Lineage-specific targets of positive selection in three leaf beetles with different defence capacity against a parasitic wasp

Xuyue Yang[1], Christopher Wheat[1], Tanja Slotte[1], and Peter Hamback[1]

[1]Stockholm University

November 4, 2022

Abstract

Parasitoid wasps are major causes of mortality of many species, and therefore traits related to host immune defence are usually favoured by natural selection. One powerful approach to detect functionally important genes under natural selection is through the analysis of directional selection acting upon protein-coding gene sequences across different species. Here, we investigated patterns of positive selection across three closely related leaf beetle species with different immune defence capacity against a shared parasitoid wasp using a Bayesian approach for the McDonald–Kreitman test. Focusing on single-copy orthologs for Coleoptera, as well as on candidate immune related genes, we detected species-specific positive selection on coding regions in each of the closely related *Galerucella* beetle species. Results indicated that more immune genes had experienced positive selection in the species with the greatest immunocompetence (*G. pusilla*) against parasitoid wasps, while almost no immune genes were under positive selection in the species with the least immunocompetence (*G. calmariensis*).

Lineage-specific targets of positive selection in three leaf beetles with different defence capacity against a parasitic wasp

Xuyue Yang[1], Christopher W. Wheat[2], Tanja Slotte[1], Peter A. Hambäck[1*]

[1]Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm 10691, Sweden,[2]Department of Zoology, Stockholm University, Stockholm 10691, Sweden

*Corresponding author

Email : peter.hamback@su.se

Abstract

Parasitoid wasps are major causes of mortality of many species, and therefore traits related to host immune defence are usually favoured by natural selection. One powerful approach to detect functionally important genes under natural selection is through the analysis of directional selection acting upon protein-coding gene sequences across different species. Here, we investigated patterns of positive selection across three closely related leaf beetle species with different immune defence capacity against a shared parasitoid wasp using a Bayesian approach for the McDonald–Kreitman test. Focusing on single-copy orthologs for Coleoptera, as well as on candidate immune related genes, we detected species-specific positive selection on coding regions in each of the closely related *Galerucella* beetle species. Results indicated that more immune genes had experienced positive selection in the species with the greatest immunocompetence (*G. pusilla*) against parasitoid wasps, while almost no immune genes were under positive selection in the species with the least immunocompetence (*G. calmariensis*).

Key words: host-parasitoid systems, positive selection, immune genes, beetles, *Galerucella*

Introduction

Parasitoids are major natural enemies of many insects, leading to high mortality rates of host species (Asgari & Rivers, 2011; Hawkins, Cornell, & Hochberg, 1997). One particularly important group are koinobiont endoparasitoid wasps, which lay eggs and complete their entire larval development inside their hosts, leading to host death. To defend themselves, insect hosts respond immunologically within hours by encapsulating and killing the wasp eggs (Carton, Poirié, & Nappi, 2008). Due to the ever-changing virulence of wasps, host species have been shown to rapidly evolve novel defence mechanisms (Kraaijeveld, Van Alphen, & Godfray, 1998), and this host-parasitoid coevolutionary dynamic likely explains the large amount of variation in defence mechanisms between host populations and species (Carton et al., 2008; Wertheim, 2022). Therefore, traits associated with host defence against parasitoids are widely regarded to be under strong selection pressures with a potentially crucial role in species divergence (Carton et al., 2008). However, while many studies have investigated the physiological and behavioural responses upon parasitoid attack in single host species (De Roode & Lefèvre, 2012; Fors, Markus, Theopold, Ericson, & Hambäck, 2016; Godfray, 1994), and evolutionary responses of hosts to various pathogens (Kim-Jo, Gatti, & Poirié, 2019), few studies have compared divergence in immune genes among closely related host species that are attacked by a common parasitoid species except for wasps attacking*Drosophila* (Wertheim, 2022).

We have previously investigated host-parasitoid interactions in a system with three closely-related beetle species (*Galerucella calmariensis* , *G. pusilla* and *G. tenella* , Coleoptera: Chrysomelidae) that differ in their hemocyte composition, in their capacity to induce potent hemocytes upon attack and in their encapsulation success against the parasitic eulophid wasp *Asecodes parviclava* (Fors et al., 2016; Fors, Markus, Theopold, & Hambäck, 2014). At the cellular level, *G. pusilla* exerts a much more potent immune response, with strong encapsulation and melanisation responses against wasp eggs, compared to *G. calmariensis* where encapsulation events of wasp eggs in host larvae are rarely observed. The immune response of *G. tenella* , in turn, is intermediate between *G. pusilla* and *G. calmariensis* . Recently, these differences in immune responses among species were also confirmed by gene expression analysis, especially in showing differential expression of genes involved in signalling, haematopoiesis, and melanisation in*G. pusilla* (Yang et al., 2020).

Insects rely on their innate immune responses to defend themselves against parasitoid and pathogen attack, as they lack an adaptive immune response. This innate immune system consists of two parts, humoral and cellular immunity. The humoral defence system is mainly active against pathogens and is mediated by genes regulating the expression of antimicrobial peptides (AMPs), wound healing, coagulation, and melanisation pathways (Bulet, Hetru, Dimarcq, & Hoffmann, 1999; Gillespie, Kanost, & Trenczek, 1997). Immune responses against parasitoid wasps, on the other hand, are usually regulated by cellular immunity, which can be classified into seven broad functional categories: recognition, signalling, effector, proteases, haematopoiesis, melanisation, and wound healing (see also Yang et al., 2020). Immune genes, particularly those involved in the recognition phase, have been found to evolve faster than other genes (Nielsen et al., 2005; Sackton et al., 2007; Schlenke & Begun, 2003; Waterhouse et al., 2007), but selection dynamics are complex and depend upon specific functions and taxa studied (Keehnen, Hill, Nylin, & Wheat, 2018; Keehnen, Rolff, Theopold, & Wheat, 2017). Recent studies based on amino-acid substitutions in *Drosophila* species suggest that some classes of immune-related genes show more evidence of positive selection than other genes (Heger & Ponting, 2007; Waterhouse et al., 2007) while other functional classes such as AMPs appear to experience more balancing or purifying selection (Unckless, Howick, & Lazzaro, 2016; Unckless & Lazzaro, 2016). These findings are however restricted to the scope of anti-microbial immune genes, whereas the selection dynamics of genes involved in the host-parasitoid specific immune response have received less attention. The observed phenotypic differences between species in our study system suggest differences in historic and current rates of evolution in key immune-related genes involved in the defence against parasitoids, and makes the system highly suitable for studying host selection dynamics in response to parasitic wasps. In particular, we expect stronger positive selection on parasitoid related immune genes in the species with the strongest immunocompetence against parasitoid wasps (*G. pusilla*) compared with the species with the least immunocompetence (*G. calmariensis*).

Apart from immunity related genes, other gene groups are also expected to differ between the three species, though the link between such genes and parasitoid attack are less clear or absent. For instance, *G. calmari-*

ensis are slightly larger in size and develop faster than the other two species, indicating potential differences in pathways concerning development and metabolic processes (Carreira, Mensch, & Fanara, 2009). Moreover, *Galerucella* larvae differ in colour; bright yellow in *G. calmariensis* , pale white-yellow in *G. pusilla* and yellow with many black/brown spots in *G. tenella*(Hambäck, 2004). Such colour differences and spotting patterns may involve gene pathways related to pigmentation and melanisation (Ito et al., 2010; Linnen, O'Quin, Shackleford, Sears, & Lindstedt, 2018), or master regulatory genes (Deshmukh, Baral, Gandhimathi, Kuwalekar, & Kunte, 2018). Finally, the species may also differ in mate finding traits, such as the occurrence and detection of pheromones and cuticular hydrocarbons, and host-plant finding traits, as *G. tenella* uses a different host plant than the two other species.

A widely-used approach for investigating positive selection based on DNA sequencing data is the McDonald–Kreitman test (MK test), which infers the direction of natural selection by comparing the ratio of non-synonymous and synonymous polymorphism (P_n/P_s) within species to the ratio of synonymous and non-synonymous divergence (D_n/D_s) between species (McDonald & Kreitman, 1991). When positive selection favours the phenotypic impact of novel amino acid changes, and the corresponding advantageous alleles go to fixation, positive selection is expected to yield$D_n/D_s > P_n/P_s$, under the assumption that synonymous mutations are evolving neutrally and there is no change in constraint over time. In contrast, if weak purifying selection is prevalent, deleterious alleles can segregate in the population for extended periods, yet rarely fix and therefore contribute little to divergence. Under this scenario$D_n/D_s < P_n/P_s$. The contribution of positive selection to amino acid divergence can be estimated usingα (= 1-$D_s P_n/D_n P_s$) (Smith & Eyre-Walker, 2002), which represents the proportion of substitutions driven by positive selection.

The classic MK test was designed in the pre-genomics era, analyzing each gene locus separately and usually limited to comparisons of pairs of species (McDonald & Kreitman, 1991). It was not designed to infer the directionality or relative strength of selection across 1000s of genes across several study species. To overcome the limitations, here we used the high-dimension McDonald-Kreitman Poisson random field method (hereafter HDMKPRF, Zhao et al., 2019). This method is an extension of the MKPRF method developed by Sawyer and Hartl (1992), applying a Bayesian model across multiple gene loci to simultaneously estimate population genetic parameters of multiple target species, including lineage specific mutation rates and relative effective population size (N_e), which improves inference of directionality and relative strength of selection along the lineages unique to each species.

To examine selection dynamics in our study system, we generated whole genome re-sequencing data from 15 individuals from each of the three beetle species (*G. calmariensis* , *G. pusilla* and *G. tenella*). We then investigated the selection dynamics acting on coding regions from single-copy orthologs using HDMKPRF as well as a set of candidate immune genes. Based on previous studies, our primary hypothesis was that immune genes involved in wasp attack would more frequently have experienced positive selection in the species with the strongest immune response (i.e. *G. pusilla* >*G. tenella* > *G. calmariensis*).

2. Methods

2.1 Study species

The three *Galerucella* species (Coleoptera: Chrysomelidae) are closely related, with recent divergence times: *G. pusilla* and*G. calmariensis* diverged around 77,000 years ago while *G. tenella* diverged around 400,000 years ago (Hambäck et al., 2013).*G. pusilla* and *G. calmariensis* are monophagous on*Lythrum salicaria* , whereas *G. tenella* is oligophagous with the primary host *Filipendula ulmaria* . The three beetle species have similar life cycles. Adults in the area appear in May and start laying eggs on leaves or stems of their host plants. It takes a few weeks for the eggs to hatch, 2-3 weeks for the larvae to pupate, and another 2-3 weeks for the adults to emerge from the pupae. Adults then overwinter until next May. The geographic distribution in Sweden differs between species: *G. pusilla* occurs in the south up to central Sweden (62°N, 17°E) whereas *G. calmariensis* and *G. tenella*occur both in the south and north along the entire Baltic seashore (Fors et al., 2014).

The three species share an endoparasitoid wasp enemy *Asecodes parviclava* (Hymenoptera: Eulophidae),

which lays one or more eggs in the beetle larvae (Stenberg & Hambäck, 2010). When successfully parasitized, wasp eggs hatch and wasp larvae turn the beetle larvae to black mummies containing the wasp pupae. However, if beetles manage to defend themselves, their immune system encapsulates and kills the wasp eggs, enabling the host larvae to continue growing and developing (Fors et al., 2014). Previous work show that the beetle species differ in their capacity to mount an efficient defence against parasitoid attack. Whereas *G. pusilla* has a strong capacity to encapsulate wasp eggs, encapsulation is rarely observed in *G. calmariensis* , and at an intermediate frequency in *G. tenella* (Fors et al., 2016; Fors et al., 2014).

We collected 45 adult male individuals, 15 samples from each *Galerucella* species, during mid-May 2019 from the following sites: three *G. calmariensis* populations: Iggön (60°52'18"N, 17°19'29"E), Våtnäs (61°32'93"N, 17°12'77"E) and Hölick (61°37'22"N, 17°27'18"E); three *G. pusilla* populations: Rastsjön (60°6'36"N, 17°53'97"E), Lörudden (62°14'14"N, 17°39'12"E) and Haversjön (59°2'31"N, 17°9'49"E); three *G. tenella* populations: Umeå-1 (63°46'72"N, 20°36'00"E), Umeå-2 (63°46'36"N, 20°37'48"E) and Umeå-3 (63°47'18"N, 20°35'89"E). For each population, five individuals were sampled.

2.2 DNA extraction and sequencing

All individual samples were snap frozen in liquid nitrogen and stored at -80° before DNA extraction. Genomic DNA were extracted from the whole adult body using KingFisher Cell and Tissue DNA Kit using the sample preparation protocol "DNA Extraction from Single Insects". After extraction, DNA concentrations were measured with a Qubit 3.0 Fluorometer using the dsDNA HS Assay Kit (Thermo Fisher Scientific) and Nanodrop 8000 to ensure an absorbance ratio at 260/280 between 1.7 and 2. We estimated DNA fragmentation using agarose gel electrophoresis stained with 2% GelRed and only retained samples with minimal degradation. Library preparation was performed with the Illumina TruSeq DNA PCR-free library preparation kit and then paired-end 2x150-bp sequenced on a NovaSeq6000 platform at SciLifeLab, Sweden. Library preparation failed for one *G. calmariensis* sample from Vatnas and this sample was excluded from downstream analysis. The total number of samples with whole-genome resequencing data was thus 44 and we generated 1.6 Gbp of sequence data (>Q30) in total (out of 1.8 Gbp), corresponding to an average of 34.8 Mbp per sample.

2.3 Population mapping and statistics

We assessed quality on the resequencing data using FastQC v0.11.5 (Andrews, 2017) before and after filtering, and only retained reads [?]50 bp with a quality score >30 in both read start and end. All sequence reads were mapped against the *Galerucella calmariensis* reference genome (Yang, Slotte, Dainat, & Hamback, 2021) using NextGenMap version 0.4.12 (Sedlazeck, Rescheneder, & von Haeseler, 2013). The reference genome had an assembly size of 588 Mbp, containing 39,255 scaffolds and 40,031 predicted proteins with 91.3% and 85.1% complete orthologs in the genome and proteome, respectively, compared with the endopterygota_odb10 database (Simao, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) (For further info on the reference genome assembly see Yang et al., 2021). Mapping rates were similar between samples (85% to 95%). We filtered the resulting bam files with Samtools v1.3.1 (Li et al., 2009) to retain alignments with mapping quality>20 (-q 20).

We next called SNPs across all samples using FREEBAYES v0.9.21 (Garrison & Marth, 2012). For SNP filtering, we only kept bi-allelic sites with a minimum read depth of 5X, a quality score >30 and a maximum proportion of missing data of 20%. To ensure there is not population genetic structure across populations within each species, we conducted a PCA analysis. For this purpose, we first conducted LD-based pruning (–indep-pairwise 50 10 0.2), followed by a principal component analysis (PCA) across all the samples using Plink v1.9 (Purcell et al., 2007) (Supporting information Figure S1). Genetic diversity (nucleotide polymorphism, π) was estimated for each species using pixy (Korunes & Samuk, 2021).

2.3 Test of selection

Multiple consensus sequences of coding sequences for all samples were extracted using bam2consensus function from BamBam v1.4 (Borowiec, 2016), allowing a minimum read coverage per site of 4X. BamBam

uses the individual bam files that were mapped to the *G. calmariensis*draft *de-novo* genome, and extracts consensus for each CDS region based upon the genome annotation. We then assessed summary statistics of the consensus sequences using AMAS v1.0 (Borowiec, 2016). Only CDS regions with the length >300 and low proportions of missing values (<10%) were kept for downstream analysis (N =11,368).

To limit out analysis to orthologous loci among our three species, we assessed orthology among *G. calmariensis* protein sets and three other Coleoptera species (Asian long-horned beetle [*Anoplophora glabripennis*], red flour beetle [*Tribolium castaneum*] and mountain pine beetle [*Dendroctonus ponderosae*]) using OrthoVenn2 (Xu et al., 2019) with default settings. A total of 4,591 single-copy-orthologs (SCO) were identified in *G. calmariensis* . Then, we identified these SCOs from the previously identified high quality consensus sequences, which resulted in 4,154 SCOs for downstream analysis.

To detect adaptive evolution, we used HDMKPRF (Zhao et al., 2019), which is an extension of MKPRF to analyse selection across multiple species (Bustamante, Wakeley, Sawyer, & Hartl, 2001; Sawyer & Hartl, 1992). To improve detection of selection, HDMKPRF pools information over many loci and thereby gains power compared with single-gene based methods. The method then uses a Poisson random field on synonymous and non-synonymous polymorphisms to estimate posterior distributions for included genes and simultaneously corrects for differences in population demography, while accounting for the phylogenetic structure (for details see table 1 in Zhao et al., 2019). The script for calculating population parameters and for performing the HDMKPRF to derive selection intensities was kindly provided by Zhao et al. Here we follow terminology from the MKPRF test (Bustamante et al., 2001) in referring to the estimated relative strength of selection per gene as a 'selection intensity', akin to a neutrality index measure (Hahn, 2019).

Estimates of positive selection using population resequencing data are usually biased downward by the segregation of slightly deleterious mutations (Bierne & Eyre-Walker, 2004; Fay, Wyckoff, & Wu, 2002). To minimize the impact of such bias, we removed singleton polymorphisms from all gene sets using a custom script (Sattath, Elyashiv, Kolodny, Rinott, & Sella, 2011). In order to test the effect of singleton removal on the power of detecting adaptive evolution, we applied HDMKPRF to gene sets before and after removing singletons with 200,000 burn-in steps, 400,000 total MCMC steps and set a thinning interval of 5. We performed the same analysis three times and retained the average value from triplicates, with these settings demonstrating high levels of concordance. A gene was considered to be under positive selection only when the 95% posterior confidence interval of selection intensity was > 0, and similarly under negative selection when the interval was < 0.

We performed a gene ontology (GO) enrichment test of biological process terms of genes under positive selection using the BIOCONDUCTOR package topGO (Alexa & Rahnefuhrer, 2010). The background genes included the 4,154 genes used in HDMKPRF and those genes that were significantly (positively or negatively) selected were tested against this background for enrichment of biological process terms (FDR<10%) using the parent–child algorithm (Grossmann, Bauer, Robinson, & Vingron, 2007). To ensure a robust result, we only analysed GO terms with at least 5 members (node size=5), and we used EggNOG v5.0 (Huerta-Cepas et al., 2019) before the analysis to assign Gene Ontology (GO) terms to the predicted protein sets using Insecta as taxonomic scope to restrict the functional inferences to an insect-related scale. Afterwards, enriched GO terms were clustered to representative functional subsets using the REVIGO *Drosophila* database (Supek, Bošnjak, Škunca, & Smuc, 2011).

2.4 Candidate gene analysis

For the immune gene analysis, we used the candidate protein set from a previous RNA-seq study (Yang et al., 2020), which contains 166 genes suggested to be important in immune response against parasitoid wasp attack in *Drosophila* (Supporting information Tables S1), subdivided into seven functional immune gene categories: recognition (N =17), signalling (N =35), effector (N =21), proteases (N =35), haematopoiesis (N =31), melanisation (N =18), and wound healing (N =9). We then used BLASTP with an E-value[?]1×10^{-6} and a bitscore>60 to identify these 166 immune proteins in the *G. calmariensis*proteome. We further assessed the accuracy of BLASTP results by comparing the BLAST hits with our functional annotation. Finally,

to achieve a more conservative assessment of selection dynamics on immune genes, we only investigated candidate immune genes that were among the previously identified 4,154 SCOs, as including other loci in the genome may be confounded with recent gene birth death dynamics and generate spurious results in molecular analyses.

3. Results

3.1. Population-level patterns

Before quality filtering, a total of 266, 262 and 256 million sites were retained for analyses of polymorphism levels, among which 8.7, 9.2 and 8.4 million were variable sites in *G. tenella* , *G. pusilla*and *G. calmariensis* , respectively (Raw data: Yang, Wheat, Slotte, & Hamback, 2022). After quality filtering, we retained 7.6 million variable sites out of 126.5 million sites in *G. tenella* , 8.6 million variable sites out of 130.6 million sites in *G. pusilla*and 7.2 million variable sites out of 122.6 million total sites in*G. calmariensis* . At the whole genome level, *G. pusilla*populations harboured the highest nucleotide diversity (0.0058),*G. calmariensis* (0.0051) had the lowest, with *G. tenella*(0.0056) having an intermediate nucleotide diversity. The posterior mean estimates of the relative effective population sizes (N_e) from HDMKPRF on the 4154 genes were N_2 =1.045 for *G. pusilla* and N_3 = 0.934 for *G. calmariensis* in units of *G. tenella*(N_1). Thus, the relative ranking of population sizes among species was *G. pusilla* > *G. tenella* > *G. calmariensis* .

3.2 Genes under Selection

The summary of synonymous and nonsynonymous polymorphisms and divergence across species indicates that singleton removal increased the estimated proportion of positively selected genes in all three species and reduced indices of negative selection (Table 1), indicating the presence of segregating weakly deleterious mutations. In *G. calmariensis* , we detected mainly positive selection ($\alpha > 0$) and a higher proportion of positively selected genes. On the other hand, in*G. tenella* , negative selection was more common ($\alpha < 0$) even after singleton removal and with a higher proportion of negatively selected genes. In the following sections, we only discuss inferences after removing singletons.

The HDMKPRF identified a similar number of genes under selection in the three *Galerucella* species (Fig. 1). In *G. pusilla* , 451 and 562 genes were identified as being under positive and negative selection, respectively. In *G. calmariensis* , 610 and 553 genes were identified as being under positive and negative selection, respectively. Finally, in *G. tenella* , 425 and 466 genes were identified as being under positive and negative selection, respectively. Because genes under positive selection are more commonly associated with lineage-specific adaptive traits (e.g., immunity genes), we focus our analyses on gene function connected to genes under positive selection.

The gene set enrichment analysis found several functions that were enriched in genes under positive selection, some of which were common among the three beetle species and some that differed (Table S2-S4). First, enriched functions common among the beetle species included functions involved in the formation of adult morphology, such as the imaginal disc pattern formation (forming the adult cuticle and appendage structures), the wing disc pattern formation (forming wing structures) and the dorsal/ventral pattern formation. Second, unique gene categories under positive selection in *G. calmariensis* included those coding for metabolic processes (e.g., processes related to carbohydrate derivatives, oligosaccharides, amino sugars, sulphur compounds and catechol-containing compounds) and those coding for processes in the nervous system (e.g., neuroblast proliferation, neuroblast differentiation, nervous system process). Third, unique gene categories under positive selection in *G. pusilla* included those coding for positive regulation of the innate immune response (e.g., positive regulation of small GTPase mediated signal transduction) and those coding for axoneme assembly (e.g., cilium movement, cilium organization). Finally, unique gene categories under positive selection in *G. tenella* included genes coding for a range of biosynthetic processes (e.g., nucleobase-containing compound biosynthetic processes, heterocycle biosynthetic processes and aromatic compound biosynthetic processes), genes coding for lipid metabolic processes (e.g., sterol metabolism, membrane lipid biosynthesis, cellular lipid metabolism, sphingolipid metabolism and lipoprotein metabolism), but also GO pathways involving the activation of immune response, pigment metabolic process involved in pigmentation,

peripheral nervous system development and response to oxidative stress.

3.3 Candidate gene analysis

Due to the conservative filtering strategy (SCO set, length >300 and missing values <10%), only 72 out of 166 candidate immune genes present in the *G. calmariensis* genome remained in our MK analysis, corresponding to 14 recognition genes, 22 signalling genes, 5 effectors, 18 protease coding genes, 26 haematopoiesis genes, 15 melanisation genes and 5 wound healing genes (Table S1). Eight of these immune genes were identified to be under positive selection in *G. pusilla* , including important genes involved in parasitoid recognition (*santa-maria* and*Corin*), Toll and JNK pathways (*grass* and *Tak1*), a protease with serine-type carboxypeptidase activity and genes involved in lamellocyte differentiation (*Raf* , *cher* and *zfh1*) (Table 2). In contrast, only one candidate immune gene (*Cyp9f2*) was positively selected in *G. calmariensis* . In *G. tenella* , four immune genes were found to be under positive selection; two recognition genes (*Corin* and *PGRP-LE*) and two genes with important roles in haematopoiesis through regulation of lamellocyte differentiation (*cher* and *Cyt-b5*) (Table 2).

4. Discussion

This study reports one of the first analyses comparing genome-wide natural selection across multiple lineages of beetle species, and the first analysis identifying host species differences in the selection dynamics of putative immune genes against parasitoid wasps. Our study system included three closely-related leaf beetle species (*Galerucella* spp.) with previously detected phenotypic differences in their capacity to encapsulate eggs from the shared parasitoid wasp (*Asecodes parviclava*), as well as differences in gene expression following wasp attack (Fors et al., 2016; Fors et al., 2014; Yang et al., 2020). Using a modified MK-test based approach and whole genome re-sequencing data from 44 individuals of the three species, our analysis identified, as hypothesized, a higher number of immune genes under positive selection in the species with the strongest capacity to encapsulate wasp eggs (*G. pusilla* , $N = 8$). Species lacking the capacity to encapsulate wasp eggs (), or having an intermediate encapsulation capacity (*G. tenella,* $N = 4$), had a lower number of immune genes under positive selection (*G. calmariensis,* $N = 1$). Even though these genes have yet to be functionally characterized in the three beetle species, previous comparative gene expression analyses (Yang et al., 2020) as well as the differential selection between beetles with high and low immune phenotype in this paper suggest they are important in mediating defence against parasitoid wasps. Additionally, a genome-wide scan based on orthologous loci among the three species characterized different selection patterns for a range of other genes, suggesting divergent directions of selective pressure between species.

Even though we could analyze only a subset of immune genes in the*Galerucella* genome (72 out of 166), the detected genes indicate the type of processes most likely under selection. Most importantly, the analysis suggests that the observed phenotypic differences in the capacity to encapsulate wasp eggs (Fors et al., 2014) are likely explained by genes involved in wasp egg detection and in the recruitment of hemocytes building the capsule (Yang et al., 2020), whereas genes involved in melanisation or wound healing processes seem to be more conserved. In the species with the strongest encapsulation capacity (*G. pusilla*), we identified two recognition genes (*santa-maria* and *Corin*), two signalling genes (*Tak1*and *grass*), one protease coding gene (*CG32483*) and three genes involved in haematopoiesis (*Raf* , *cher* and*zfh1*) that were all under positive selection. This set of eight genes can be contrasted against the set of four immune genes that were identified as under positive selection in *G. tenella* , the species with intermediate encapsulation capacity, or the single positively selected immune gene in *G. calmariensis* , the species with the weakest encapsulation capacity. It is however notable that two genes (*Corin* and *cher*) were under positive selection in both*G. tenella* and *G. pusilla* , whereas two other genes (recognition gene *PGRP-LE* and haematopoiesis gene *Cyt-b5*) were only found as being under positive selection in *G. tenella* , indicating both parallel and diverging selection patterns. The importance of these genes for the immune system in this and other species is also evident as the two genes *Corin* and *cher*were similarly differentiated among *Drosophila* species with different encapsulation ability (Salazar-Jaramillo et al., 2014), whereas three other genes (*Grass* , *zfh1* and *Cyt-b5*) were differentially expressed following wasp infection involving*A. parviclava* and *G. pusilla* (Yang et al., 2020).

The specific functions for these genes in *Galerucella* have not been identified, but some studies exist from *Drosophila* . First, both *santa-maria* and *Corin* have been suggested to encode proteins with scavenger receptor activities, whereas *PGRP-LE*encodes an intracellular protein that binds to diaminopimelic acid-type peptidoglycans to activate the IMD/Relish pathway (Bosco-Drayon et al., 2012). Second, both *Tak1* and *Grass* have been predicted to be involved in the Toll signalling pathway, which may be the most important signalling pathway in the immune response against wasp attack in *Drosophila* (Carton et al., 2008; Lemaitre & Hoffmann, 2007; Wertheim et al., 2005). Third, the protease coding gene *CG32483*is predicted to enable serine-type carboxypeptidase activity but there is no known defense function against wasp attack. Finally, the four haematopoiesis genes are widely known to be involved in the regulation of lamellocyte differentiation and proliferation, which is the hemocyte type that makes up most of the capsule around the wasp egg in both*Galerucella* (Fors et al., 2014) and *Drosophila* (Kim-Jo et al., 2019). Among these genes, *Raf* is suggested to play a crucial role in the transition from pro-hemocytes to lamellocytes (Luo, Rose, Roberts, & Dearolf, 2002), *cher* is predicted to be involved in the negative regulation of lamellocyte differentiation (Rus et al., 2006), *zfh1* is one element in a transcription factor cascade that plays a role as the switch between plasmatocyte and lamellocyte fate, and *Cyt-b5* encodes a conserved hemoprotein that is required for hemocyte regulation (Kleinhesselink, Conway, Sholer, Huang, & Kimbrell, 2011). The specific immune genes selected for likely vary between species (Wertheim, 2022), but our results indicates that phenotypic differences in the immune response in these species is related to multiple functional genes.

When examining other gene functions identified as experiencing positive selection, we found several interesting candidates, where genes coding for imaginal disc pattern formation and wing disc pattern formation were under positive selection in all species. Imaginal discs are epithelial sacs found in insect larva that later develop into cuticular structures (e.g., head, wing, limbs, thorax) of adult insects, and the wing disc is among the largest imaginal discs in insects (Blair, 2009). These findings are not surprising as the beetle species are morphologically differentiated, both in size and colour. Other genes varied between species, but those under positive selection in *G. calmariensis*specifically involved genes coding for metabolic functions such as the metabolism of carbohydrate derivatives, oligosaccharides, amino sugars, sulphur compounds and catechol-containing compounds. These sets of positively selected gene functions indicate the importance of energy allocation and dietary transitions during evolution in this species. Moreover, pathways related to the nervous systems were also found to be positively selected in *G. calmariensis* . Positive selection on nervous system-related genes were previously documented in social insects such as bees and ants (Roux et al., 2014; Woodard et al., 2011) but have rarely been reported in beetles. These patterns may indicate changes either in the capacity to detect host plants or mates and may thus be involved in the species differentiation. In *G. tenella* , several pathways related to lipid metabolic processes were enriched in positively selected genes. Some of these processes, such as sterol metabolism, may be linked to needs to handle differences in plant chemistry from the different host plants (Rosaceae vs. *L. salicaria*). Other lipids, such as sphingolipids, have been suggested to be involved in cell defences and could be interesting to study in relation to wasp attack. Finally, a unique pathway under positive selection in *G. tenella* was involved in the pigmentation metabolic process, which may potentially explain the different spotting and pigmentation patterns in the species.

Some methodological concerns are relevant in relation to our study. First, we were only able to include 4,154 orthologous genes in the analysis, which may have reduced the power to detect important GO terms. However, the removals should not have biased our conclusions as genes were likely removed randomly across the genome. Moreover, by discarding genes of low coverage and removing bias due to recent gene birth death dynamics, this smaller set of genes provides a robust gene set for inferring selection dynamics and is sufficient for estimating genome wide selection dynamics. Second, a general problem in MK tests is their sensitivity to the demographic history (Eyre-Walker, 2002; McDonald & Kreitman, 1991). In our analysis, we approached this problem by using HDMKPRF, which attempts to correct for demography effects during the detection of selection patterns by accounting for the changes in effective population size and other population genetic parameters. One potential concern, however, is that the proportion of substitutions driven by positive selection were highest in *G. calmariensis* even though this species had the smallest N_e . The correction for

demography in this species may have been insufficient as nearly neutral nonsynonymous substitutions have fixed during the recent period of low N_e and thereby incorrectly indicate evidence for positive selection. Third, an additional issue when examining selection patterns is slightly deleterious mutations. If such mutations are segregating in the population, it becomes difficult to detect positive selection and the degree of positive selection will be underestimated. Removing deleterious mutations during selection analysis is therefore widely accepted, by either setter a threshold for removing minor alleles (Fay, Wyckoff, & Wu, 2001; Fay et al., 2002) or by removing singleton polymorphism (Bierne & Eyre-Walker, 2004; Proschel, Zhang, & Parsch, 2006). We used the latter strategy, which resulted in a skew towards positively selected sites while keeping the relative ranking of α among species. Thus, while absolute α-estimates in our analysis are certainly underestimated, the relative ranking among species is likely correct.

To summarize, our study species diverged recently (Hambäck et al., 2013) but have during a limited time evolved key phenotypic differences in the defence against parasitoid wasps. These phenotypic differences allowed us to pinpoint important links between natural selection on immune genes and host-parasitoid interactions where the species with the highest immunocompetence had the highest number of positively selected immune genes. When examining candidate immune-related genes, our results are consistent with the variable selection pattern of immune genes in other studies (Heger & Ponting, 2007; Waterhouse et al., 2007), but also show that selection acts on multiple genes in the immune pathways and not only on a few major genes (Kraaijeveld et al., 1998). While the evolutionary process underlying the speciation and immune system differences is unknown, comparisons with studies involving experimental evolution in microbial systems provide potential hypotheses. These studies show that natural enemies generally are not expected to cause host populations to differentiate sympatrically, and that the role of natural enemies instead seems to be that the arms-race between host and their natural enemies increases the rate of divergence between allopatric populations (Buckling & Hodgson, 2007; Buckling & Rainey, 2002). If these processes are similar for the evolution of host-parasitoid interactions in this beetle system, it is likely that the phenotypic differences in encapsulation capacity initially evolved during periods when the beetles where geographically isolated and thereby coincided with an allopatric speciation process. In either case, it seems likely that the observed phenotypic and genotypic differences and the different relationships among species with the parasitoid wasps have been pivotal during the speciation process.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

The authors thank L. Ericson for collecting beetle samples from the field and M. de la Paz Celorio Mancera for helpful guide in DNA extraction. This work was supported by grant VR-2015-4232 (to PAH) from the Swedish Research Council Vetenskapsrådet, and by the National Genomics Infrastructure (NGI) and Uppmax through the Science for Life Laboratory, which is funded by the Knut and Alice Wallenberg Foundation, and the Swedish Research Council Vetenskapsrådet. Science for Life Laboratory are also acknowledged for bioinformatics support through the PhD advisory program. Computational analyses were performed on resources provided by SNIC through Uppsala Multidisciplinary Centre for Advanced Computational Science (UPPMAX) under project b2013018. SNIC is partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

References

Alexa, A., & Rahnefuhrer, J. (2010). *topGO: enrichment analysis for gene ontology. R Package Version 2(0)* .

Asgari, S., & Rivers, D. B. (2011). Venom proteins from endoparasitoid wasps and their role in host-parasite interactions. *Annual Review of Entomology, 56* , 313-335. doi:10.1146/annurev-ento-120709-144849

Bierne, N., & Eyre-Walker, A. (2004). The genomic rate of adaptive amino acid substitution in *Drosophila* . *Molecular Biology and Evolution, 21* (7), 1350-1360. doi:10.1093/molbev/msh134

Blair, S. S. (2009). Imaginal discs. In V. H. Resh (Ed.),*Encyclopedia of insects* (pp. 489-492): Academic Press Inc.

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *Peerj, 4* . doi:10.7717/peerj.1660

Bosco-Drayon, V., Poidevin, M., Boneca, I. G., Narbonne-Reveau, K., Royet, J., & Charroux, B. (2012). Peptidoglycan sensing by the receptor PGRP-LE in the *Drosophila* gut induces immune responses to infectious bacteria and tolerance to microbiota. *Cell Host & Microbe, 12* (2), 153-165. doi:10.1016/j.chom.2012.06.002

Buckling, A., & Hodgson, D. J. (2007). Short-term rates of parasite evolution predict the evolution of host diversity. *Journal of Evolutionary Biology, 20* (5), 1682-1688. doi:DOI 10.1111/j.1420-9101.2007.01402.x

Buckling, A., & Rainey, P. B. (2002). The role of parasites in sympatric and allopatric host diversification. *Nature, 420* (6915), 496-499. doi:10.1038/nature01164

Bulet, P., Hetru, C., Dimarcq, J. L., & Hoffmann, D. (1999). Antimicrobial peptides in insects; structure and function.*Developmental and Comparative Immunology, 23* (4-5), 329-344. doi:10.1016/s0145-305x(99)00015-4

Bustamante, C. D., Wakeley, J., Sawyer, S., & Hartl, D. L. (2001). Directional selection and the site-frequency spectrum. *Genetics, 159* (4), 1779-1788. doi:10.1093/genetics/159.4.1779

Carreira, V. P., Mensch, J., & Fanara, J. J. (2009). Body size in*Drosophila* : genetic architecture, allometries and sexual dimorphism. *Heredity, 102* (3), 246-256. doi:10.1038/hdy.2008.117

Carton, Y., Poirié, M., & Nappi, A. J. (2008). Insect immune resistance to parasitoids. *Insect Science, 15* (1), 67-87. doi:10.1111/j.1744-7917.2008.00188.x

De Roode, J. C., & Lefèvre, T. (2012). Behavioral immunity in insects.*Insects, 3* , 789-820. doi:10.3390/insects3030789

Deshmukh, R., Baral, S., Gandhimathi, A., Kuwalekar, M., & Kunte, K. (2018). Mimicry in butterflies: co-option and a bag of magnificent developmental genetic tricks. *Wiley Interdisciplinary Reviews-Developmental Biology, 7* (1). doi:10.1002/wdev.291

Eyre-Walker, A. (2002). Changing effective population size and the McDonald-Kreitman test. *Genetics, 162* (4), 2017-2024. doi:10.1093/genetics/162.4.2017

Fay, J. C., Wyckoff, G. J., & Wu, C. I. (2001). Positive and negative selection on the human genome. *Genetics, 158* (3), 1227-1234. doi:10.1093/genetics/158.3.1227

Fay, J. C., Wyckoff, G. J., & Wu, C. I. (2002). Testing the neutral theory of molecular evolution with genomic data from *Drosophila* .*Nature, 415* (6875), 1024-1026. doi:10.1038/4151024a

Fors, L., Markus, R., Theopold, U., Ericson, L., & Hambäck, P. A. (2016). Geographic variation and trade-offs in parasitoid virulence.*Journal of Animal Ecology, 85* (6), 1595-1604. doi:10.1111/1365-2656.12579

Fors, L., Markus, R., Theopold, U., & Hambäck, P. A. (2014). Differences in cellular immune competence explain parasitoid resistance for two coleopteran species. *Plos One, 9* (9), e108795. doi:10.1371/journal.pone.0108795

Garrison, E., & Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing.* ArXiv, 1207.3907.

Gillespie, J. P., Kanost, M. R., & Trenczek, T. (1997). Biological mediators of insect immunity. *Annual Review of Entomology, 42* , 611-643. doi:10.1146/annurev.ento.42.1.611

Godfray, H. C. J. (1994). *Parasitoids: behavioral and evolutionary ecology* . Princeton: Princeton University Press.

Grossmann, S., Bauer, S., Robinson, P. N., & Vingron, M. (2007). Improved detection of overrepresentation of Gene-Ontology annotations with parent-child analysis. *Bioinformatics, 23* (22), 3024-3031. doi:10.1093/bioinformatics/btm440

Hahn, M. W. (2019). *Molecular Population Genetics, 1st ed.* : Oxford University Press.

Hambäck, P. A. (2004). Why purple loosestrife in sweet gale shrubs are less attacked by herbivorous beetles? (in swedish with english abstract). *Entomologisk Tidskrift, 125* , 93-102.

Hambäck, P. A., Weingartner, E., Ericson, L., Fors, L., Cassel-Lundhagen, A., Stenberg, J. A., & Bergsten, J. (2013). Bayesian species delimitation reveals generalist and specialist parasitic wasps on *Galerucella* beetles (Chrysomelidae): sorting by herbivore or plant host. *BMC Evolutionary Biology, 13* . doi:10.1186/1471-2148-13-92

Hawkins, B. A., Cornell, H. V., & Hochberg, M. E. (1997). Predators, parasitoids, and pathogens as mortality agents in phytophagous insect populations. *Ecology, 78* , 2145-2152. doi:10.2307/2265951

Heger, A., & Ponting, C. P. (2007). Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. *Genome Research, 17* (12), 1837-1849. doi:10.1101/gr.6249707

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., . . . Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research, 47* (D1), D309-D314. doi:10.1093/nar/gky1085

Ito, K., Katsuma, S., Yamamoto, K., Kadono-Okuda, K., Mita, K., & Shimada, T. (2010). Yellow-e determines the color pattern of larval head and tail spots of the silkworm *Bombyx mori* . *Journal of Biological Chemistry, 285* (8), 5624-5629. doi:10.1074/jbc.M109.035741

Keehnen, N. L. P., Hill, J., Nylin, S., & Wheat, C. W. (2018). Microevolutionary selection dynamics acting on immune genes of the green-veined white butterfly, *Pieris napi* . *Molecular Ecology, 27* (13), 2807-2822. doi:10.1111/mec.14722

Keehnen, N. L. P., Rolff, J., Theopold, U., & Wheat, C. W. (2017). Insect antimicrobial defences: A brief history, recent findings, biases, and a way forward in evolutionary studies. In P. Ligoxygakis (Ed.),*Insect Immunity* (Vol. 52, pp. 1-33).

Kim-Jo, C., Gatti, J. L., & Poirié, M. (2019). *Drosophila*cellular immunity against parasitoid wasps: A complex and time-dependent process. *Frontiers in Physiology, 10* , 603. doi:10.3389/fphys.2019.00603

Kleinhesselink, K., Conway, C., Sholer, D., Huang, I., & Kimbrell, D. A. (2011). Regulation of hemocytes in *Drosophila* requires dappled Cytochrome b5. *Biochemical Genetics, 49* (5-6), 329-351. doi:10.1007/s10528-010-9411-7

Korunes, K. L., & Samuk, K. (2021). pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data.*Molecular Ecology Resources, 21* (4), 1359-1368. doi:10.1111/1755-0998.13326

Kraaijeveld, A. R., Van Alphen, J. J. M., & Godfray, H. C. J. (1998). The coevolution of host resistance and parasitoid virulence.*Parasitology, 116* , S29-S45. doi:10.1017/S0031182000084924

Lemaitre, B., & Hoffmann, J. (2007). The host defense of*Drosophila melanogaster* . *Annual Review of Immunology, 25* , 697-743. doi:10.1146/annurev.immunol.25.022106.141615

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data, P. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25* (16), 2078-2079. doi:10.1093/bioinformatics/btp352

Linnen, C. R., O'Quin, C. T., Shackleford, T., Sears, C. R., & Lindstedt, C. (2018). Genetic basis of body color and spotting pattern in redheaded pine sawfly larvae (*Neodiprion lecontei*).*Genetics, 209* (1), 291-305. doi:10.1534/genetics.118.300793

Luo, H., Rose, P. E., Roberts, T. M., & Dearolf, C. R. (2002). The Hopscotch Jak kinase requires the Raf pathway to promote blood cell activation and differentiation in *Drosophila* . *Molecular Genetics and Genomics, 267* (1), 57-63. doi:10.1007/s00438-001-0632-7

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the ADH locus in *Drosophila* . *Nature, 351* (6328), 652-654. doi:10.1038/351652a0

Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., . . . Cargill, M. (2005). A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biology, 3* (6), 976-985. doi:10.1371/journal.pbio.0030170

Proschel, M., Zhang, Z., & Parsch, J. (2006). Widespread adaptive evolution of *Drosophila* genes with sex-biased expression.*Genetics, 174* (2), 893-900. doi:10.1534/genetics.106.058008

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., . . . Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics, 81* (3), 559-575. doi:10.1086/519795

Roux, J., Privman, E., Moretti, S., Daub, J. T., Robinson-Rechavi, M., & Keller, L. (2014). Patterns of positive selection in seven ant genomes. *Molecular Biology and Evolution, 31* (7), 1661-1685. doi:10.1093/molbev/msu141

Rus, F., Kurucz, E., Markus, R., Sinenko, S. A., Laurinyecz, B., Pataki, C., . . . Ando, I. (2006). Expression pattern of Filamin-240 in*Drosophila* blood cells. *Gene Expression Patterns, 6* (8), 928-934. doi:10.1016/j.modgep.2006.03.005

Sackton, T. B., Lazzaro, B. P., Schlenke, T. A., Evans, J. D., Hultmark, D., & Clark, A. G. (2007). Dynamic evolution of the innate immune system in *Drosophila* . *Nature Genetics, 39* (12), 1461-1468. doi:10.1038/ng.2007.60

Salazar-Jaramillo, L., Paspati, A., van de Zande, L., Vermeulen, C. J., Schwander, T., & Wertheim, B. (2014). Evolution of a cellular immune response in *Drosophila* : A phenotypic and genomic comparative analysis. *Genome Biology and Evolution, 6* (2), 273-289. doi:10.1093/gbe/evu012

Sattath, S., Elyashiv, E., Kolodny, O., Rinott, Y., & Sella, G. (2011). Pervasive adaptive protein evolution apparent in diversity patterns around amino acid substitutions in *Drosophila simulans* .*Plos Genetics, 7* (2). doi:10.1371/journal.pgen.1001302

Sawyer, S. A., & Hartl, D. L. (1992). Population genetics of polymorphism and divergence. *Genetics, 132* (4), 1161-1176. doi:10.1093/genetics/132.4.1161

Schlenke, T. A., & Begun, D. J. (2003). Natural selection drives*Drosophila* immune system evolution. *Genetics, 164* (4), 1471-1480. doi:10.1093/genetics/164.4.1471

Sedlazeck, F. J., Rescheneder, P., & von Haeseler, A. (2013). NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics, 29* (21), 2790-2791. doi:10.1093/bioinformatics/btt468

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31* (19), 3210-3212. doi:10.1093/bioinformatics/btv351

Smith, N. G. C., & Eyre-Walker, A. (2002). Adaptive protein evolution in *Drosophila* . *Nature, 415* (6875), 1022-1024. doi:10.1038/4151022a

Stenberg, J. A., & Hambäck, P. A. (2010). Host species critical for offspring fitness and sex ratio for an oligophagous parasitoid: implications for host coexistence. *Bulletin of Entomological Research, 100* (6), 735-740. doi:10.1017/S0007485310000143

Supek, F., Bošnjak, M., Škunca, N., & Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *Plos One, 6* (7). doi:10.1371/journal.pone.0021800

Unckless, R. L., Howick, V. M., & Lazzaro, B. P. (2016). Convergent balancing selection on an antimicrobial peptide in *Drosophila* .*Current Biology, 26* (2), 257-262. doi:10.1016/j.cub.2015.11.063

Unckless, R. L., & Lazzaro, B. P. (2016). The potential for adaptive maintenance of diversity in insect antimicrobial peptides.*Philosophical Transactions of the Royal Society B-Biological Sciences, 371* (1695). doi:10.1098/rstb.2015.0291

Waterhouse, R. M., Kriventseva, E. V., Meister, S., Xi, Z. Y., Alvarez, K. S., Bartholomay, L. C., . . . Christophides, G. K. (2007). Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science, 316* (5832), 1738-1743. doi:10.1126/science.1139862

Wertheim, B. (2022). Adaptations and counter-adaptations in*Drosophila* host–parasitoid interactions: advances in the molecular mechanisms. *Current Opinion in Insect Science, 51* , 100896. doi:10.1016/j.cois.2022.100896

Wertheim, B., Kraaijeveld, A. R., Schuster, E., Blanc, E., Hopkins, M., Pletcher, S. D., . . . Godfray, H. C. J. (2005). Genome-wide gene expression in response to parasitoid attack in *Drosophila* .*Genome Biology, 6* (11). doi:10.1186/gb-2005-6-11-r94

Woodard, S. H., Fischman, B. J., Venkat, A., Hudson, M. E., Varala, K., Cameron, S. A., . . . Robinson, G. E. (2011). Genes involved in convergent evolution of eusociality in bees. *Proceedings of the National Academy of Sciences of the United States of America, 108* (18), 7472-7477. doi:10.1073/pnas.1103457108

Xu, L., Dong, Z. B., Fang, L., Luo, Y. J., Wei, Z. Y., Guo, H. L., . . . Wang, Y. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species.*Nucleic Acids Research, 47* (W1), W52-W58. doi:10.1093/nar/gkz333

Yang, X., Fors, L., Slotte, T., Theopold, U., Binzer-Panchal, M., Wheat, C. W., & Hambäck, P. A. (2020). Differential expression of immune genes between two closely related beetle species with different immunocompetence following attack by *Asecodes parviclava* .*Genome Biology and Evolution, 12* (5), 522-534. doi:10.1093/gbe/evaa075

Yang, X., Slotte, T., Dainat, J., & Hambäck, P. A. (2021). Genome assemblies of three closely related leaf beetle species (*Galerucella* spp.). *G3-Genes Genomes Genetics, 11* (8). doi:10.1093/g3journal/jkab214

Yang, X., Wheat, C. W., Slotte, T., & Hambäck, P. A. (2022). Lineage-specific targets of positive selection in three leaf beetles with different defence capacity against a parasitic wasp. Available from https://www.ebi.ac.uk/ena/browser/home (Accession: PRJEB56839)

Zhao, S. L., Zhang, T., Liu, Q., Wu, H., Su, B., Shi, P., & Chen, H. (2019). Identifying lineage-specific targets of natural selection by a Bayesian analysis of genomic polymorphisms and divergence from multiple species. *Molecular Biology and Evolution, 36* (6), 1302-1315. doi:10.1093/molbev/msz046

Data accessibility: Sequence data generated and analyzed during the current study have been deposited at the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under the accession PRJEB56839.

Author contributions: XY and PAH conceived the study, PAH collected the field material, XY performed the lab work and the bioinformatic analysis with guidance of TS and CWW, XY wrote the first draft and all authors provided comments on the manuscript.

Table 1. Summary results from the HDMKPRF test among three beetle species. Table shows the number of synonymous/ nonsynonymous sites variable among (D_s, D_n) and within (P_s, P_n) species summed across all 4154 single copy orthologs. The proportion of nonsynoymous sites fixed by positive selection (α = 1-D_sP_n/D_nP_s), and the proportion of genes under positive and negative selection estimated by HDMKPRF, for *G. calmariensis* , *G. pusilla* and*G. tenella* . Summary statistics were calculated before and after singleton

removal. Counts are for synonymous and non-synonymous sites divergent between (D_s andD_n , respectively) and polymorphic within (P_s, P_n , respectively) species.

Species	D_n	D_s	P_n	P_s
Before singleton removal	Before singleton removal	Before singleton removal	Before singleton removal	Before singleto
G. calmariensis	6211	12192	17752	30417
G. pusilla	4371	8273	17800	29955
G. tenella	3802	7128	20882	33347
After singleton removal	After singleton removal	After singleton removal	After singleton removal	After singleton
G. calmariensis	6434	12475	9537	18886
G. pusilla	4503	8444	11469	21630
G. tenella	3947	7287	13758	23184

Table 2. Positively selected immune genes in *Galerucellapusilla* (Gp), *G. tenella* (Gt) and *G. calmariensis*(Gc).

Species	**Recognition**	**Signalling**	**Effector**	**Protease**	**Haemato-poiesis**	**Melanisation**	**Wound healing**
Gp	*santa-maria*	*Tak1*		*CG32483*	*Raf*		
	Corin	*grass*			*cher*		
					zfh1		
Gt	*Corin*				*cher*		
	PGRP-LE				*Cyt-b5*		
Gc						*Cyp9f2*	

Figure legend

Figure 1. Posterior distributions of selection intensity for the three*Galerucella* species (A. *G. tenella* , B. *G. pusilla* , C. *G. calmariensis*) (Blue, 95% CI<0, Yellow, 95% CI>0, dotted line = zero selection intensity).

Figure 1