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

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

[1]Stockholm University

June 28, 2023

Abstract

Parasitoid wasps are major causes of mortality of many species, resulting in host immune defences commonly being the target of adaptive evolution, though such targets outside model species are poorly understood. Here we compare the power of different molecular tests of selection to provide such insights in novel species. We combined our understanding of variation in immune defence capacity among three closely related Galerucella leaf beetles with a shared parasitoid wasp, with information on genomic targets of parasitoid attacks from exemplar insect species. Based on this, we predicted that these genomic targets would vary in their evolutionary history across three closely related leaf beetle species, such that genomic targets would experience stronger positive selection in the species with strongest immune response to attack. Codon based tests revealed variation among species in positive selection genome wide, and showed that parasitoid-relevant immune genes experienced more positive selection in the species with the greatest immunocompetence (G. pusilla), while almost no immune genes were under positive selection in the species with the least immunocompetence (G. calmariensis). Genome wide analyses of the haplotype frequency spectrum also identified genes experiencing positive selection across the species, though few were parasitoid-relevant immune genes and no species was particularly enriched for them. Thus, our codon based test, which summarizes all sweep events since the last common ancestor, found results consistent with our a priori hypothesis, providing a series of targets for future functional genomic study.

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

Xuyue Yang[1], Kalle Tunström[2], Tanja Slotte[1#], Christopher W. Wheat[2#]and 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

[#]Shared senior author

Abstract

Parasitoid wasps are major causes of mortality of many species, resulting in host immune defences commonly being the target of adaptive evolution, though such targets outside model species are poorly understood. Here we compare the power of different molecular tests of selection to provide such insights in novel species. We combined our understanding of variation in immune defence capacity among three closely related *Galerucella* leaf beetles with a shared parasitoid wasp, with information on genomic targets of parasitoid attacks from exemplar insect species. Based on this, we predicted that these genomic targets would vary in their evolutionary history across three closely related leaf beetle species, such that genomic targets would experience stronger positive selection in the species with strongest immune response to attack. Codon based tests revealed variation among species in positive selection genome wide, and showed that parasitoid-relevant

immune genes experienced more positive selection in the species with the greatest immunocompetence (*G. pusilla*), while almost no immune genes were under positive selection in the species with the least immunocompetence (*G. calmariensis*). Genome wide analyses of the haplotype frequency spectrum also identified genes experiencing positive selection across the species, though few were parasitoid-relevant immune genes and no species was particularly enriched for them. Thus, our codon based test, which summarizes all sweep events since the last common ancestor, found results consistent with our *a priori* hypothesis, providing a series of targets for future functional genomic study.

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

Introduction

Parasitoid attack always leads to the death of either host or parasitoid, and genes affecting the host capacity to defend themselves against the intruder should therefore evolve rapidly (Kraaijeveld, Van Alphen, & Godfray, 1998). In insects being attacked by endoparasitoid wasps, which inject the egg inside the host body, host survival depends on an immunological reaction that ends in the encapsulation of the enemy egg (Carton, Poirié, & Nappi, 2008). Hence, a major task in evolutionary studies of host-parasitoid systems is to identify the genes underlying phenotypic differences in immunocompetence (Wertheim, 2015). Work on model organisms, such as *Drosophila melanogaster* , and a range of non-model organisms have identified key pathways involved in the immunological process (Carton et al., 2008). We also know that host species or genotypes may differ greatly in their defence capacity (Carton et al., 2008; Fors, Markus, Theopold, & Hambäck, 2014; Wertheim, 2022), but we lack information outside of the *Drosophila* system on the importance of canonical immune genes as determinants of defence phenotypes.

In our work, we have observed large phenotypic differences between three closely related leaf beetle species (Chrysomelidae: *Galerucella*spp.) in defense capacity when attacked by the same parasitoid wasp species (*Asecodes parviclava* , Eulophidae). Whereas *Galerucella pusilla* shows a high capacity to encapsulate wasp eggs, this capacity is much lower in *G. tenella* and almost absent in *G. calmariensis* (Fors et al., 2014). These defense phenotypes have also been connected to differences in the induction of those immune cells involved in the encapsulation of wasp eggs (Fors et al., 2014) and in the expression of immune-related genes following parasitoid attack (Yang et al., 2020), suggesting a genetic basis for the observed differences in immune performance. Importantly, all three of these closely related species are attacked by the same parasitoid, suggesting that these large differences in immunocapacity evolved recently. Thus, this system provides a promising opportunity outside of *Drosophila* to connect defense phenotypes and genotypic differences by identifying genes under positive selection. Within this top-down framework, here we ask whether species-level differences in phenotypes with clear connections to Darwinian fitness in response to a common, strong selective pressure are mirrored in molecular tests of selection, with the expectation that parasitoid-relevant immune genes in the most defended species should be enriched for positive selection.

The insect immune system consists of two parts, where the humoral system is mainly active against pathogens (Bulet, Hetru, Dimarcq, & Hoffmann, 1999; Gillespie, Kanost, & Trenczek, 1997), whereas the cellular immune system seems more important when encapsulating parasitoid wasp eggs (Carton et al., 2008). The cellular immunity can be separated into seven broad functional categories: recognition, signalling, effector, proteases, haematopoiesis, melanisation, and wound healing (see also Yang et al., 2020) and particularly genes 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). However, selection dynamics are complex and depend on specific functions and taxa studied (Keehnen, Hill, Nylin, & Wheat, 2018; Keehnen, Rolff, Theopold, & Wheat, 2017), but some classes show more evidence of positive selection while other classes such as AMPs appear to experience more balancing or purifying selection (Heger & Ponting, 2007; Unckless, Howick, & Lazzaro, 2016; Unckless & Lazzaro, 2016; Waterhouse et al., 2007). So far, however, studies focused upon the selection dynamics of parasitoid specific immune genes have received less attention. Here we make the prediction that species exhibiting the strongest immunocompetence against wasp attack (*G. pusilla*) will also exhibit stronger positive selection on relevant immune genes compared with the species with lower immunocompetence, as we expect immunocompetence to have evolved via historical selection upon

relevant genes.

Apart from defence phenotypes, there are also other phenotypic differences between the three species of unknown or absent relation to parasitoid defence capacity. First, *G. calmariensis* is slightly larger and develops faster than the other two species, indicating potential differences in development and metabolic pathways. Second, 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), which may be traced to differences in gene pathways related to pigmentation and melanisation (Ito et al., 2010; Linnen, O'Quin, Shackleford, Sears, & Lindstedt, 2018). Finally, the species likely 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. Thus, each species is expected to exhibit unique signatures of selection, highlighting the importance of our *a priori* hypothesis framework with regards to parasitoid-relevant immune genes.

To detect selection, we employ two molecular tests that are reliant upon population level genetic variation but primarily differ in their genomic focus (coding regions vs. genome wide), and the time scale over which they have the most statistical power. For coding regions, 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 nonsynonymous and synonymous polymorphism (P_n/P_s) within species to the ratio of nonsynonymous and synonymous fixed differences (D_n/D_s) between species (McDonald & Kreitman, 1991). When positive selection favours the phenotypic impact of a novel amino acid change, the corresponding advantageous allele goes to fixation, and when this happens repeatedly, 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 at genes can be estimated using α (= 1-D_sP_n/D_nP_s) (Smith & Eyre-Walker, 2002), which represents the proportion of non-synonymous substitutions driven by positive selection during divergence between focal species and an outgroup, allowing for a powerful test of selection dynamics.

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 these limitations, we here 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.

For regulatory regions, we employed a genome wide scan of positive selection dynamics that leverages information in the frequency of genetic variation found within populations of a given species to detect both soft and hard selective sweeps (aka positive selection). Due to the reliance on the site frequency spectrum (SFS), this test covers a more recent time history of selection compared to the aforementioned MK test, yet allows for the detection of selection anywhere in the genome, such as in the regulatory region flanking genes. We investigated genome signatures of positive selection using LASSI Plus (Harris & DeGiorgio, 2020). This method estimates haplotype frequency spectrum statistics within sliding windows of population genomic data to detect soft and hard sweeps, with stronger power for the latter category due to the more dramatic signature left in the SFS (Harris, Garud, & DeGiorgio, 2018).

To test our primary hypothesis 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*), we quantified selection dynamics in our study system using the two molecular tests of selection (HDMKPRF and LASSI plus) using whole genome re-sequencing data from 15 individuals from

each of the three beetle species (*G. calmariensis* , *G. pusilla* and *G. tenella*).

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 (Supplementary Information, Fig. S1).

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, Markus, Theopold, Ericson, & Hambäck, 2016; Fors et al., 2014). These differences also match selection by wasp females to larval odour cues (Fors, Mozūraitis, Blažytė-Čereškienė, Verschut, & Hambäck, 2018)

We collected 45 adult individuals, 15 samples from each*Galerucella* species, during May and June 2019 from the following sites: three *G. calmariensis* populations: Norrfjärden (62°3'28"N, 17°26'18"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, which was the least fragmented 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 of all sites, 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) using Plink v1.9 (Purcell et al., 2007) across all the samples and for each species separately (Supporting information Figures S1 and S2). Genetic diversity (nucleotide polymorphism, π) was estimated for each species using pixy (Korunes & Samuk, 2021).

2.3 Codon based 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 our 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). A comprehensive justification for using HDMKPRF to detect genes under selection can be found in Okamura et al (2022), but a main advantage is the higher power for detecting weak and moderate selection. Compared with the classical MK-test, the higher power of both MKPRF and HDMKPRF is a consequence of adopting a Poisson random-field framework (Sawyer & Hartl, 1992) where per gene selection intensities are estimated using a Bayesian approach that combine information across multiple loci and derive posterior distributions. The meaning of selection intensities is adopted from Bustamante et al. (2001) and is akin to a neutrality index measure (Hahn, 2019). The HDMKPRF improves the MKPRF-test by simultaneously analysing polymorphism and divergence across multiple species, which allow the test to determine in which lineage that selection has occurred (Zhao et al., 2019). During this process, HDMKPRF additionally estimates population genetic parameters for each species, such as effective population sizes and mutation rates, which it uses to account for lineage specific differences in these parameters in its estimation of selection intensity per locus (for details see Zhao et al., 2019). The script (Table S1) for performing the HDMKPRF to derive selection intensities was kindly provided by Zhao et al. The input data for the analysis included all 4,154 SCOs and the analysis was implemented by first running 200,000 burn-in steps and thereafter posterior parameter distributions were estimated from 400,000 steps in a Markov chain Monte Carlo process with a thinning interval of 5, based on author recommendations. A gene was considered to be under positive selection when the 95% posterior credibility interval for the selection intensity was >0, and similarly under negative selection when the interval was <0. Because estimates of positive selection using population resequencing data are usually biased downward by the segregation of slightly deleterious mutations and some singleton errors (Bierne & Eyre-Walker, 2004; Fay, Wyckoff, & Wu, 2002), we removed

singleton polymorphisms from all gene sets using a custom script (Sattath, Elyashiv, Kolodny, Rinott, & Sella, 2011) and then tested the effect on the power of detecting adaptive evolution by comparing HDMKPRF to gene sets before and after singleton removal.

We performed a gene ontology (GO) enrichment test of biological process terms of genes under 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

To analyse immune genes specifically, we first used BLASTP to detect the set of candidate genes identified in a previous RNA-seq study (Yang et al., 2020) in the *G. calmariensis* proteome. This gene set contains 166 genes suggested to be important in immune response against parasitoid wasp attack in *Drosophila* (Table S2), 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). The threshold used in BLASTP was an E-value[?]1×10^{-6} and a bitscore>60, which identified a set of 96 immune genes in our genomic dataset. When multiple hits were recovered during BLAST, we always used the one with the highest bit-score. When working to include these into our SCO gene analysis, only 40 of these genes passed this conservative threshold (they are part of the 4154 SCOs gene set). In order to more fully focus upon the 96 immune genes themselves, we performed a second HDMKPRF analysis. Because this second analysis is run on a smaller gene set, HDMKPRF has less power. We therefore compared the estimated selection intensities for the 40 SCOs between the two analyses, and confirmed that these were almost identical, suggesting this set size was sufficient for model parameter estimates.

2.5 Genome wide test of selection

To detect positive selection outside of the coding regions of genes, we used a maximum likelihood analysis of the haplotype frequency spectrum across the genome in order to identify putative the targets of positive selection via signatures of both soft and hard sweeps. For this purpose, we used LASSI Plus (Harris & DeGiorgio, 2020) and the saltiLASSI statistic (DeGiorgio & Szpiech, 2022). This approach is capable of using unphased sequencing data to infer haplotypes and identify genomic regions within population samples that exhibit greater than expected changes in their haplotype allele frequencies given background genomic patterns that are taken as neutrality. This method is able to both estimate the likelihood of a given haplotype sweeping, as well as the inferred width and number of haplotypes sweeping within a given species. To avoid reference bias, we aligned our short read sequencing data from 15 individuals of each species to their respective reference genomes (Yang et al., 2021) using the bwa-mem2 v.2.0pre2 (Vasimuddin, Misra, Li, & Aluru, 2019), and then called variants using bcftools v.1.13-35-ge3ba077 to generate an all-site vcf (Danecek et al., 2021). The resulting vcfs were filtered for low quality calls (QUAL > 30), read depth (5-50), no indels and to no more than 2 alternative alleles at a given site. Inferences of selective sweeps were made using the salti statistic in the LASSI Plus software package (k=10, window size 52, step size = 12). To identify any outlier windows across the genome, we extracted all windows with a salti statistic (L) higher than 4 standard deviations above the mean. L is a composite likelihood ratio test statistic of the haplotype frequency spectra in a given window being distorted relative to genomic background.

Finally, in order to compare the distribution and location of sweeps between our different species, while avoiding reference bias from aligning the samples to a single reference, we scaffolded each species genome against a common chromosome level assembly from a species in a sister genus, the beetle *Lochmaea crataegi*

(NCBI: GCA_947563755.1). Scaffolding was performed using Ragtag v.2.1.0, with default settings using minimap2 (Alonge et al., 2022), with alignments filtered to remove any contigs shorter than 50kb. To identify outlier loci (i.e. those likely to have experienced a sweep), we extracted all windows with a salti statistic (L) greater than 4 standard deviations above the mean.

2.6 Comparison of results from molecular tests of selection

Comparisons of genome-wide targets of positive selection across species used two approaches. First, we identified genes within any of the identified sweep regions, by intersecting the outlier sweep locations with gene annotations using bedtools intersect, v2.27.1 (Quinlan & Hall, 2010), for each species genome (Yang et al., 2021). The protein sequences of any gene, from start to stop, overlapping any sweep region interval, was then retrieved for each species. We then assessed the overlap of the sets of proteins identified for the three species using Orthovenn2 (Xu et al., 2019), which clusters proteins based upon sequence similarity. This approach provides a means of identifying any targets of selection that might be shared across the three species, without also requiring targeting of the same gene region (because this analysis allows for independent members of a given gene family to be targeted). In order to assess whether any species appeared to have a higher proportion of immune genes among the putative targets of positive selection, we also included the 96 candidate immune genes previously identified from *G. calmariensis* in the Orthovenn2 analysis. Second, we repeated the analysis above, but extended the candidate gene region to included 5kb on either side of the gene body (i.e. 5 kb before start and after stop), in order to detect any putative signatures of positive selection associated with the regulatory regions of a given gene. We chose a 10kb flanking region targeted for regulatory evolution based on information from other insect groups (Ghavi-Helm et al., 2014; Lewis & Reed, 2019).

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). Species were strongly separated (Fig. S2) and there were some additional within-species separation between sites (Fig. S3). 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 Codon based test of 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* , weak 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.

Out of the 4154 genes, the HDMKPRF identified a similar number of genes under selection in the three *Galerucella* species (Fig. 1). In*G. pusilla* , 469 and 562 genes were identified as being under positive and negative selection, respectively. In *G. calmariensis* , 665 and 598 genes were identified as being under positive and negative selection, respectively. Finally, in *G. tenella* , 442 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, we focus our analyses on 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 (Tables S3-S5). 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) (Table S3). Third, unique gene categories under positive selection in *G. pusilla* included those coding for positive regulation of the innate immune response, including positive regulation of small GTPase mediated signal transduction, and those coding for axoneme assembly (e.g., cilium movement, cilium organization) (Table S5). 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 (Table S4).

3.3 Candidate gene analysis

The 96 candidate immune genes (out of 166) that remained in our second HDMKPRF analysis include 13/17 recognition genes, 21/35 signalling genes, 4/21 effectors, 16/35 protease coding genes, 24/31 haematopoiesis genes, 13/18 melanisation genes and 5/9 wound healing genes (Table S2). Seven of these immune genes were found to be under positive selection in*G. pusilla* , which was significantly more than the single gene identified under positive selection in *G. calmariensis* (Fisher's Exact test: Odds ratio = 0.10, $p<0.02$), but not significantly greater than the four genes under positive selection in *G. tenella* (Odds ratio = 0.61, $p>0.3$), when related to all genes under positive selection in respective species. However, while the number of genes under positive selection in *G. tenella* was marginally higher than the genes under positive selection in *G. calmariensis* , this was not significant (Odds ratio = 0.17, $p<0.09$). Positively selected immune genes in *G. pusilla*include 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 (*cher* and *zfh1*) (Table 2, Table S6). The four immune genes under positive selection in *G. tenella* included 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, Table S6). Finally, the single gene under positive selection in *G. calmariensis* was *Cyp9f2* , which is involved in melanization.

3.4. Genome wide test of selection

While we were able to detect signatures of selective sweeps in all three species, the patterns were strikingly different between *G. calmariensis* and the other two species (Fig. 2). In *G.calmariensis* , the outlier loci were concentrated in a handful of locations, which had strong signatures of selection large and were large (Fig. 3). In contrast, both *G. tenella* and *G. pusilla*exhibited signatures of selection that were generally weaker, with identified genomic regions that were more evenly distributed across their genomes (Fig. 2) and narrower in *G. tenella* and *G. pusilla* than in *G. calmariensis* (Fig. 3). To make a direct comparison with the HDMKPFR test that is focused upon the coding region of genes, we identified genes having a direct overlap of the sweep region with any part of a gene body, from the start to stop, and in between. This intersection of the outlier haplotypes with annotated genes revealed a total of 192, 115, and 154 genes in *G.calmariensis, G. pusilla* and *G. tenella,* respectively. When we included 10 kb on either side of each gene when detecting overlap with outlier haplotypes, a total of 216, 305, and 430 genes were intersected, suggesting that in the

latter two species, there was more overlap with potential regulatory regions. We next sought to assess to what extent the identified candidate targets of positive selection from the genome wide analysis overlapped with our candidate immune genes. Unlike our results for the HDPRFMK test, the distribution of these outlier genes (with 10kb on either side) were roughly equally distributed among the three species (7, 6, and 6 genes in *G.calmariensis, G. pusilla* and *G. tenella,* respectively).

3.5. Comparison of results from molecular tests of selection

In order to compare results from our two tests of positive selection, via the HDMKPRF and LASSI Plus analyses, we necessarily must focus upon*G. calmariensis* . In order to ensure locus orthology in our HDMKPRF analyses, we both identified 1:1:1 orthologs among the three species and mapped all of our population data to this reference and used a single annotation (i.e. all gene information in with reference to the*G. calmariensis* genome). However, for our LASSI Plus analysis, we mapped reads to each species genome, in order to be able to accurately assess the haplotype frequency spectrum outside of coding regions. Thus, we are only able to accurately compare results from both analyses for *G. calmariensis* . While 595 genes were identified as having experienced positive selection via the first HDMKPRF, that included all 4151 genes, only 10 of these overlapped with the set of 216 genes found to be overlapping with haplotype sweep regions (with genes having a 10kb flaking region for overlap assessment) and none of these immune genes. However, the comparison with the second HDMKPRF that included all 96 candidate immune genes identified a larger overlap, with 7,6 and 6 genes identified respectively.

4. Discussion

We have investigated whether documented differences in defence phenotypes between three closely related species that are attacked by the same parasitoid wasp species (*G. pusilla* >*G. tenella* >*G. calmariensis*), are reflected in genomic signatures of positive selection among them. We found that the relative ranking of species by immune performance was predictive of the number of immune genes exhibiting signatures of positive selection. However, this pattern differed depending on the type of test that we used, as well as by the region of the genome analysed by a given test.

Our analysis of the coding region of genes with a modified MK-test based approach showed that higher numbers of immune genes ($N = 7$) were under positive selection in the species with the strongest capacity to encapsulate wasp eggs (*G. pusilla*), compared with the species lacking the capacity to encapsulate wasp eggs (*G. calmariensis,* $N= 1$) or having an intermediate encapsulation capacity (*G. tenella,* $N = 4$). These findings suggest that evolutionary adaptations of host defences are not only a consequence of changes in regulatory genes but are also due to changes in protein-coding regions. Even though the identified genes have yet to be functionally characterized in these species, their importance in mediating defence against parasitoids was partly confirmed by previous comparative gene expression analyses (Yang et al., 2020). This results contrasts with the pattern from LASSI Plus, where an equal number of immune related genes were identified as outliers in the three species. While this pattern may in part be due to differences in the power of the two methods to detect selection at different time scales, it is also notable that most genes (4/6) identified as outliers in *G. pusilla* were also previously found as being differentially expressed following wasp attack whereas none of the genes were such for *G. calmariensis* .

Even though we could analyze only a subset of immune genes in the*Galerucella* genome (96 out of 166), the detected genes indicate the type of processes most likely under selection. Most importantly, the analysis suggested 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 melanization or wound healing processes seem to be more conserved. A key observation from the previous gene expression analyses was that attack by *A. parviclava* increased expression of multiple genes in *G. pusilla* and almost none in *G. calmariensis*(Yang et al., 2020). This finding indicates that differences in defense phenotypes between these species can best be explained by genes coding for proteins involved during the early phases of the immune process. The current analysis identified several promising candidates, such as*santa-maria* , *Corin* , *PGRP-LE* , *Tak1*

and *Grass* , that have been suggested to be involved in either the recognition or signalling processes. In addition, two of these genes (Grass and PGRP-LE) were are among those identified as outliers in LASSI Plus when including flanking regions. The specific functions for these five genes have not been identified in *Galerucella* , but some studies exist from *Drosophila* . First, both *santa-maria* and *Corin* have been suggested to encode proteins with scavenger receptor or serine-type endopeptidase activities (Cao & Jiang, 2018), where *Corin* has a transmembrane domain that may function as a receptor to the Toll pathway (Irving et al., 2001). Second, *PGRP-LE* encodes an intracellular protein that binds to diaminopimelic acid-type peptidoglycans to activate the IMD/Relish pathway (Bosco-Drayon et al., 2012), but the function may primarily be anti-bacterial (Libert, Chao, Chu, & Pletcher, 2006). Third, *Grass* has 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), whereas *Tak1* may be involved in the change between the JNK and IMD pathways. Among these genes, a particularly interesting candidate is *Corin* , which was under positive selection in both *G. pusilla* and *G. tenella* , the two species with relatively more efficient encapsulation responses. Even though *G. tenella* was not included in the gene expression analysis, we know from other studies that this species has an encapsulation capacity that is intermediate between the two other species (Fors et al., 2016). Previous work on *Drosophila* show upregulation of *Corin* following bacterial infections (Irving et al., 2001), but not following parasitoid attack (Salazar Jaramillo et al., 2017). Unfortunately, this gene was data deficient in our gene expression analysis.

These recognition and signaling genes are sufficient for explaining differences in defense phenotypes, but we also detected positive selection on genes involved in downstream regulation of hemocytes. First, *cher* was under positive selection in both *G. tenella* and *G. pusilla* , and has previously been shown to negatively regulate lamellocyte differentiation (Rus et al., 2006). Lamellocytes are key elements in the capsules killing parasitoid eggs of *Galerucella* and are induced from precursors after attack by *A. parviclava* (Fors et al., 2016; Fors et al., 2014), and similarly induced by Drosophila following parasitoid attack (Kim-Jo, Gatti, & Poirié, 2019). Two other genes involved in lamellocyte regulation similarly have evidence for positive selection, one in *G. pusilla* (*zfh1*) and one on *G. tenella* (*Cyt-b5*). *zfh1* is one element in a transcription factor cascade that plays a role as the switch between plasmatocyte and lamellocyte fate (Frandsen, Gunn, Muratoglu, Fossett, & Newfeld, 2008), and *Cyt-b5* encodes a conserved hemoprotein that is required for hemocyte regulation (Kleinhesselink, Conway, Sholer, Huang, & Kimbrell, 2011). These two latter genes were also differentially expressed following wasp infection involving *A. parviclava* and *G. pusilla* (Yang et al., 2020).

When examining other gene functions identified as experiencing positive selection, we found that 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.

With regards to our HDPRFMK tests, there are some methodological concerns warranting attention. 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. Whether these removals have biased our conclusions is unclear, but 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 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 . Third, an additional issue when examining selection patterns is slightly deleterious mutations. If such mutations are segregating in the population, the degree of positive selection will be underestimated. Removing deleterious mutations during selection analysis is therefore widely accepted, by setting a threshold for removing minor alleles (Bierne & Eyre-Walker, 2004; Fay, Wyckoff, & Wu, 2001; Fay et al., 2002). We only removed singleton polymorphisms, which increased the proportion of negatively selected sites but retained 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.

Our investigation of signatures of positive selection from genome wide analysis in each of the three species failed to reveal an enrichment of immune genes biased towards the most defended species, similar to our predictions and what was observed in the HDPRFMK test. Because we assessed the overlap of the identified sweep regions with both gene bodies, as well as up to 10kb flanking regions as a proxy for gene regulatory regions, these findings are not because we limited our sweep detection to either regulatory or coding regions; we captured putative candidate sweep regions for both. Thus, we conclude that differences in the results from the two tests are more reflective of their inherent differences, which are diverse.

The HDPRFMK test could be considered to be much more conservative than LASSI Plus, as it requires the fixation of multiple amino acid mutations and hence multiple rounds of selective sweeps in order to result in significant departures from neutral expectations. However, because the test uses all such events since a last common ancestor was shared with the other species in the analysis, while conservative, this length of time provides many opportunities for selective sweeps, making it a powerful test, yet narrow in the underlying scenario it can detect. In contrast, LASSI Plus, like nearly all tests reliant upon outliers in the site frequency spectrum, is only able to detect positive selection events on a much more recent time horizon, likely a small fraction of the time since the last common ancestor. While limited in the time horizon such a test can query over, most substantial selective sweeps are likely to be detected and these are identifiable nearly anywhere in the genome; LASSI Plus should be able to detect one of the many selective sweeps that build up the amino acid fixations underlying HDPRFMK tests, as well as sweeps happening anywhere else. In sum, while the HDPRFMK test is narrowly focused upon selection acting on amino acid variation, it gains power by covering a deep time horizon. LASSI Plus, while able to detect positive selection nearly anywhere in the genome, can only see events in the relatively recent past.

Conclusion

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 and other parasites. 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 focused upon coding gene evolution over a deeper time scale. In contrast, analyses of recent selection dynamics across the genome detected no differential evolution of the immune system among these species. 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 (Kraaijeveld et al., 1998). While the arms-race between host and their natural enemies is expected to increase the rate of divergence between allopatric populations (Buckling & Hodgson, 2007; Buckling & Rainey, 2002), we find a varying degree of support evidence for this hypothesis varies when assessed via molecular tests of selection, with tests covering older events being more concordant with expectations. It is likely that the phenotypic differences in encapsulation capacity initially evolved during periods when the beetles were geographically isolated and thereby coincided with an allopatric speciation process. As such, the expected selection dynamics shaping the immune system of the beetles are likely old, where the observed phenotypic and genotypic differences have arisen during or after the speciation process in a coevolutionary process between hosts and parasitoid wasps.

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)* .

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., . . . Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing.*Genome Biol, 23* (1), 258. doi:10.1186/s13059-022-02823-7

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

Cao, X., & Jiang, H. (2018). Building a platform for predicting functions of serine protease-related proteins in Drosophila melanogaster and other insects. *Insect Biochem Mol Biol, 103* , 53-69. doi:10.1016/j.ibmb.2018.10.006

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

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., . . . Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience, 10* (2). doi:10.1093/gigascience/giab008

DeGiorgio, M., & Szpiech, Z. A. (2022). A spatially aware likelihood test to detect sweeps from haplotype distributions. *Plos Genetics, 18* (4), e1010134. doi:10.1371/journal.pgen.1010134

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

Fors, L., Mozūraitis, R., Blažytė-Čereškienė, L., Verschut, T. A., & Hambäck, P. A. (2018). Selection by parasitoid females among closely related hosts based on volatiles: Identifying relevant chemical cues.*Ecology and Evolution, 8* (6), 3219-3228. doi:10.1002/ece3.3877

Frandsen, J. L., Gunn, B., Muratoglu, S., Fossett, N., & Newfeld, S. J. (2008). Salmonella pathogenesis reveals that BMP signaling regulates blood cell homeostasis and immune responses in Drosophila. *Proc Natl Acad Sci U S A, 105* (39), 14952-14957. doi:10.1073/pnas.0808208105

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

Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., & Furlong, E. E. M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase.*Nature, 512* (7512), 96-+. doi:10.1038/nature13417

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

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

Harris, A. M., & DeGiorgio, M. (2020). A Likelihood Approach for Uncovering Selective Sweep Signatures from Haplotype Data.*Molecular Biology and Evolution, 37* (10), 3023-3046. doi:10.1093/molbev/msaa115

Harris, A. M., Garud, N. R., & DeGiorgio, M. (2018). Detection and Classification of Hard and Soft Sweeps from Unphased Genotypes by Multilocus Genotype Identity. *Genetics, 210* (4), 1429-1452. doi:10.1534/genetics.118.301502

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

Irving, P., Troxler, L., Heuer, T. S., Belvin, M., Kopczynski, C., Reichhart, J. M., . . . Hetru, C. (2001). A genome-wide analysis of immune responses in Drosophila. *Proc Natl Acad Sci U S A, 98* (26), 15119-15124. doi:10.1073/pnas.261573998

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

Lewis, J. J., & Reed, R. D. (2019). Genome-Wide Regulatory Adaptation Shapes Population-Level Genomic Landscapes in Heliconius.*Molecular Biology and Evolution, 36* (1), 159-173. doi:10.1093/molbev/msy209

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

Libert, S., Chao, Y., Chu, X., & Pletcher, S. D. (2006). Trade-offs between longevity and pathogen resistance in Drosophila melanogaster are mediated by NFkappaB signaling. *Aging Cell, 5* (6), 533-543. doi:10.1111/j.1474-9726.2006.00251.x

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

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

Okamura, Y., Dort, H., Reichelt, M., Tunstrom, K., Wheat, C. W., & Vogel, H. (2022). Testing hypotheses of a coevolutionary key innovation reveals a complex suite of traits involved in defusing the mustard oil bomb. *Proceedings of the National Academy of Sciences of the United States of America, 119* (51). doi:10.1073/pnas.2208447119

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

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26* (6), 841-842. doi:10.1093/bioinformatics/btq033

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., Jalvingh, K. M., de Haan, A., Kraaijeveld, K., Buermans, H., & Wertheim, B. (2017). Inter- and intra-species variation in genome-wide gene expression of Drosophila in response to parasitoid wasp attack. *Bmc Genomics, 18* . doi:10.1186/s12864-017-3697-3

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

Vasimuddin, M. D., Misra, S., Li, H., & Aluru, S. (2019).*Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems.* Paper presented at the IEEE Parallel and Distributed Processing Symposium (IPDPS).

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. (2015). Genomic basis of evolutionary change: evolving immunity. *Frontiers in Genetics, 6* . doi:10.3389/fgene.2015.00222

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, XY and KT performed the bioinformatic analysis with guidance of TS and CWW, XY and PAH 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 non-synonymous and synonymous polymorphism sites (P_n, P_s) within species to the ratio of nonsynonymous and synonymous fixed differences (D_n, D_s) between species summed across all 4154 single copy orthologs. The proportion of nonsynonymous 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.

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*	*cher*		
	Corin	*grass*			*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 2. Manhattan plot of haplotype frequency distortion across the genome for three beetle species. Y-axis in each plot is a measure of positive selection, captured by L, a measure of the haplotype frequency distortion. Chromosomes are coloured in alternatingly light and dark grey. Outlier loci (red) were calculated as being more than 4 standard deviations above the mean. Loci on autosomes and sex chromosomes for each species' scaffolds are plotted relative to the beetle*Lochmaea crataegi* , from a sister genus.

Figure 3. Histograms of the widths of identified sweep haplotypes in the genome of 3 species. **a.** A plot of the width all sweep haplotypes (significant and non-sig.; all dots in Fig. 2). **b.** A plot of only haplotypes whose frequency was significantly different from the mean (red dots in Fig. 2).

Figure 1

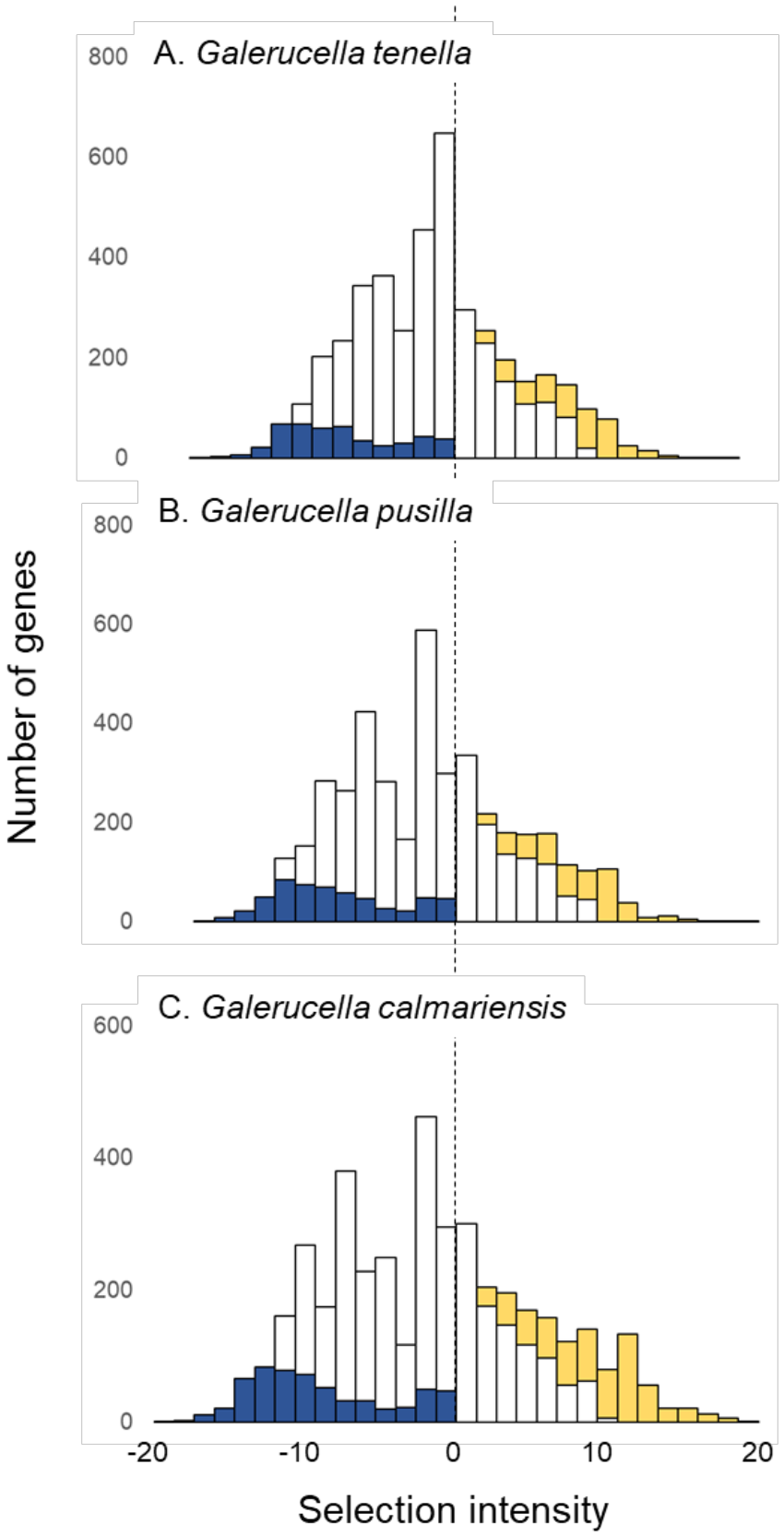

Figure 2.

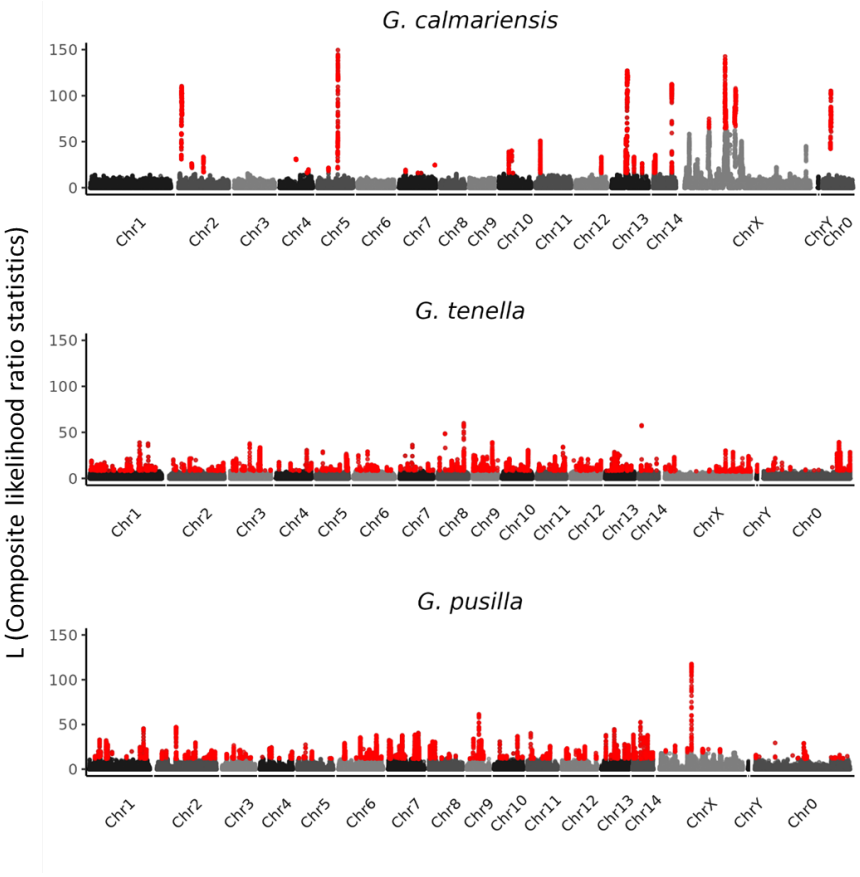

Figure 3.