk in the egg has been reported in many species. This phenotypic change likely evolved as a response to the shift from ectoparasitism to endoparasitism. However, the genetic basis of this trait and the impact of its loss on genome evolution remain poorly understood. Here, we performed a comparative genomic analysis of 64 hymenopteran insects. The conserved insect yolk protein gene *vitellogenin* (*Vg*) underwent five independent loss events in four families, involving 23 of the analyzed endoparasitoid species. Whole-genome alignment suggested that *Vg* loss occurred during genome rearrangement events. Analysis of *Vg* receptor gene (*VgR*) loss, selection, and structural variation in lineages

lacking *Vg* demonstrated functional biases in the patterns of gene loss. The ectoparasitism to endoparasitism transition did not appear to be the primary driver of *Vg* loss or the subsequent *VgR* evolution. A number of parallel and convergent genomic changes were observed between *Vg*-loss lineages, including gene family evolution and selection of genes related to transport, development, and metabolism. These changes may have facilitated embryonic development without the yolk in these lineages. Together, these findings reveal the genomic basis underlying a unique trait loss in parasitoid wasps. More broadly, this study enhances our understanding of yolk loss evolution outside the class Mammalia, highlighting a potential evolutionary trend arising from the availability of an alternative nutrient source for embryonic development.

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

Introduction

A fundamental goal in the field of evolutionary biology is to understand the genetic bases underlying phenotypic shifts and evolutionary novelties throughout the tree of life. Trait regression or loss (the reduction or complete loss, respectively, of an ancestral trait) are phenotypic shifts that have occurred repeatedly throughout evolutionary history. These shifts likely occur in response to alterations in the environment, habitat, or species interaction network[1–3]. Multiple genome sequencing and comparative genomic studies have been conducted to uncover the genetic bases of evolutionary trait loss in animals, including eye and pigmentation loss among animals living in darkness[4–8], tooth loss in birds[9–11], limb loss in snakes[12], and tail loss in humans and apes[13]. Such

studies have highlighted the importance of gene loss and inactivating mutations in the evolution of trait loss or regression. These investigations have also demonstrated the power of comparative genomic approaches to reveal the genetic mechanisms underlying complex trait evolution.

Trait loss and the associated genetic bases are topics of great research interest in the context of parasite evolution, particularly because long-term host dependence may lead to trait loss in processes such as nutrient acquisition[14,15]. For example, many parasitic plants show a loss of genes associated with photosynthesis, nutrient uptake, flowering regulation, defense responses, and root development[16–19]. A large-scale genomic analysis of parasitic nematodes and platyhelminths revealed the widespread losses of metabolic pathways and differential pathway coverage across different evolutionary clades[20]. Other studies have revealed the loss of homeobox gene families[21] and cilium-related genes[22] in parasitic worms, which may be related to their simplified morphology.

Insects are the largest and most diverse group of organisms on Earth. Parasitoids, defined as parasitic species that eventually kill their hosts, have evolved repeatedly and independently in several insect orders[23]. The most common and representative parasitoid species are parasitoid wasps of the order Hymenoptera; these are extremely biologically diverse, and a rapid accumulation of genomic data has made them model organisms in the field of evolutionary biology[24–31]. Comparative genomic analyses of parasitoid wasps and their free-living relatives can reveal novel evolutionary trait losses or regressions, providing new insights into the evolution of parasitism.

Some parasitoid wasp species exhibit an intriguing trait loss with a visible phenotype: loss of the egg yolk. Eggs with little or no yolk were first recorded in the endoparasitoid *Chalcis sispes*, and are widespread

among parasitoid wasps from various lineages[33–37] (Supplementary Table 1). Such yolkless eggs have been termed “hydropic”, referring to their method of nutrient acquisition from host hemolymph, which distinguishes them from anhydropic eggs[32]. Moreover, the phenomenon of radiolabeled amino acids absorption from host hemolymph was found in *Microplitis croceipes*, rather than protein taken[33], implying that yolkless insect eggs have evolved to obtain the nutrients necessary for development directly from the host. Indeed, yolk trait loss has been likely linked to the evolution of parasitic strategies (e.g., the shift from ectoparasitism to endoparasitism). Thus, yolk loss has likely evolved as a result of embryonic development inside the host[23,38,39], although this has not been proven.

To uncover the genetic basis of egg yolk loss in parasitoid wasps, we here conducted a comparative genomic analysis using *vitellogenin* (*Vg*), a conserved gene encoding an egg yolk protein precursor in insects[38,40]. Using a broad, representative selection of hymenopteran species, we identified lineages with *Vg* loss events and established a putative loss mechanism. Furthermore, we examined loss of a *Vg*-related gene to assess co-evolution in multiple lineages. These findings enhance our knowledge of yolk loss and parasitoid wasp genome evolution, contributing more broadly to an improved understanding of the genomic changes associated with shifts in parasitic strategies and embryonic developmental conditions.

Results

Vg loss was recurrent in Hymenoptera but confined to endoparasitoid wasps

We first sought to identify *Vg*s and homologous genes in members of the order Hymenoptera. Thus, a comparative genomic investigation was conducted in 64 representative hymenopteran species, including the major lineages: sawflies, wasps, ants, and bees (Fig. 1b). A comprehensive homology-based search was followed by protein domain screening and manual curation (see **Methods**), yielding a total of 237 genes. These genes were categorized into three groups based on the presence or absence of specific functional domains: 64 complete classic *Vg* genes (*Vg*s), 169 *Partial Vg* genes (*PVg*s), and four *Vg-like* genes (*Vgl*s) (Fig. 1a; Supplementary Table 2). The classic *Vg*s encoded proteins with three major domains[38,41], namely Vitellogenin_N or LPD_N (PF01347), Vit_openb-sht (PF09172), and von Willebrand factor type D (vWD) (PF00094); notably, we identified seven *Vg*s in seven species that were incorrectly annotated in prior genome annotations (Supplementary Fig. 1). The *PVg*s encoded other proteins containing the Vitellogenin_N domain, whereas *Vgl*s encoded proteins without the Vitellogenin_N domain but with the Vit_openb-sht and/or vWD domain.

To place our findings in an evolutionary context, a phylogenetic tree was constructed for the 64 species using 1116 single-copy orthologous genes. The *Vg* identification results were then mapped onto the resulting tree. This revealed a great deal of variation in *Vg* copy number within the order Hymenoptera. Interestingly, *Vg* was absent from the genomes of 24 parasitoid wasps in three superfamilies (Ichneumonoidea, Chalcidoidea and Cynipoidea) that contain diverse parasitoid species (Fig. 1b; Supplementary Table 2). In contrast, one to four copies of *Vg* were conserved in all other non-parasitoid hymenopteran species, including sawfly (*Athalia*), vespid wasps (*Vespa* and *Polistes*), ants (*Ooceraea*, *Atta*, and *Solenopsis*), and bees (*Colletes*, *Megalopta*, *Apis*, and *Bombus*).

140 *Vg* loss was presumed in the *Drosophila* parasitoid *Asobara japonica*
141 (Braconidae), but this may have been an artifact due to the presence of a
142 *PVg* gene located at the end of the scaffold (Supplementary Fig. 2; Fig.
143 1b).
144 *Vg* loss appeared to have occurred independently in five different
145 parasitoid wasp lineages, leading to the observed lack of *Vg* in 23 species
146 (excluding *A. japonica*). Specifically, *Vg* was lost in the common ancestor
147 of two *Leptopilina* species in the superfamily Cynipoidea; the common
148 ancestor of two *Trichogramma* species in the superfamily Chalcidoidea;
149 the common ancestor of *Diadegma semiclausum* and *Hyposoter*
150 *didymator* in the family Ichneumonidae; the common ancestor of three
151 aphid parasitoids of the genera *Aphidius* and *Lysiphlebus* in the family
152 Braconidae; and the common ancestor of 14 braconid parasitoids from
153 the genera *Cotesia*, *Microplitis*, *Chelonus*, *Macrocentrus*, and
154 *Eumacrocentrus* in the family Braconidae (Fig. 1b). In addition, the
155 phylogenetic tree demonstrated extensive variations in *PVg* and *Vgl* gene
156 copy number across hymenopteran species, suggesting complex
157 evolutionary histories.
158 To understand the relationships between *Vg* loss and parasitic
159 strategies, the 64 species were divided into three categories based on the
160 location in which their eggs are laid and embryos develop:
161 ectoparasitoids (those that lay eggs on the external host surface),
162 endoparasitoids (those that lay eggs inside the host)[23], and unknown.
163 These categorizations were then mapped to the phylogenic tree (Fig. 1b).
164 Based on the key role of Vg proteins in yolk formation (vitellogenesis)[40],
165 we hypothesized that *Vg* losses would be restricted to endoparasitoids due
166 to the nutrient-rich environment (host hemolymph) [42,43] in which the eggs
167 reside. Indeed, all of the parasitoid wasps with *Vg* loss were

168 endoparasitoids. However, not all of the analyzed endoparasitoids
169 demonstrated *Vg* loss; 35% retained copies of *Vg*. This suggested that
170 endoparasitism may not have been the primary factor driving *Vg* loss in
171 parasitoid wasps.
172
173 ***PVg* was not derived from *Vg* disruption in *Vg*-loss species**
174 We hypothesized that recently mutated *Vg*s retaining the Vitellogenin_N
175 domain (i.e., *PVg*s) may have retained some *Vg* functionality due to the
176 importance of the N-domain for oocyte uptake[44,45]. Thus, *Vg* loss may not
177 have directly resulted in a lack of vitellogenesis if *PVg* was present. To
178 investigate the evolutionary origins of *PVg*s and to test whether these
179 genes arose from recent *Vg* disruption among *Vg*-loss species, a
180 phylogenetic analysis was performed using all *Vg*s and *PVg*s identified in
181 this study. If the *PVg*s had primarily originated from *Vg* mutation, *PVg*s
182 would cluster together with *Vg*s. Instead, the Vg sequences showed a
183 distinct evolutionary history from the PVg sequences (Fig. 2a). Notably,
184 the N-terminal domain sequences of eight PVgs in seven species were
185 located in the Vg clade, suggesting that these specific PVgs may have
186 resulted from Vg disruption. However, none of the corresponding species
187 had undergone *Vg* loss events, with the exception of the ambiguous *Vg*
188 loss in *A. japonica* as noted above.
189 The functions of *PVg*s in insects have not yet been characterized. To
190 establish whether there was putative functional overlap between *Vg*s and
191 *PVg*s, we examined *Vg* and *PVg* expression in 10 hymenopteran species
192 for which there were publicly available RNA sequencing data generated
193 from adults of both sexes. *Vg* was generally expressed at higher levels in
194 females than in males, whereas the opposite was true of *PVg* (Fig. 2b;
195 Supplementary Table 3). These distinct expression patterns between *Vg*s

196 and *PVg*s in adult hymenopterans suggested that the genes have divergent
197 functionality. Overall, these results revealed dramatic differences between
198 *Vg*s and *PVg*s with respect to both evolutionary history and gene
199 expression. This supported the hypothesis that vitellogenesis was largely
200 arrested in endoparasitoid wasps with *Vg* loss, consistent with previous
201 reports of yolkless eggs[32,35] (Supplementary Table 1).
202
203 ***Vg* loss was likely caused by genome rearrangement**
204 To explore the potential causes of *Vg* loss in endoparasitoid wasps, the
205 genomes of species with *Vg* loss were examined for regions that were
206 syntenic with *Vg* loci in their *Vg*-retaining relatives. Because *Vg* loss
207 events span large evolutionary distances across hymenopteran phylogeny,
208 syntenic regions containing *Vg* were examined only within each family or
209 superfamily. In Ichneumonidae, the genes encoding chondroitin sulfate
210 N-acetylgalactosaminyltransferase and alanine-glyoxylate
211 aminotransferase were identified as marker genes for the *Vg* region; these
212 two genes flanked the genomic region containing *Vg* in wasps of the *Vg*-
213 retaining genera *Ophion* and *Venturia*. A search for remnants of *Vg* (e.g.,
214 a pseudogene or inactive gene) in the corresponding region of *H.*
215 *didymator*, which had lost *Vg*, revealed no homologous sequences (Fig.
216 3a; see **Methods**). Similar searches did not reveal *Vg* remnants in other
217 *Vg*-loss species of the families/superfamilies Braconidae, Cynipoidea, or
218 Chalcidoidea (Fig. 3b, c; Supplementary Fig. 3). These results suggested
219 that *Vg* loss was not due to gene-inactivating mutations. However, it is
220 also possible that rapid sequence evolution prevented homologous
221 sequence detection.
222 We next asked whether genome rearrangements had contributed to *Vg*
223 losses. This was addressed through detection of genome rearrangement

224 breakpoints between *Vg*-loss species and their *Vg*-retaining relatives in
225 three independent lineages (Cynipoidea, Chalcidoidea, and Braconidae)
226 via whole-genome alignment (Fig. 3d; see **Methods**). Ichneumonid
227 species were excluded from this analysis because the genome assemblies
228 were highly fragmented, with over 13,000 scaffolds for *Vg*-loss species,
229 which could have negatively affected the accuracy of genome-wide
230 alignments. In Cynipoidea (Group 1), 2151 genome rearrangement
231 breakpoints were identified between *Leptopilina boulardi* (*Vg* loss) and
232 *Belonocnema kinseyi* (*Vg* retaining). Notably, the breakpoints
233 significantly overlapped with the *Vg* locus in the *B. kinseyi* genome (Fig.
234 4e). Similar results were observed for species in Chalcidoidea and
235 Braconidae (Fig. 4e; Observed breakpoints vs. a set of control regions by
236 10,000 times randomly shuffling breakpoints on the same chromosome, p
237 = 0, one-tailed permutation test). Taken together, these findings indicated
238 strong contributions of genome rearrangement to *Vg* loss events in
239 endoparasitoid wasps.
240

241 **Gene loss and relaxed selection on the *Vg* receptor gene (*VgR*) after**
242 ***Vg* loss**

243 Evolutionary patterns of gene loss are usually not stochastic, but rather
244 display a clear bias with respect to gene function or genomic position[46–48].
245 This fact prompted us to investigate the mechanisms by which genes
246 associated with *Vg* function evolved in *Vg*-loss lineages. We first focused
247 on *VgR*, which is crucial for incorporation of Vg proteins into insect
248 oocytes[38,40,41,49,50]. We hypothesized that *VgR*s may be absent or subject to
249 distinct evolutionary pressure in *Vg*-loss species. To test this hypothesis,
250 *VgR* genes were identified in the genomes of all 64 examined
251 hymenopteran species. *VgR* was absent from 19 of the genomes and was

252 present as a single-copy gene in the other 45 (Fig. 4a). Based on *VgR*
253 presence or absence, we identified four independent *VgR* loss events
254 throughout hymenopteran evolution; these were strongly correlated with
255 *Vg* loss events (Fig. 4a). A high proportion of *Vg*-loss species, excluding
256 *A. japonica*, had also lost *VgR* (78%), with only five species retaining
257 *VgR* (namely two *Leptopilina* species, two *Trichogramma* species, and *H.*
258 *didymator*). Interestingly, only one species (*Telenomus remus* of the
259 superfamily Platygastroidea) retained *Vg* but lost *VgR*.
260 We next examined whether *VgR* was under relaxed selective pressure
261 after *Vg* was lost. Using the selection intensity parameter K, where $K < 1$
262 indicates relaxed selection and $K > 1$ indicates strong purifying selection,
263 we screened for signatures of relaxed selection on all test branches
264 leading to all *Vg*-loss *VgR*-retaining species, under the RELAX model[51]
265 in HyPhy[52]. As expected, the result revealed a significant relaxed
266 selection pattern on these *VgR*s (Fig. 4b; $K = 0$, $p = 0$). Independent tests
267 for each of the *Vg*-loss lineages indicated significant relaxed selection on
268 *VgR* in *Leptopilina* wasps (Fig. 4c; $K = 0$, $p = 0$) and in *H. didymator*
269 (Fig. 4e; $K = 0.72$, $p = 0.004$), but not in *Trichogramma* spp. (Fig. 4d; K
270 $= 1.06$, $p = 0.916$) or *A. japonica* (Fig. 4f; $K = 1.21$, $p = 0.094$). Overall,
271 these findings suggested that *Vg* loss in endoparasitoid wasps resulted in
272 subsequent loss or relaxed selection on *VgR*, a novel example of co-
273 elimination of functionally linked genes[46].
274
275 **Structural variations in the conserved VgR domain of *Leptopilina***
276 **may have led to functional loss**
277 Relaxed selection often leads to amino acid substitutions that alter the
278 three-dimensional structures (and therefore functions) of the encoded
279 proteins. To determine whether this was true of VgRs in *Vg*-loss lineages,

we predicted the structures of all 45 VgRs identified in this study with AlphaFold2. Based on the confidence scores (pLDDT), we focused only on the three most well-structured regions, designated A, B, and C (Fig. 5b–d). Each region contained one YWTD repeat (six YWTD motifs) and two epidermal growth factor (EGF)-like repeats (Fig. 5a). Comparative analysis suggested that the structures of these three regions were generally conserved between species (Fig. 5b–d). Notably, a portion of region A in the two *Leptopilina* VgRs differed substantially from those of all other VgRs (Fig. 5f); this region overlapped with the second EGF-like repeat of region A. Previous studies have shown that mutations in the EGF-like repeat of VgR severely disrupt the process of yolk formation[53] and that this functional domain is essential for Vg dissociation under acidic conditions[54]. We therefore hypothesized that structural variants in this EGF-like repeat region may have led to loss of the original VgR function in *Leptopilina* species, consistent with the results of a recent study[55].

The transition from ectoparasitism to endoparasitism did not directly cause *Vg*/*VgR* loss or relaxed selection

The transition from ectoparasitism to endoparasitism is considered a key step and a major trend in the evolutionary history of parasitoid wasps[23,56]. As noted above, although *Vg* and *VgR* loss events were restricted to endoparasitoids, a number of endoparasitoids retained one or both genes. This implied that the transition from ectoparasitism to endoparasitism was not the primary force driving these gene losses. To gain deeper insights into the reasons for *Vg*/*VgR* loss, we next tested whether the *Vg*s and *VgR*s in endoparasitoids evolved under relaxed purifying selection. If the transition to endoparasitism drove *Vg* or *VgR* evolution, there should

be clear signals of relaxed selection on *Vg*s/*VgR*s in other *Vg*-retaining endoparasitoids compared to both ectoparasitoids and free-living species. However, the results showed no evidence of relaxed selection on *Vg*s or *VgR*s in *Vg*-retaining endoparasitoids (K = 1.04, $p = 0.561$ for *Vg*; K = 1.21, $p = 0$ for *VgR*). We therefore concluded that the ectoparasitism to endoparasitism transition was not the primary factor underlying either *Vg* loss or the subsequent loss/relaxed selection on *VgR*s in parasitoid wasps.

Changes throughout the genome were consistent in *Vg*-loss lineages

Two genome-wide screens were next conducted to identify genetic changes in *Vg*-loss species. These screens were designed to clarify specific evolutionary patterns after *Vg* loss in different parasitoid wasp lineages and to determine whether parallel or convergent genomic changes had occurred. We first investigated changes in the sizes of gene families, also known as orthogroups (OGs). Expanded and contracted OGs were identified in the most ancient branches leading to the *Vg*-loss lineages. In each of the five independent branches associated with *Vg* loss, there were 39–462 expanded OGs and 120–713 contracted OGs (Fig. 6a).

Comparative analysis showed that most changes in OG size were branch-specific; only two expanded and 40 contracted OGs were likely to have occurred in parallel (i.e., were shared between three or more lineages) (Fig. 6b, c; Supplementary Table 4, 5). The expanded OGs included a multidrug resistance-associated gene encoding a member of the ATP-binding cassette (ABC) transporter family that transports numerous substrates across the cellular membrane[57]. We inferred that parallel expansion of this OG may have enhanced the capacity of endoparasitoids to utilize host nutrients in the early development stage,

compensating for the lack of nutrients from the yolk. Parallel contracted OGs included lipase, spaetzle, farnesol dehydrogenase, and trypsin. Functional annotation was conducted for the expanded and contracted OGs between the five lineages to search for signatures of convergent functional evolution (see **Methods**). The results showed that convergently expanded or contracted OGs (those that were shared by four or all five lineages) were primarily associated with transport (e.g., organic substance transport or monoatomic ion transport), metabolism (e.g., steroid, amine, or peptide metabolic processes), and development (e.g., embryo or cuticle development) (Fig. 6d, e; Supplementary Table 6, 7).

The second genome-wide screen entailed analyzing the 2652 protein-coding one-to-one orthologous genes in the five ancestral branches leading to *Vg*-loss lineages for signatures of positive or relaxed selection (see **Methods**). One gene of unknown function was under positive selection in all five branches; five genes were under positive selection in four branches; and 27 genes were under positive selection in three branches (Supplementary Table 8). These genes were annotated with diverse functions, including “double-stranded RNA-specific editase”, “histone lysine N-methyltransferase”, and “zinc finger protein”. There were 199 genes with relaxed selection signals in all five branches. Functional enrichment analysis indicated that these genes were related to nervous system development, embryo development, locomotion, and stress responses (false discovery rate-adjusted $p < 0.05$; Supplementary Table 9, 10; Supplementary Fig. 4). Together, these results demonstrated parallel or convergent changes in specific gene families in addition to shared gene selection events across five independent *Vg*-loss lineages. These genomic changes may have been related to adaptation to the endoparasitic lifestyle (Fig. 7).

Discussion

In this study, we used a comparative genomic approach to explore the genetic basis of egg yolk loss in endoparasitoid wasps, the origin and spread of which constitutes a long-standing puzzle in the field of evolutionary biology. We focused our analysis on *Vg*, an egg yolk precursor gene found in nearly all oviparous animals, including fish, birds, and insects[58]. To better understand *Vg* evolution, we annotated *Vg*s and homologs (*PVg*s and *Vgl*s) in 64 representative hymenopteran genomes. Although Vg structures and functions have been well characterized[38,41,50,59–62], the molecular functions of PVgs and Vgls in insects are largely unknown. Here, phylogenetic and gene expression analyses revealed distinct evolutionary histories and expression patterns of *Vg* compared to *PVg*, suggesting divergent functionality of these genes. Similarly, in *Apis mellifera* and *Formica* ants, *PVg* reportedly functions in inflammation and/or oxidative stress processes rather than vitellogenesis due to sub-functionalization or neofunctionalization[63,64]. Further studies will be required in other species to clarify the functionality and evolutionary patterns of *PVg*s.

We discovered notable *Vg* loss events in 23 species from five independent lineages of endoparasitoid wasps in three superfamilies (Ichneumonoidea, Chalcidoidea, and Cynipoidea). Of the species with *Vg* loss, 14 (namely *Aphidius ervi*, *Aphidius gifuensis*, *Cotesia congregata*, *Cotesia glomerata*, *Cotesia vestalis*, *Diachasma alloeum*, *Diadegma semiclausum*, *Fopius arisanus*, *Leptopilina boulardi*, *Lysiphlebus fabarum*, *Macrocentrus cingulum*, *Microplitis demolitor*, *Microplitis mediator*, and *Venturia canescens*) have previously been identified as having yolkless eggs (Supplementary Table 1); relevant studies of egg

392 morphology have not been conducted for the remaining 9 species. None
393 of the species containing *Vg* have been reported to have yolkless eggs,
394 demonstrating a strong correlation between *Vg* loss and egg yolk
395 disappearance in the species analyzed here. In contrast to the
396 endoparasitoids, all of the ectoparasitoids and free-living hymenopterans
397 were found to retain at least one copy of *Vg*, demonstrating restriction of
398 *Vg* loss to endoparasitoids.

399 Genome alignment and synteny analyses were conducted to identify
400 remnants of *Vg* genes, which could provide insights into the evolutionary
401 mechanism of *Vg* loss. Even using a low alignment threshold, *Vg*
402 pseudogenes (e.g., those with clearly disabling insertions, deletions,
403 frameshifts, or truncations) were not identified. This was likely due to the
404 relatively distant timing of *Vg* loss (estimated at ~24.32–124.58 million
405 years ago) and the high evolutionary rates of hymenopteran genomes[65,66].
406 Interestingly, in three independent lineages, *Vg* loci were frequently
407 found at genome rearrangement breakpoints, implying that a genome
408 rearrangement event in the *Vg* region was likely the cause of gene loss in
409 many of these species (Fig. 7). Comparison of a greater number of
410 genomes and identification of recent *Vg* loss events would provide more
411 robust evidence supporting this proposed mechanism of *Vg* loss.

412 Previous genomic studies have revealed apparent functional biases in
413 gene loss events[46], likely resulting from co-elimination of functionally
414 related genes. This may be caused by changes in selective pressure on a
415 specific trait due to adaption to a new living environment or to new
416 biological characteristics[46,67–69]. For example, eight skin-related genes
417 have been inactivated in both cetaceans and hippopotami over the course
418 of evolution, potentially due to aquatic skin adaptations[70]. Fruit bats have
419 lost functionality of three renal transporter genes (*SLC22A12*, *SLC2A9*,

and *SLC22A6*) and a renal ammonium secreting transporter gene (*RHBG*), likely contributing to efficient excretion of the excess water obtained through their unique diet[71]. Similarly, we here discovered co-elimination of *Vg* and *VgR* in 21 endoparasitoid wasp species. Based on current knowledge of *Vg* and *VgR* evolution in Hymenoptera, we propose that loss or relaxed selection on *VgR* was likely a direct consequence of *Vg* loss, which rendered *VgR* dispensable (Fig. 7).

Despite the widespread loss of *VgR* among endoparasitoids, five *Vg*-loss species retained *VgR*, and the gene was under significantly relaxed selection in two *Leptopilina* wasp species (*L. boulardi* and *L. heterotoma*) and in *H. didymator*. The structure of one *VgR* functional domain was altered in the two *Leptopilina* species, suggesting a loss of the original VgR function. There are several hypotheses for *VgR* maintenance among *Vg*-loss species. First, the relatively short evolutionary time may not have been sufficient to produce inactivating mutations or complete gene loss. Second, *VgR*s may have undergone neofunctionalization. Consistent with the latter hypothesis, a recent study of *L. boulardi VgR*, which is under relaxed selection, showed that this gene no longer performs its original function in ovary development but is involved in host-searching and mating behavior[55]. The genomic sequences of additional endoparasitoid wasps from different families and superfamilies should be analyzed in the future to further explore the coevolutionary histories of *Vg* and *VgR*.

The transition from ectoparasitism to endoparasitism may have led to relaxed selection on genes associated with nutritional status during early developmental stages (e.g., vitellogenesis genes). This is because ectoparasitoid embryos have strict nutrient requirements of the eggs (primarily the yolk proteins), whereas endoparasitoid embryos are directly exposed to a nutrient-rich external environment (the insect host

448 hemolymph)[42,43]. Thus, the dispensability of yolk-related functions could
449 lead to the loss of related genes in endoparasitoids. However, we found
450 no evidence for direct selection or co-elimination of *Vg* and *VgR* as a
451 result of the ectoparasitism to endoparasitism transition. This was
452 supported by previous reports of functional Vg proteins in endoparasitoid
453 wasps such as *Pteromalus puparum*[72–75]. Nevertheless, we did find that
454 co-elimination of *Vg* and *VgR* was restricted to endoparasitoids. This may
455 have been because endoparasitism leads to an alternative method of
456 embryonic nutrient acquisition, which is a prerequisite for co-elimination
457 of *Vg* and *VgR*. These two genes are crucial for formation of the yolk,
458 which supports embryonic development in the absence of external
459 nutrient availability. Loss of these genes could therefore lead to
460 dependence on external nutrients, hindering a return to the ectoparasitoid
461 state due to the generally irreversible nature of gene loss[71].
462 Determination of the mechanism(s) underlying trait loss or regressive
463 trait evolution remains challenging[46]. Here, it proved difficult to assess
464 whether loss of the yolk protein in endoparasitoids had a neutral or
465 positive effect on fitness. The hypothesis for positive selection suggests
466 that yolk loss could save energy costs and increase fecundity[76], but this
467 hypothesis has not been directly tested. Even if loss of the yolk protein in
468 endoparasitoids was determined to be a neutral process, other genomic
469 changes, such as the transporter gene duplication identified through
470 comparative genomics, could be adaptive. Detailed functional studies of
471 closely related *Vg*-loss and *Vg*-retaining species could shed light on the
472 impacts of *Vg* loss on fitness. In mammals, loss of the gene encoding the
473 egg yolk protein (*VIT*) has previously been reported as coinciding with
474 the origin of lactation and placentation[77]. We therefore propose that loss
475 of the egg yolk is likely a consequence of regressive evolution, which

may occur repeatedly during animal evolution when an alternative nutrient source becomes available during embryonic development. A study of the model species *Caenorhabditis elegans* has shown that the embryos can develop without a yolk[78]. Thus, future studies should also investigate yolk loss in nematode species, especially endoparasitic nematodes.

In conclusion, we here conducted an extensive comparative genomic analysis of 64 hymenopteran species to study loss of the vitellogenesis gene *Vg*. We found evidence of five independent *Vg* loss events throughout the order Hymenoptera, primarily occurring in species with endoparasitic strategies. *Vg* loci in species that retained the gene were frequently found in regions corresponding to genome rearrangements in *Vg*-loss species, suggesting that loss of the gene was due to genome rearrangement. Furthermore, *Vg* loss often coincided with loss or relaxed selection of *VgR*. Overall, these findings enhance our understanding of genome evolution and embryonic development across an important order of insects, revealing key genomic changes associated with yolk loss.

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 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 *Vg*s and *VgR*s. 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 12). And two refence sequences of vitellogenin receptor of *Drosophila melanogaster* and *Solenopsis invicta* were retrieved from Uniprot database (see Supplementary Table 13). 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: 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*-retaining 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

558 HyPhy package v2.5.33[52] was used for detecting gene selection signals.
559 Indeed, aBSREL[97] was used for detecting positively selection genes while
560 RELAX model[51] was used for identifying genes evolving under relaxed
561 selection ($p < 0.05$, the Bonferroni-Holm corrected). In RELAX result, K
562 was introduced as a parameter to assess the degree of relaxed selection,
563 specifically K < 1 indicates that the selection was relaxed, while K > 1
564 indicates that the selection was strengthened. Whereas, in the result of
565 positive selection test, the value of dN/dS greater than 1 indicated that the
566 test branches were positively selected, otherwise they were under
567 purifying selection.
568 Briefly, we used RELAX for testing relaxed selection on *Vg*s and
569 *VgR*s. To test whether the *VgR*s of the *Vg* loss species were under relaxed
570 selection, we first performed a general test in which all branches leading
571 to the *Vg* loss species (excluding *A. japonica*) were set as test branches,
572 while the remaining branches were set as reference. In addition, we also
573 performed a separate test on each branch leading to the *Vg* loss species;
574 To test whether the ecto-to-endo transition is a major driving force to
575 relax the purifying selection on *Vg*s and *VgR*s and then eventually lead to
576 their losses, we compared branches leading to endoparasitoid wasps (test)
577 and non-endoparasitoid wasps (reference) for both *Vg*s and *VgR*s.
578 We also searched for common positively/relaxedly selected genes on
579 the branches leading to the *Vg* loss lineages, which may reflect the
580 parallel or convergent genomic changes in the *Vg* loss lineages. In total,
581 2652 one-to-one orthologue protein-coding genes from 17 species were
582 involved in this selection analysis.
583
584 **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 1116 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] 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[108,109] (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[110], which implements the AlphaFold algorithm[111]. The initial multiple sequence alignments (MSAs) for each protein were generated with the MMSeqs2 server[112]. 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[113] in ChimeraX[114]. A detailed list of sequences used for predictions is provided in the Supplementary Table 14. All figures were made with ChimeraX[114].

Data availability

All data used in this study were downloaded from public database and sources were listed in Supplementary Tables. The data supporting the findings of this study are available from the corresponding authors upon reasonable request.

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 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, JY., 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. Preprint at https://www.biorxiv.org/content/10.1101/2021.09.14.460388v1 (2021).

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. Tsai, I. *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**, 3514–521.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, JH. *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, JS., 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, SZ. *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, JY. *et al.* Oosorption in the endoparasitoid, *Pteromalus puparum*. *J. of Insect Sci.* **11**, 90 (2011).

74. Ye, GY. *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, SZ., Ye, GY., Guo, JY., Yu, XP. & 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 germline 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. *Mol. Biol. Evol.* **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. Balart-García, P. *et al*. Parallel and convergent genomic changes underlie independent subterranean colonization across beetles. *Nat. Commun.* **14**, 3842 (2023).

109. Aristide, L. & Fernández, R. Genomic innovation, convergence and parallelism facilitated the ecological transition to life out of the sea in molluscs. Preprint at https://doi.org/10.21203/rs.3.rs-2382076/v1 (2022).

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

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

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

113. 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).

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

Figures

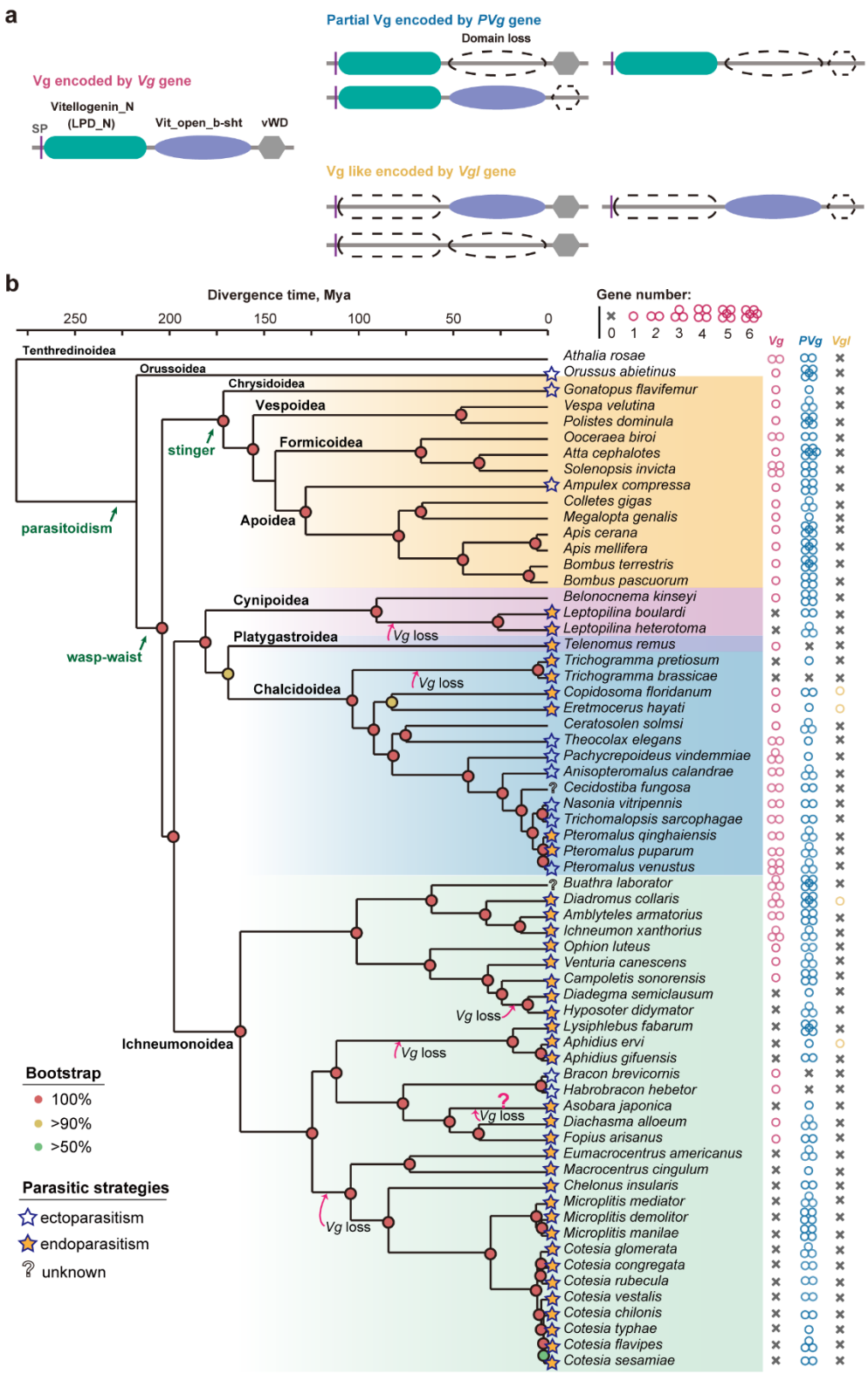

a
Vg encoded by Vg gene
Vitellogenin_N
(LPD_N)
SP
Vit_open_b-sht
vWD
Partial Vg encoded by PVg gene
Domain loss
Vg like encoded by Vgl gene
b
Divergence time, Mya
250
200
150
100
50
0
Gene number:
0 1 2 3 4 5 6
Vg
PVg
Vgl
Tenthredinoidea
Orussoidea
Chrysidoidea
Vespoidea
Formicoidea
Apoidea
Cynipoidea
Platygastroidea
Chalcidoidea
Ichneumonoidea
stinger
parasitoidism
wasp-waist
Vg loss
Athalia rosae
Orussus abietinus
Gonatopus flavifemur
Vespa velutina
Polistes dominula
Ooceraea biroi
Atta cephalotes
Solenopsis invicta
Ampulex compressa
Colletes gigas
Megalopta genalis
Apis cerana
Apis mellifera
Bombus terrestris
Bombus pascuorum
Belonocnema kinseyi
Leptopilina boulardi
Leptopilina heterotoma
Telenomus remus
Trichogramma pretiosum
Trichogramma brassicae
Copidosoma floridanum
Eretmocerus hayati
Ceratosolen solmsi
Theocolax elegans
Pachycrepoideus vindemmiae
Anisopteromalus calandrae
Cecidostiba fungosa
Nasonia vitripennis
Trichomalopsis sarcophagae
Pteromalus qinghaiensis
Pteromalus puparum
Pteromalus venustus
Buathra laborator
Diadromus collaris
Amblyteles armatorius
Ichneumon xanthorius
Ophion luteus
Venturia canescens
Campoletis sonorensis
Diadegma semiclausum
Hyposoter didymator
Lysiphlebus fabarum
Aphidius ervi
Aphidius gifuensis
Bracon brevicornis
Habrobracon hebetor
Asobara japonica
Diachasma alloeum
Fopius arisanus
Eumacrocentrus americanus
Macrocentrus cingulum
Chelonus insularis
Microplitis mediator
Microplitis demolitor
Microplitis manilae
Cotesia glomerata
Cotesia congregata
Cotesia rubecula
Cotesia vestalis
Cotesia chilonis
Cotesia typhae
Cotesia flavipes
Cotesia sesamiae
Bootstrap
100%
>90%
>50%
Parasitic strategies
ectoparasitism
endoparasitism
unknown

Fig. 1 | Evolutionary analyses of vitellogenin (*Vg*) and *Vg*-related genes. (**a**) Domain schematics of Vg and homologs, namely Partial Vg (PVg) and Vg-like (Vgl) proteins. Vg proteins had three major domains: Vitellogenin_N (LPD_N) (green), Vit_open_b-sht (purple), and von Willebrand factor type D (vWD) (gray). Other Vg-related proteins were classified as PVgs or Vgls based on the presence or absence, respectively, of the Vitellogenin_N domain. SP, signal peptide. (**b**) Phylogenic analysis of hymenopteran insects. The tree suggested the existence of at least five independent *Vg* loss events. Divergence times were estimated using 12 calibration points. The colored solid circle on each tree node indicates the bootstrap value. Stars to the left of the species name indicate the parasitic strategy of that species, endoparasitism (yellow) or ectoparasitism (white). Question marks to the left of the species name indicate an unknown parasitic strategy. The numbers of *Vg* and *Vg*-related genes in each genome are indicated at right; the pink question mark for *Vg* in *Asobara japonica* indicates ambiguity. This loss may have been an artifact from genome assembly (Supplementary Fig. 2).

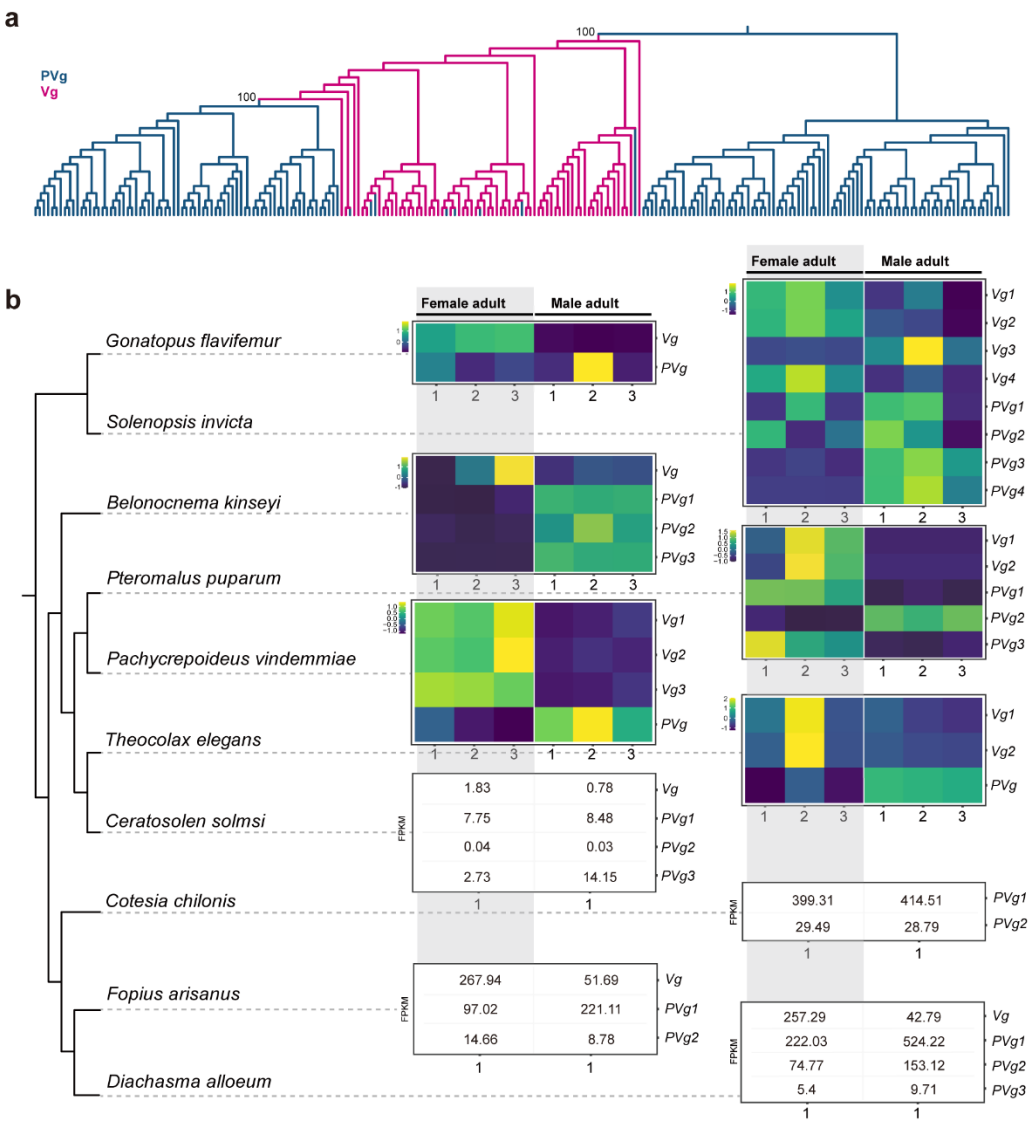


Fig. 2 | Phylogenetic and gene expression analyses of *Vg*s and *PVg*s. (**a**) Phylogenetic tree of *Vg*s (red) and *PVg*s (blue) in hymenopteran species. All of the *Vg*s clustered together with some of the *PVg*s (~48%); the remainder of the *PVg*s clustered together as a monophyletic group. (**b**) Expression patterns of *Vg*s and *PVg*s in male and female adults of 10 hymenopteran species. Columns indicate individual samples; each row represents expression of a *Vg* or *PVg* (as noted at right). Expression values are represented with color and were calculated in fragments per kilobase of transcript per million mapped reads (FPKM). For experiments without biological replicates, gene expression values are shown in numerical form. The phylogeny of the 10 selected hymenopteran insects was extracted from the data shown in **Fig. 1**.

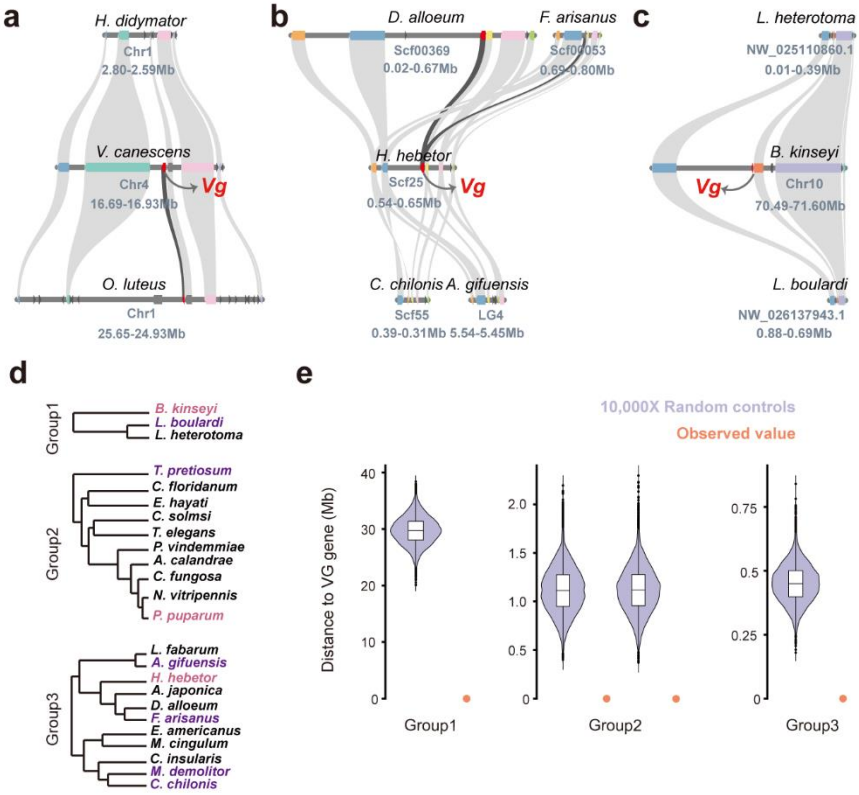
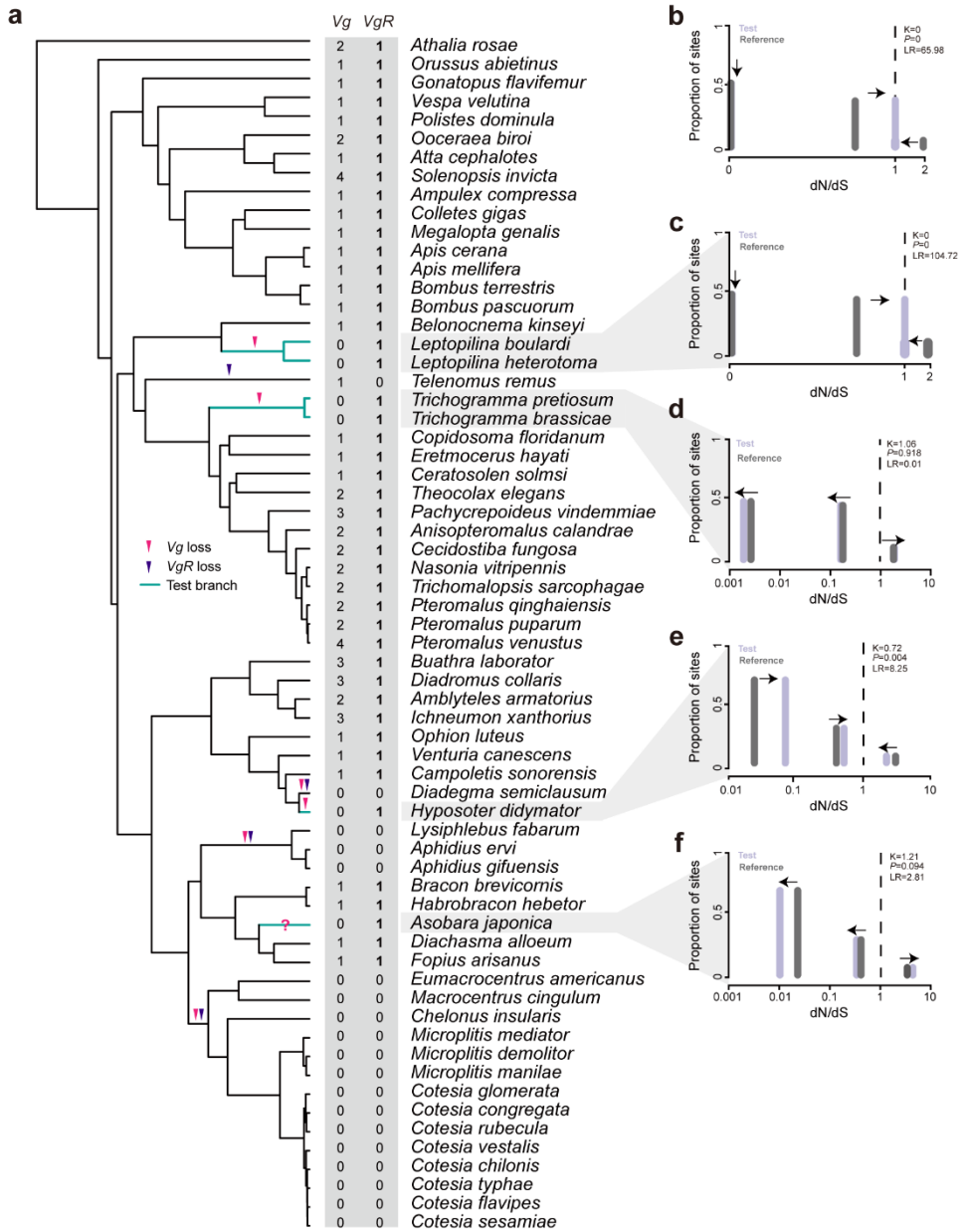


Fig. 3 | *Vg* was likely lost from endoparasitoid Hymenoptera due to genome rearrangement. (**a–c**) Synteny of *Vg* and the flanking genes among representative *Vg*-loss and *Vg*-retaining species within the (**a**) family Ichneumonidae, (**b**) family Braconidae, and (**c**) superfamily Cynipoidea. *Vg*s are indicated with red. Grey lines indicate orthologous relationships. (**d**) Phylogenetic analyses of three independent lineages with *Vg* loss. Group1, Cynipoidea; Group2, Chalcidoidea; Group3, Braconidae. Pairwise genome alignments were performed within each group to identify genomic rearrangement breakpoints. Species shown in pink and purple were *Vg*-retaining (reference) and *Vg*-loss (query) species, respectively. (**e**) Distance between *Vg* loci and observed genomic rearrangement breakpoints (orange points) or randomly shuffled control bins (purple violins).


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
4 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
3 1 Pachycrepoideus vindemmiae
2 1 Anisopteromalus calandrae
2 1 Cecidostiba fungosa
2 1 Nasonia vitripennis
2 1 Trichomalopsis sarcophagae
2 1 Pteromalus qinghaiensis
2 1 Pteromalus puparum
4 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

1002

Fig. 4 | Relaxed selection and complete loss of genes encoding the Vg receptor (*VgR*) in many lineages with *Vg* loss. (**a**) Phylogenetic tree (from **Fig. 1b**) showing instances of *Vg* and *VgR* loss (pink and purple arrows, respectively). Branches shown in green were used in molecular selection analyses. (**b–f**) Selection signatures of *VgR*s from independent analyses of several *Vg*-loss branches or clades, namely (**b**) all *Vg*-loss branches or clades, excluding *Asobara japonica*; (**c**) *Leptopilina* spp.; (**d**) *Trichogramma* spp.; (**e**) *Hyposoter didymator*; and (**f**) *A. japonica*. Nonsynonymous to synonymous substitution rate (dN/dS) distributions across sites in the *VgR*s are shown for test (branches listed above in **b–f**, light purple) and reference (branches left from test branches, dark grey) branches. dN/dS values of 1, > 1, and < 1 represent neutral evolution, positive selection, and purifying selection, respectively. Arrows show the direction of selective pressure between the test and reference branches. K values < 1 or > 1 indicate relaxed or intensified selection, respectively. The Bonferroni-Holm corrected *p*-values were showed.

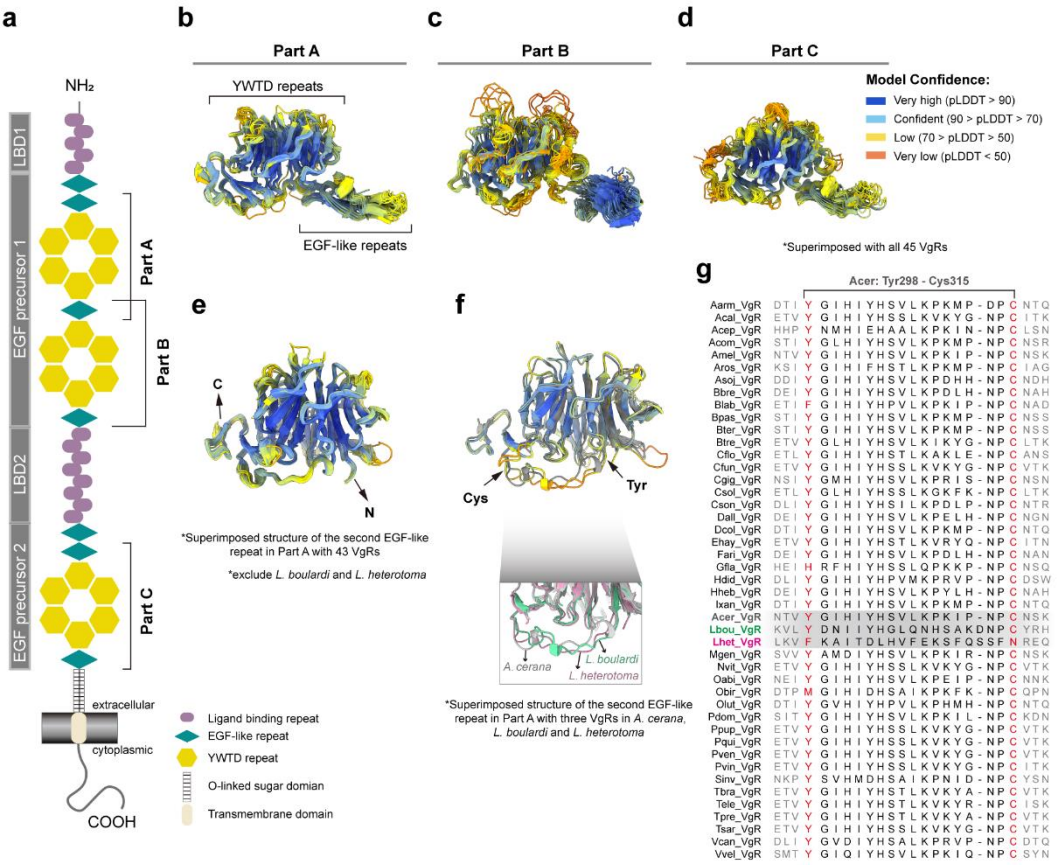


Fig. 5 | Predicted three-dimensional structures of vitellogenin receptors (VgRs). (**a**) Domain schematic of a classic VgR protein. The ligand binding domains (LBD1 and LBD2) and the EGF precursor domains (EGF precursor1 and EGF precursor2) are indicated at left. Three regions with well-defined structures based on the confidence (pLDDT) scores were designated regions A, B, and C (labeled at right). (**b–d**) Predicted structures of the VgR regions A, B, and C with AlphaFold2. Each structure is a superimposed representation based on 45 VgR sequences. pLDDT scores for each residue are indicated with color. (**e**) A superimposed structure of region A based on 43 VgRs (excluding *Leptopilina boulardi* and *L. heterotoma*) showing the highly conserved structure of the second EGF-like repeat. (**f**) Structural comparison showing variations in the second EGF-like repeat of region A in *L. boulardi* and *L. heterotoma*. (**g**) Multiple sequence alignment of the second EGF-like repeat of region A from all 45 VgRs, demonstrating the sequence variations in *L. boulardi* and *L. heterotoma*.

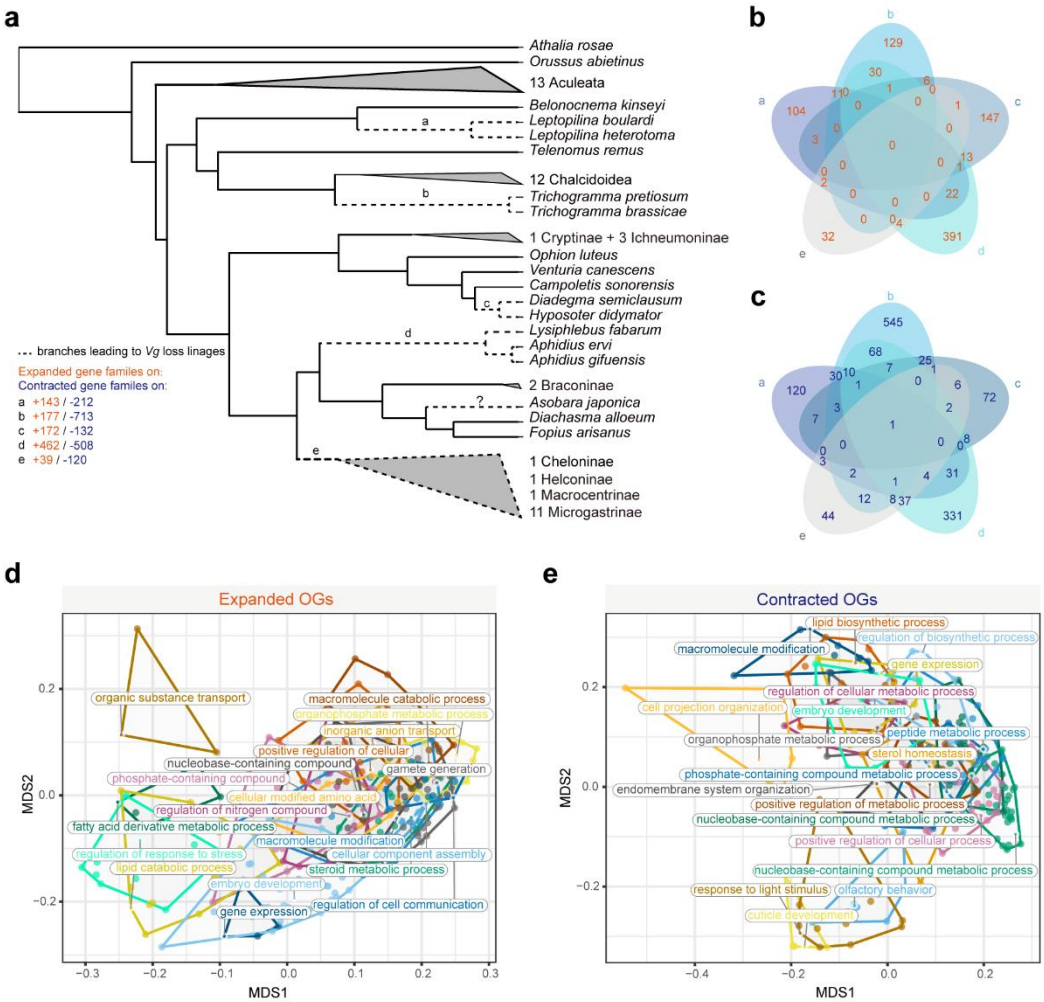


Fig. 6 | Parallel or convergent gene family evolution in the *Vg*-loss lineages. (**a**) Expansion and contraction of gene families, also known as Orthogroups (OGs) in branches leading to *Vg*-loss lineages. The phylogenetic tree was taken from **Fig. 1b**. The five branches affected by *Vg* loss events are indicated with dashed lines and marked a–e. (**b, c**) Unique and overlapping (**b**) expanded and (**c**) contracted OGs between the five independent *Vg*-loss branches. (**d, e**) Functional convergence of (**d**) expanded and (**e**) contracted OGs in the five independent *Vg*-loss branches (see **Methods**). Each point represents an expanded or contracted OG shared by four or five of the independent *Vg*-loss lineages. The centroid OGs were labeled with putative functions in each polygon. Gene Ontology terms associated with transport, development, and metabolism were functionally converged.

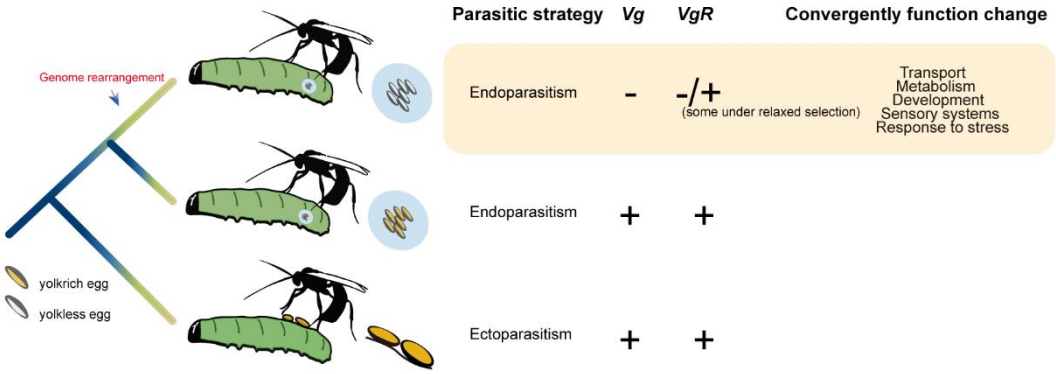
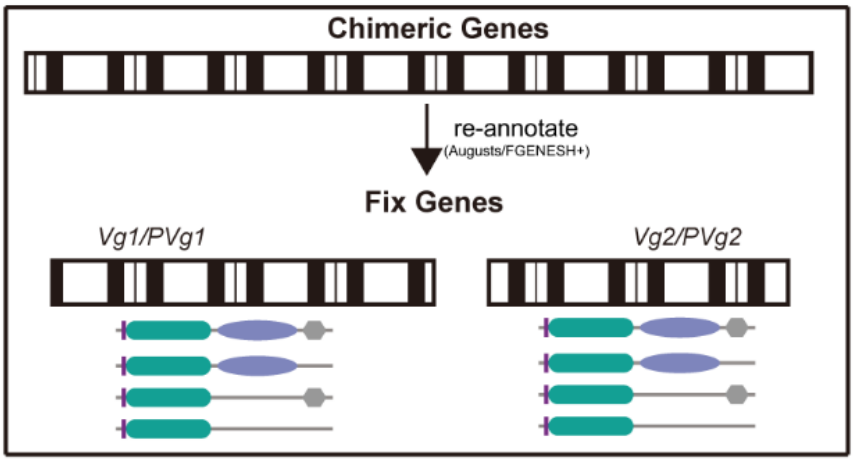


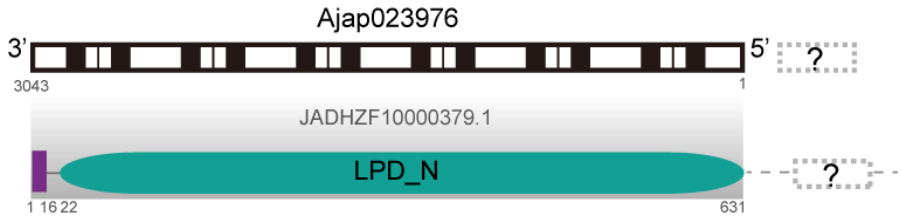
Fig. 7 | Schematic of the evolution of *Vg* and *VgR* in parasitoid wasps. The ancestral parasitoid wasp is thought to have been an ectoparasitoid that retained both the *Vg* and *VgR* genes and laid yolk-rich eggs. Hymenopteran evolution involved several independent shifts from ectoparasitism to endoparasitism. Some endoparasitoid wasps (e.g., *Pteromalus puparum*) retain *Vg* and *VgR* genes and lay yolk-rich eggs. Others have lost *Vg*, likely through genome rearrangement, leading to loss or relaxed selection of *VgR* genes due to disruption of vitellogenesis. These wasps often produce yolkless eggs (**Supplementary Table 1**). There were several shared genomic changes related to transport, metabolism, development, sensory systems, and stress resistance among the five independent *Vg*-loss lineages; these may have been related to endoparasitoid adaptation and/or contributed to loss of the yolk as an available nutrient source to the early embryo.

Supplementary Figures

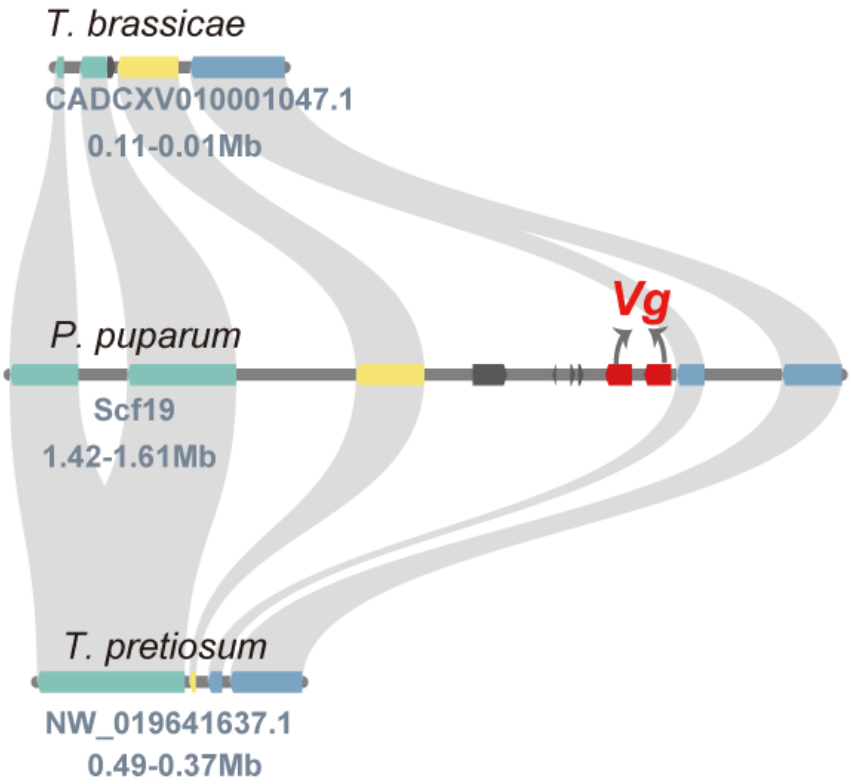


Species	Chimeric genes ID	Fixed genes	
Amblyteles armatorius	ENSAAYP00000011660	*AarmVg1*	*AarmVg2*
Anisopteromalus calandrae	Acal20629-TA	*AcalVg1*	*AcalVg2*
Ooceraea biroi	XP_019887836.2	*ObirVg1*	*ObirVg2*
Pachycrepoideus vindemmiae	Pvin01909-TA	*PvinVg1*	*PvinVg2*
Solenopsis invicta	XP_039305184.1	*SinvVg1*	*SinvVg2*
Trichomalopsis sarcophagae	OXU21803.1	*TsarVg1*	*TsarVg2*
Diadromus collaris	DCO15731-TA	*DcolPVg1*	*DcolPVg2*

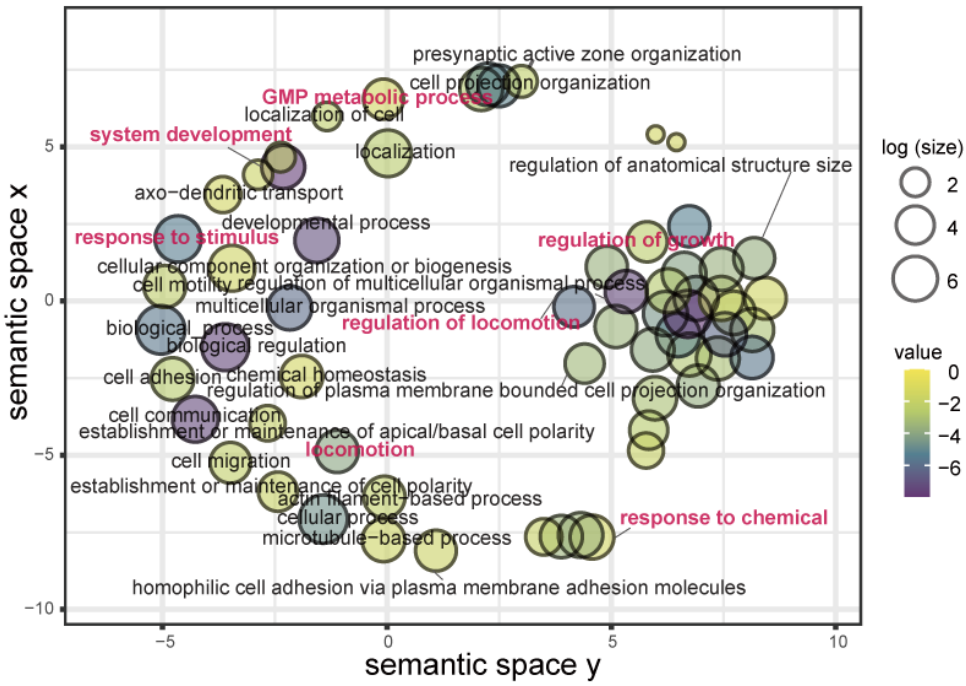
Supplementary Fig. 1 Schematic of fixation of chimeric genes. For genes joined together but owned repeat domain structures of *Vg* were split into two genes which composed domains of *Vg* or *PVg*. These genes were re-annotate using Augusts or FGENESH+ webserver, and resulted genes were further validated from using BLAST and HMMER.



Supplementary Fig. 2 Domain structure and location of *PVg* in *Asobara japonica*. *PVg* (*Ajap023976*) of *A. japonica* was located at the start terminal on the scaffold JADHZF10000379.1 from scanning 3,043bp while the LPD_N domain predicted using SMART webserver was covered whole protein-coding sequences from 22 to 631 site with 1 to 16 was signal peptide. However, whether this *PVg* was a result of genome misassemble or bad annotation still unknown.



Supplementary Fig. 3. Gene synteny of *Vg*s and flanking genes among representative *Vg*-loss species and *Vg*-retaining wasps in Chalcidoidea. *Pteromalus puparum* as the *Vg*-retaining species was showed in the middle and two *Vg*s were highlighted in red, while *Trichogramma brassicae* and *T. pretiosum* were showed above and below. Results showed the two *Vg*s in *Pteromalus puparum* have no homologous genes in this syntenic block region.



Supplementary Fig. 4. Enrichment visualization of 199 gene under relaxed selection using REVIGO. The enrichment results generated from GOATOOLS were used as input, and value C was pre-set as 0.5 which corresponding to the "small" list size, then REVIGO will clustered listed GO terms and reduce redundancy of them. Based on the information in Supplementary Table 10, dispensability of GO terms less than 0.2 were labelled in the figure and some interested GO terms were marked in pink.