


# Signatures of Adaptation, Constraints, and Potential Redundancy in the Canonical Immune Genes of a Key Pollinator

Sarah J. Larragy <sup>1,\*</sup>, Jannik S. Möllmann<sup>2</sup>, Jane C. Stout<sup>3</sup>, James C. Carolan<sup>1,†</sup>, and Thomas J. Colgan<sup>2,4,\*</sup>

<sup>1</sup>Department of Biology, Maynooth University, Maynooth, Co. Kildare, Ireland

<sup>2</sup>Institute for Organismic and Molecular Evolution, Johannes Gutenberg University Mainz, Mainz, Germany

<sup>3</sup>School of Natural Sciences, Trinity College Dublin, College Green, Co. Dublin, Ireland

<sup>4</sup>School of Biological, Earth and Environmental Sciences, University College Cork, Co. Cork, Ireland

<sup>†</sup>Joint senior authors.

\*Corresponding authors: E-mails: Sarah.Larragy@mu.ie; tcolgan@uni-mainz.de.

Accepted: 20 February 2023

## Abstract

All organisms require an immune system to recognize, differentiate, and defend against pathogens. From an evolutionary perspective, immune systems evolve under strong selective pressures exerted by fast-evolving pathogens. However, the functional diversity of the immune system means that different immune components and their associated genes may evolve under varying forms of selection. Insect pollinators, which provide essential ecosystem services, are an important system in which to understand how selection has shaped immune gene evolution as their populations are experiencing declines with pathogens highlighted as a potential contributing factor. To improve our understanding of the genetic variation found in the immune genes of an essential pollinator, we performed whole-genome resequencing of wild-caught *Bombus terrestris* males. We first assessed nucleotide diversity and extended haplotype homozygosity for canonical immune genes finding the strongest signatures of positive selection acting on genes involved in pathogen recognition and antiviral defense, possibly driven by growing pathogen spread in wild populations. We also identified immune genes evolving under strong purifying selection, highlighting potential constraints on the bumblebee immune system. Lastly, we highlight the potential loss of function alleles present in the immune genes of wild-caught haploid males, suggesting that such genes are potentially less essential for development and survival and represent redundancy in the gene repertoire of the bumblebee immune system. Collectively, our analysis provides novel insights into the recent evolutionary history of the immune system of a key pollinator, highlighting targets of selection, constraints to adaptation, and potential redundancy.

**Key words:** bumblebee, insect immunity, population genomics, selection, pollinator health.

## Significance

Pathogens have been identified as a potential contributing factor to the decline of many beneficial insect pollinators, including bumblebees. Although we know pathogens place strong selective pressures on their hosts, we still lack a thorough understanding of the genetic variation found in immune genes of wild pollinator populations and the types of selection shaping the insect immune system. Here, we find in a contemporary bumblebee population that functional immune gene classes are not only evolving under different selective pressures, identifying signatures of positive selection, but also evidence of evolutionary constraints and potential redundancy, which may have implications for their ability to respond to ongoing and future threats. Collectively, our analyses provide important and novel insights into the evolution of the immune system of an essential pollinator, furthering our understanding of the ability of a wild bumblebee to defend itself against pathogens.

## Introduction

Pathogens threaten all living organisms, leading to the evolution of host mechanisms to resist, tolerate or avoid pathogen interaction (Barreiro and Quintana-Murci 2010; Balloux and van Dorp 2017). Insects, in particular, have evolved several effective traits to prevent infection, including avoidance and hygienic behaviors, physical barriers, such as a tough exoskeleton, and chemical defenses (Siva-Jothy et al. 2005). However, if a pathogen can overcome these defenses and successfully invade its host, the last barrier to the establishment of infection is the insect innate immune system (Tsakas and Marmaras 2010). The insect immune system is comprised of two arms: the cellular response, coordinated by hemocytes, which implements phagocytosis, encapsulation, and melanization (Mavrouli et al. 2005; Strand 2008) and the humoral response, a protein-derived defense involved in the recognition of and response to pathogen-associated molecular patterns. The recognition of such molecules activates pathway cascades, such as the Toll pathway, which is activated on host recognition of gram-positive (Rutschmann et al. 2002) and fungal invaders (Roh et al. 2009). These cascades stimulate the production of effector molecules, such as antimicrobial peptides that target and eliminate invaders (Ali Mohammadi Kojour et al. 2020).

Genes involved in the immune system are often subject to strong selective pressures to defend against and co-evolve along with pathogens while also being able to distinguish pathogens from beneficial foreign products (e.g., ingested food), as well as symbiotic microbes (Royet et al. 2011; Nyholm and Graf 2012). Similarly, components of the immune system have evolved to recognize and differentiate between self and nonself to protect the host from its own defenses (Cooper and Eleftherianos 2017). The strong selective pressures placed on immune genes are reflected in the routine identification of these genes as fast evolving across diverse taxa through comparative genomic scans (Roux et al. 2014; Shultz and Sackton 2019). However, there are different types of selection, which often

produce characteristic patterns in the host genome. For example, positive selection produces genomic regions of reduced diversity through “selective sweeps,” whereby selected beneficial alleles, as well as neighboring sites, can increase in frequency within a population to the point of fixation (Nielsen et al. 2005). Recent studies on insects have applied population genetic approaches to investigate the genetic variability of immune genes, as well as to determine the different forms of selection acting upon them (Palmer et al. 2018; Tan et al. 2021). However, with a few exceptions, studies employing genome-wide population genomic techniques to better understand how selection has acted on different functional classes of the immune system are still lacking, especially for nonmodel organisms within the field of genetics (but see Tan et al. 2021).

For many ecologically and commercially beneficial insect species, such as pollinators, it is vitally important to understand the selective pressures acting on their immune genes as pathogens and associated disease are one of several factors suggested to contribute to population declines (Goulson 2010; Wagner 2020). Pathogen spread is hypothesized to be exacerbated by the use of commercially reared bumblebees. These bumblebee colonies, usually *Bombus terrestris* or *Bombus impatiens*, are reared in commercial facilities and exported around the globe to provide pollination services to crops, such as tomatoes (Velthuis and Van Doorn 2006). However, they are known to often carry bumblebee and honeybee pathogens (Graystock et al. 2013; Murray et al. 2013), and there is evidence supporting pathogen spread from commercial bumblebees to wild populations of many bee species (Colla et al. 2006; Murray et al. 2013). Declines in pollinator populations, including bumblebees (Fitzpatrick et al. 2007; Kosior et al. 2007; Cameron et al. 2011), have led to concern about the sustainability of pollination services provided by insects which greatly contribute to global economies, food security, and ecosystem functioning (Aizen et al. 2009; Stout et al. 2019). Our understanding of the capacity of wild bee populations to respond to ongoing environmental threats has

been greatly informed by recent studies using population genomics, which have highlighted the genetic variation harbored in a population, as well as the targets of recent positive selection acting throughout the genome (Harpur and Rehan 2021; Colgan et al. 2022; Heraghty et al. 2022; Kardum Hjort et al. 2022), including immune genes (Kent et al. 2018). However, despite the importance of hymenopteran insects such as bees, wasps, and ants, focused population genetic studies on their immune genes are limited. Addressing this knowledge gap is essential in understanding how these genes are evolving in response to the risks currently facing many valuable insects in their natural environments.

A valuable study system to examine how recent selection has shaped immune gene evolution is the earth or buff-tailed bumblebee, *Bombus terrestris*. Bumblebees (*Bombus* sp.), in general, are a particularly beneficial group of pollinators for a variety of commercial crops due to their variation in tongue length, ability to buzz pollinate, and tolerance to temperate climates (Velthuis and Van Doorn 2006). Bumblebees are also a classical system for studying insect immunity as they are social insects that live in thermoregulated, densely packed nests of closely related individuals, providing prime conditions for pathogen spread among nest mates (Shykoff and Schmid-Hempel 1991). This fact is further reflected by bumblebees serving as hosts for a taxonomically diverse group of pathogens, including viruses, bacteria, fungi, protozoans, nematodes, insect parasitoids, and social parasites (Rutrecht and Brown 2008). Bumblebees, like other holometabolous insects, also have morphologically and behaviorally distinct life stages that face different pathogenic threats and risks (Schmid-Hempel 1998). Similarly, being social insects, bumblebees have caste differentiation in the female sex with the evolution of queen and worker castes, which perform different colony-level tasks that likely influence their risk of pathogen exposure and infection (Colgan et al. 2011). At the genetic level, bees have a decreased set of canonical immune genes in comparison with model organisms, such as *Drosophila melanogaster*, suggesting that bees may have a reduced immune capacity to combat infection (Evans et al. 2006; Barribeau et al. 2015). Finally, like other Hymenoptera, sex determination is haplodiploid, with females developing from fertilized, diploid eggs and males from unfertilized, haploid eggs (Cook and Crozier 1995). The haploid status of the male means if a deleterious allele is inherited, they do not have an extra gene copy to rescue function (Hartl and Brown 1970). Therefore, males automatically express deleterious alleles that they carry and, therefore, selection acts strongly to remove these from a population (Joseph and Kirkpatrick 2004). Additionally, haploid individuals may carry alleles that are predicted to impact gene function and, simultaneously, exhibit no obvious impediment on fitness. The presence of these alleles in

what appears to be robust, healthy individuals may highlight genes with reduced, redundant, or nonessential roles in important biological processes.

In this study, we performed whole-genome resequencing of a natural population of *B. terrestris*, to understand the genetic variation maintained in their canonical immune genes and the selection pressures acting on them. We sampled male bumblebees across the island of Ireland and applied a population genomics-based approach to examine patterns of positive selection through estimating genome-wide levels of nucleotide diversity and extended haplotype homozygosity to provide evidence for recent selective sweeps acting on canonical immune genes (Barribeau et al. 2015; Sadd et al. 2015). Furthermore, we investigated evidence of evolutionary constraints through the characterisation of immune genes evolving under strong purifying selection. Lastly, to understand potential nonfunctional or redundant components of the bumblebee immune system, we investigated the presence and frequency of putative loss of function alleles carried by wild-caught adult males.

## Results

### Genetic Diversity Within and Across Bumblebee Canonical Immune Genes

As a provisional step to examine different selective pressures acting on canonical immune genes, we assessed the genetic diversity found within the sequenced genomes of 33 wild-sampled *B. terrestris* males. We identified 1,103,300 million high-quality biallelic single-nucleotide polymorphisms (SNPs) with a sliding window-based analysis (window size: 100 kb) estimating mean genome-wide nucleotide diversity ( $\pi$ ) as  $1.71 \times 10^{-3}$  (supplementary table S2a, Supplementary Material online), similar to diversity estimates calculated for

**Table 1**

Functional Annotation of Polymorphic Sites Found in the Immune Genes of a Wild-caught Population of *Bombus terrestris*

Variant category	Total number of SNPs	Total number of SNPs located in immune genes	Total number of SNPs located in nonimmune genes
Intronic region	627,882	8,442	619,440
5 kb Upstream/ downstream of a gene	209,505	3,038	206,467
Untranslated region (UTR)	34,078	641	33,437
Coding region (CDS)	34,428	726	33,702

NOTE.—The total number of genome-wide SNPs, the numbers of immune and nonimmune SNPs are shown for variant categories, including SNPs found in intronic regions, 5 kb up or downstream of a gene, within 5' or 3' untranslated regions (UTRs) and coding regions.

a British population of *B. terrestris* ( $1.51 \times 10^{-3}$ ; Colgan et al. 2022). In terms of the canonical immune genes that were analyzed ( $n = 166$  genes found on chromosomal scaffolds), we found no significant difference in terms of mean nucleotide diversity between them and either all other nonimmune genes (Wilcoxon rank sum test;  $W = 789815$ ,  $P = 0.098$ ) or a similar number of size-matched nonimmune genes (Wilcoxon rank sum test:  $W = 13668$ ,  $P = 0.397$ ). The most abundant variant type found in the immune genes was intronic ( $n = 8,442$  SNPs, 66% of total SNPs found in or within 5 kb of a canonical immune gene), which was also the most abundant SNP found genome-wide (table 1). We found immune genes have significantly less intronic SNPs compared with nonimmune genes ( $\chi^2_{df=1} = 541.6$ ,  $P < 2e^{-16}$ ) but significantly more SNPs in untranslated ( $\chi^2_{df=1} = 51.7$ ,  $P < 6.4e^{-10}$ ) and coding ( $\chi^2_{df=1} = 146.9$ ,  $P < 2e^{-16}$ ) regions.

### Canonical Immune Genes in Genomic Regions of Reduced or Elevated Diversity

As positive selection can produce genomic regions of reduced diversity generated by “hard” selective sweeps, we assessed genome-wide nucleotide diversity in 100 kb sliding windows. We found 77 windows with significantly reduced diversity ( $z$ -score  $< -1.96$ ;  $P < 0.05$ ), within which we found 23 canonical immune genes (supplementary table S2a, Supplementary Material online), which was not more than expected by random chance ( $\chi^2_{df=1} = 0.88$ ;  $P = 0.35$ ). Members of several immune gene families were identified in these low-diversity regions including: the small RNA regulatory pathway (SRRP: LOC100642865, LOC100644688, LOC100647119, LOC100647746, LOC100648299); autophagy-related genes (APHAG: LOC100642312, LOC100642580, LOC100650799); inhibitors of apoptosis (IAP: LOC100647366, LOC100647488); the IMD pathway (IMD\_PATH; LOC100631093, LOC100645550); and C-type lectins (CTL: LOC100644559, LOC100645018; fig. 1; supplementary table S2a, Supplementary Material online). Gene Ontology (GO) term enrichment analysis for these immune genes identified the top five most significantly enriched biological processes-associated terms (Fisher’s exact test;  $P < 0.05$ ) as “defense response to other organism,” “cellular catabolic process,” “regulation of defense response,” “immune response,” and “regulation of immune system process” (supplementary table S3, Supplementary Material online).

In these low-diversity regions, we also found three canonical immune genes that lacked evidence of polymorphic sites. We further identified an additional five immune genes in other parts of the genome with no SNPs (i.e., no genetic variation among sampled bees; supplementary table S2b, Supplementary Material online). In total, these eight genes included *FAS-associated factor 1* (LOC100645550), *ubiquitin-like-conjugating enzyme ATG3* (LOC100642312),

*caspase-1* (LOC100645000), *TGF-beta activated kinase 1* (LOC100631093), *plasminogen activator inhibitor 1 RNA-binding protein* (LOC100649628), *32 kDa beta-galactoside-binding lectin* (LOC100649028), *baculoviral IAP repeat-containing protein 5* (LOC100647478), and *trypsin-3* (LOC100642942).

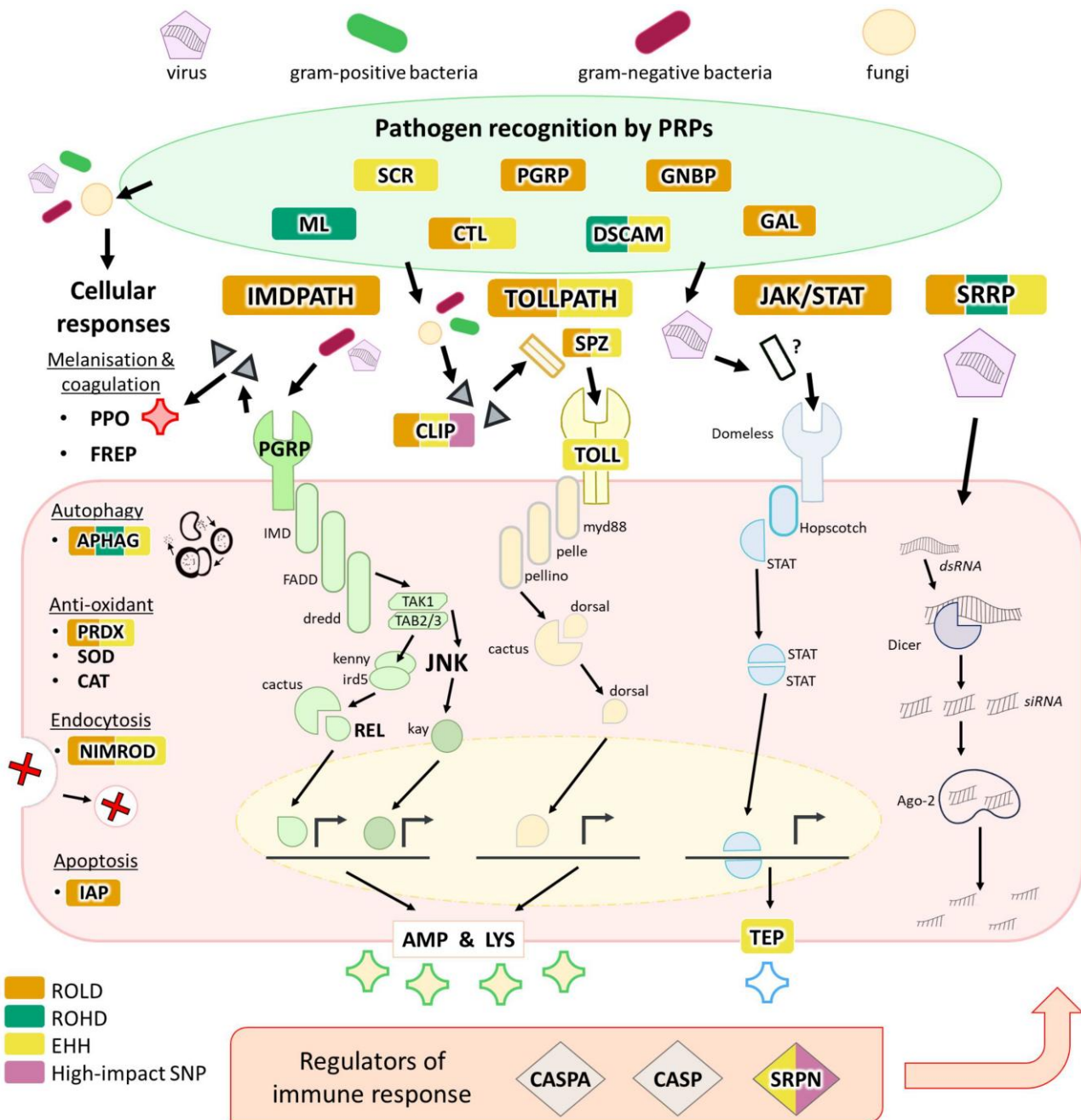
Given that many immune genes, especially those involved in pathogen recognition, may evolve under balancing selection and therefore have higher than expected polymorphisms, we investigated the presence of immune genes in regions of significantly elevated nucleotide diversity compared with the background genome. We found 127 genomic regions with elevated nucleotide diversity ( $z$ -score  $> 1.96$ ;  $P < 0.05$ ), which included five canonical immune genes (fig. 1); two APHAG gene family members (LOC100647690, LOC100644291); *Down syndrome cell adhesion molecule-like protein Dscam2* (immunoglobulin or IGG family; LOC100649765); *ecdysteroid-regulated 16 kDa protein* (MD-2-related lipid recognition or ML family; LOC100643802); and *GTP-binding nuclear protein Ran* (SRRP; LOC100646748). We found significantly more windows with elevated nucleotide diversity than expected by random chance (Binomial test:  $P < 0.0006$ ).

### Recent Positive Selection Acting on Autophagy and Scavenger Receptor Proteins

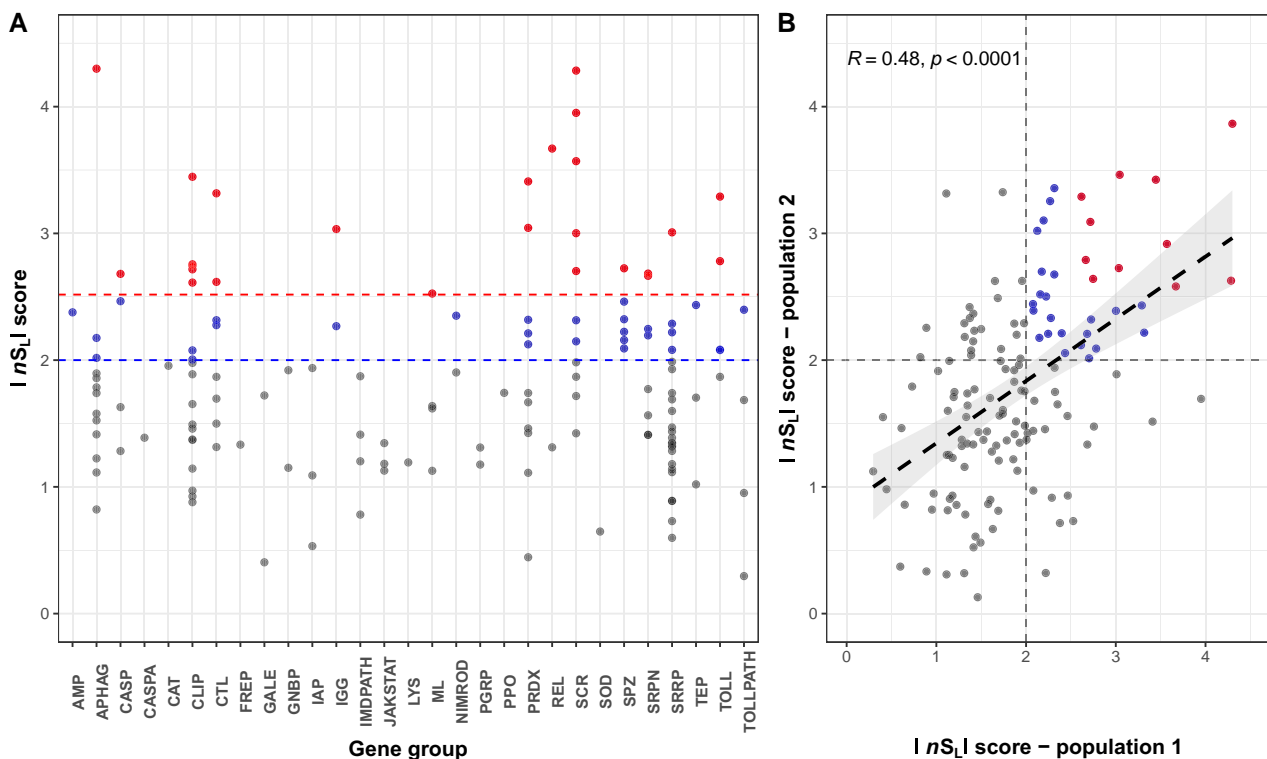
To identify immune genes with evidence of recent positive selection due to “soft” selective sweeps, whereby selection acts on standing genetic variation, we investigated genes with absolute  $nS_L$  ( $|nS_L|$ ) scores  $> 2$ , an indicator of recent selection (Ferrer-Admetlla et al. 2014; supplementary table S4a, Supplementary Material online). We identified 38 canonical immune genes with evidence of positive selection (fig. 2A; supplementary table S4b, Supplementary Material online) associated with immune gene groups involved in immune signaling ( $n = 9$  genes total: CLIP serine proteases [ $n = 5$ ]; spätzle proteins [ $n = 4$ ]; and Toll and Toll pathway [ $n = 4$ ]) and pathogen recognition genes ( $n = 11$  genes total: scavenger receptor proteins [SCR;  $n = 5$ ]; C-type lectins [ $n = 4$ ]; immunoglobulins [ $n = 2$ ]; and ML family member [ $n = 1$ ]; fig. 2). Signatures of positive selection were also identified for genes involved in detoxification (peroxidases [ $n = 5$ ]), immune regulation (serine protease inhibitors [ $n = 3$ ] and small RNA regulatory pathway [ $n = 1$ ]), autophagy ( $n = 1$ ), and a thioester containing protein ( $n = 1$ ).

The strongest evidence for positive selection was found for the *autophagy-related protein 16* gene ( $|nS_L| = 4.299$ ), which had, across all genome-wide SNPs, the 17th highest  $|nS_L|$  score (supplementary table S4a, Supplementary Material online). The immune gene group with the strongest signature of selection was the SCR family, which had five genes with signatures of strong positive selection (each containing SNPs within the top 1% of genome-wide





**FIG. 1.**—Signatures of selection acting on canonical immune genes and associated immune pathways of *Bombus terrestris*. Overview of genes and associated pathways of the bumblebee immune system with genes colored to reflect whether they display one or more of the following characteristics: found in regions of low genetic diversity (ROLD); found in regions of high genetic diversity (ROHD); evidence of extended haplotype homozygosity (EHH); or contain a predicted high-impact SNP. Diagram displays the IMD/JNK, Toll, JAK/STAT, SRRP pathways, and cellular responses of the bumblebee immune response including all functional immune gene groups outlined by Sadd et al. (2015): AMP, antimicrobial peptides; APHAG, autophagy; CASP, caspase; CASPA, caspase A; CAT, catalase; CLIP, CLIP serine protease; CTL, C-type lectin; FREP, fibrinogen like; GAL, galectin; GGBP, gram-negative binding protein/beta-glucan recognition protein; IAP, IAP repeat; IGG, immunoglobulin; IMDPATH, IMD pathway; JAKSTAT, JAK/STAT pathway; LYS, lysozyme; ML, MD-2-related lipid recognition; NIMROD, Nimrod; PGRP, peptidoglycan recognition protein; PPO, prophenoloxidase; PRDX, peroxidase; REL, relish; SCR, scavenger; SOD, superoxide dismutase; SPZ, spätzle; SRPN, serine protease inhibitor; SRRP, small RNA-regulatory pathway; TEP, thioester-containing protein; TOLL, Toll genes; TOLLPATH, Toll pathway.



**FIG. 2.**—Signatures of positive selection vary across canonical immune genes. (A) Scatterplot displaying the top  $|nS_L|$  scores assigned to each canonical immune gene with such genes categorized by gene group (x-axis). Dashed horizontal lines indicate thresholds of positive selection ( $|nS_L| = 2$ ) and those within the top 1% of genome-wide SNPs ( $|nS_L| = 2.51$ ). (B) The correlation of absolute  $|nS_L|$  scores calculated for the present data set consisting of *Bombus terrestris* males sampled in Ireland (“population 1”) and publicly available scores for a British population (“population 2”) generated by Colgan et al. (2022), highlighting conservation in signatures of positive selection in geographically separated populations. Dashed vertical and horizontal lines indicate thresholds of selection ( $|nS_L|$  score  $\geq 2$ ) in each population. Dots above both threshold lines represent canonical immune genes with signatures of recent positive selection found in both populations. The thick dashed line indicates a best fit regression line complete with 95% confidence intervals. Gene groups adapted from Sadd et al. (2015): AMP, antimicrobial peptides; APHAG, autophagy; CASP, Caspase; CASPA, caspase A; CAT, catalase; CLIP, CLIP serine protease; CTL, C-type lectin; FREP, fibrinogen like; GALE, galectin; GNBP, gram-negative binding protein/beta-glucan recognition protein; IAP, IAP repeat; IGG, immunoglobulin; IMDPATH, IMD pathway; JAKSTAT, JAK/STAT pathway; LYS, lysozyme; ML, MD-2-related lipid recognition; NIMROD, Nimrod; PGRP, peptidoglycan recognition protein; PPO, Prophenoloxidase; PRDX, peroxidase; REL, relish; SCR, scavenger; SOD, superoxide dismutase; SPZ, spätzle; SRPN, serine protease inhibitor; SRRP, small RNA regulatory pathway; TEP, thioester-containing protein; TOLL, Toll genes; TOLLPATH, Toll pathway.

$|nS_L|$  scores [ $|nS_L| \geq 2.51$ ]), which was greater than expected by random chance (Fisher’s exact test;  $P < 0.05$ ). Across all immune genes, however, we did not find significant differences (Wilcoxon rank sum test:  $W = 698704$ ,  $P = 0.179$ ) in the mean  $|nS_L|$  scores assigned to immune genes (mean  $|nS_L| = 1.6$ ) compared with the rest of the genome (mean  $|nS_L|$  for nonimmune genes = 1.5). Similarly, with a GO term enrichment analysis for immune genes ranked by  $|nS_L|$  score, we found no significant enrichment of immune-associated GO terms (KS test;  $P > 0.05$ ).

To determine if canonical immune genes are experiencing similar forms of selection in other populations, we compared the  $|nS_L|$  scores assigned to such genes in our data set with those previously published by Colgan et al. (2022), which involved analysis of male *B. terrestris* sampled across the island of Great Britain. We found a

significant positive correlation (Pearson correlation coefficient [ $R$ ] = 0.48,  $P < 0.0001$ ; fig. 2B) in terms of  $|nS_L|$  scores assigned to immune genes in both data sets, highlighting conserved patterns of strong positive selection acting on approximately one-fifth of immune genes ( $n = 34/151$  genes with  $|nS_L| \geq 2$ ) in two geographically separated populations. Despite this overlap, this number of genes was not more than expected by random chance ( $\chi^2_{df=1} = 0.95$ ,  $P = 0.33$ ).

#### Potential Loss of Function SNPs in Immune Genes

As haploid males automatically express the alleles they carry, we investigated the presence and frequency of potential loss of function polymorphic sites within our wild-sampled population, which may highlight nonfunctional or potentially nonessential immune genes. Across all genome-

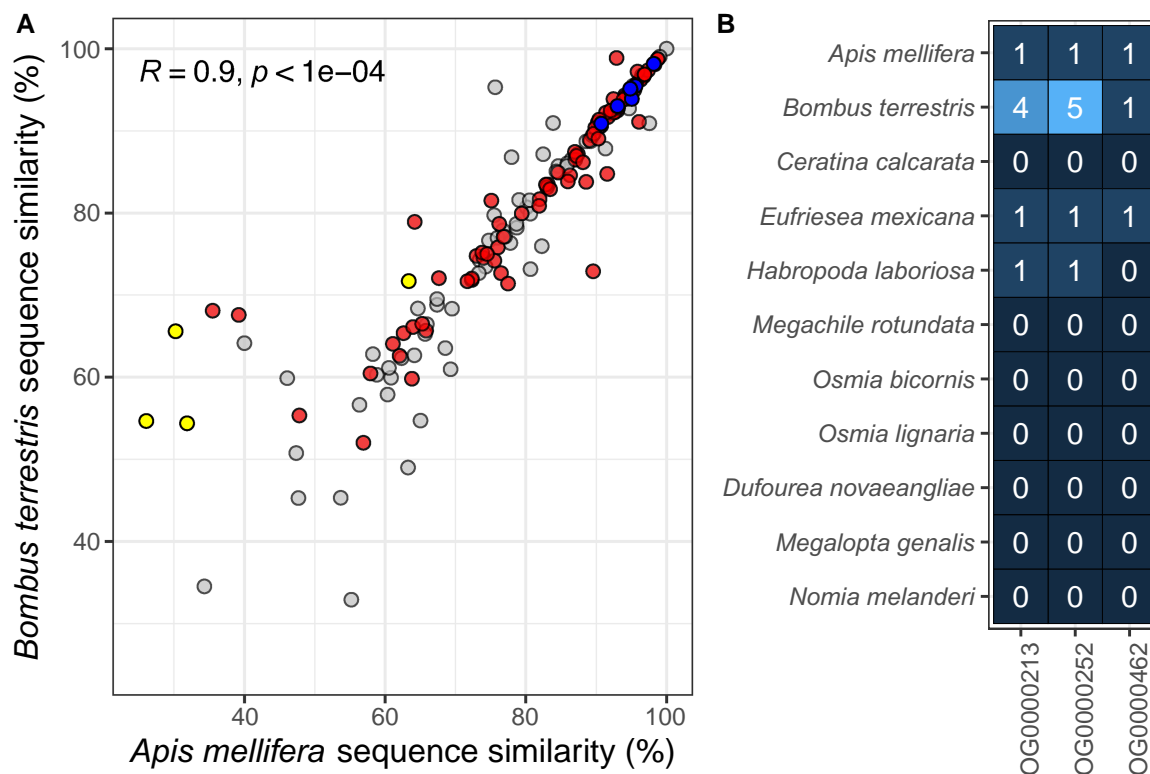
**Table 2**

Predicted Loss of Function or High-Impact Variants in the Genomes of Wild-caught Male Bumblebees

High-impact variant	Total number of SNPs in genome	Total number of SNPs in canonical immune genes
Stop gained	103	3
Splice donor variant & intron variant	81	1
Splice acceptor variant & intron variant	48	1
Stop lost	38	0
Start lost	11	0
Stop lost & splice region variant	3	0
Stop gained & splice region variant	2	0

NOTE.—The total number of high-impact SNPs categorized by snpEff-annotated variant type across the entire genome and among the canonical immune genes from a population of *Bombus terrestris*.

wide polymorphic sites, we identified 286 SNPs, affecting potentially 247 genes, which were annotated as being putatively high impact (table 2) whereby the variant was predicted to impact transcription or produce truncated proteins during translation. The gaining of an early stop codon (“Stop gained”) was the most common predicted high-impact variant, with 103 occurrences detected. We found four canonical immune genes with high-impact SNPs (fig. 3A), including predicted early stop codons in a serine protease LOC100652157 (found in 36% of all sampled males) and the gene *transmembrane protease serine 9* (LOC100652036; found in 24% of all sampled males). We also detected a splice donor variant within an intronic region in the gene *serine proteinase stubble* (LOC100648752; detected in 24% of all males). The final gene, *antichymotrypsin 2* LOC100648602, a member of the serine protease inhibitor gene family, contained a rare splice acceptor variant (present in 6% of sampled males) and an early stop codon variant (present in 18% of sampled



**FIG. 3.**—Comparative genomics of potentially nonessential canonical immune genes. (A) For the four canonical immune genes with predicted high-impact SNPs in wild-caught bumblebees, protein sequence similarity was compared between homologues of *Bombus terrestris* and the closely related honeybee *Apis mellifera*. Each dot represents individual homologues with colored dots indicating genes that are single-copy orthologues (red), those that lack polymorphism in the wild-sampled population (blue) or are genes with predicted high-impact SNPs (yellow). The strength (Pearson correlation coefficient,  $R$ ) and significance ( $P$ -value) of correlation between sequence similarity is provided. (B) Homology-based analysis resolved the four genes of interest into three orthogroups. For two out of three orthogroups, *B. terrestris* had multiple gene copies (designated by the number inside each box), whereas for each orthogroup, the presence of homologues was largely restricted to other members of the Apidae (*Apis mellifera*, *Eufriesea mexicana*, and *Habropoda laboriosa*), with the exception of *Ceratina calcarata*, and was missing in members of other bee families (Megachilidae: *Megachile rotundata*, *Osmia bicornis*, *Osmia lignaria*; and Halictidae: *Dufourea novaeangliae*, *Megalopta genalis*, *Nomia melanderi*).

males). For these four genes, pairwise protein sequence similarity was lower and more variable compared with other canonical immune genes shared with the closely-related honeybee *Apis mellifera* (fig. 3A). A homology-based analysis across a phylogenetically broader set of bee species identified these four genes as largely taxonomically restricted to members of the Apidae (fig. 3B).

We found evidence of 74 putative duplicated regions (supplementary table S5, Supplementary Material online) across the wild-collected bumblebees based on read alignment information and elevated read depth (three standard deviations from the genome-wide median), including evidence of a putative duplicated exon in the detoxification enzyme *catalase* (LOC100651198), which was found in more than half of the sampled males ( $n = 17$ ) and may potentially impact function.

## Discussion

Insect pollinators, such as bumblebees, face increasingly challenging environments, with pathogens among the key proposed contributors to their documented decline. Despite this, we know comparatively little about the types of recent selective forces specifically shaping the immune system in bumblebees. To address this, we used population genomics to identify patterns of selection acting on canonical immune genes in a natural population of the key bumblebee pollinator, *B. terrestris*. We found that within and across functional categories of immune genes, different selective pressures are acting, shaping the diversity of roles and activities that such genes play in the immune response. We found evidence of adaptation with strong signatures of recent selection acting on pathogen recognition receptors, essential for the detection of diverse pathogenic threats, the coordination of the immune response, as well as distinguishing self from nonself. We also found evidence of putative evolutionary constraints acting on bumblebee immunity with certain immune genes lacking polymorphic sites, suggesting that such genes evolve under strong purifying selection. Lastly, we find potential redundancy or nonessential components of the bumblebee immune gene repertoire through identification of putative loss of function SNPs in the genomes of wild-caught haploid males. Our study provides novel insight into the evolution of the genes underlying the hymenopteran immune defense and the capacity and putative constraints of bumblebees to respond to ongoing and emerging pathogenic threats.

### Signatures of Positive Selection Acting on Genes Involved in Pathogen Recognition and Viral Defense

The insect immune system represents the final obstacle to the establishment of pathogen infection and disease onset. Bumblebees, like other insects, face a taxonomically diverse range of pathogenic threats ranging from viruses to

macroparasites (Schmid-Hempel 1998) which likely require different components of the immune system for detection and generation of appropriate responses. The co-evolutionary nature of host–parasite interactions means that both parties can exert selective pressures on each other, which can result in rapid changes at the genomic level (Roux et al. 2014; Shultz and Sackton 2019). In the present study, we found strong signatures of recent positive selection acting on members of the SCRs, a family of cell surface proteins that bind to a variety of ligands, such as pathogen-associated molecular patterns, and play a fundamental role in pathogen recognition (Krieger 1997). The SCR family consists of 17 genes in the *B. terrestris* genome (Barribeau et al. 2015; Sadd et al. 2015). SCR family members have been shown to have elevated gene expression in response to infection by common parasites, such as the trypanosome, *Crithidia bombi* (Riddell et al. 2014) and the nematode *Sphaerularia bombi* (Colgan et al. 2020). Evidence of recent adaptive selection acting on SCRs has been described in other insects, such as the fruit fly *D. melanogaster* (Lazzaro 2005), indicating an important role of SCR genes in innate immunity and highlighting them as potentially conserved targets of positive selection across taxonomically diverse insect groups. The exact role of SCRs in bees remains largely unknown, although studies in *A. mellifera* show increased expression of an SCR protein in bees with a higher resistance to deformed wing virus when compared with individuals susceptible to the virus (Weaver et al. 2021). The importance of recognition proteins is further highlighted by the identification of genes coding for ligands and receptors associated with the Toll signaling pathway also with signatures of recent positive selection (fig. 1). Although we may expect genes involved in intracellular signaling to evolve under purifying selection in order to maintain function, selection acting on genes coding for receptors of the Toll signaling cascade may allow for adaptation to novel or re-emergent threats. Consequently, initiators and ligands of the Toll pathway, such as CLIP serine proteases and spätzle proteins, have been previously reported to have accelerated rates of evolution at the macroevolutionary scale across *Apis* and *Bombus* species (Barribeau et al. 2015) as well as in *Drosophila* species (Jiggins and Kim 2007; Levin and Malik 2017; Wang et al. 2022), highlighting the importance of such genes to contribute to the immune response. Our results support previous findings of positive selection acting on initiators and receptors of these pathways suggestive that such genes may help bumblebees cope with pathogens in their natural landscape.

A particular group of pathogens that have received attention of late are viruses, which can cause extensive morbidity and increased mortality within at-risk pollinator populations (Manley et al. 2015). More specifically, viruses have raised concerns from an environmental perspective



given that shared resources between commercial and wild pollinators can lead to increased prevalence and exposure in wild populations (Fürst et al. 2014). Within regions of potential hard sweeps, we find genes involved in antiviral defense, such as three ATP-dependent RNA helicases and Dcr-1, as well as enrichment of the GO term “helicase activity” suggestive of strong selection acting on these genes. Although our analysis did not involve a screen for viruses, accelerated rates of evolution have been described for classic antiviral RNAi pathway genes (Dcr2, R2D2, and Ago2) in wild populations of several fruit fly species (Obbard et al. 2006). However, given the essential functions such genes play in cellular defense, RNA metabolism and degradation, purifying selection may also contribute to the low nucleotide diversity found in these genes (Jin and Xie 2007; Xing et al. 2019). Roles in bee immunity have been investigated at the molecular level for RNA helicase Dbp2 and Dcr-1, which change in expression in response to viral and other parasitic infections in *A. mellifera* (Huang et al. 2016; Pizzorno et al. 2021). Our data also show the likely importance of several types of recognition genes in *B. terrestris* and that they may be facing similar selective pressures from pathogens in their environment.

In comparison with regions of low diversity, we found less canonical immune genes in regions of significantly elevated nucleotide diversity. The relatively low number detected ( $n = 5$ ; figs. 1 and 3) in such regions is surprising as many immune genes are predicted to evolve under balancing selection, whereby numerous alleles are maintained at higher than expected frequencies in a population and can occur under different scenarios, such as heterozygote advantage and negative frequency selection (Unckless and Lazzaro 2016; Chapman et al. 2019). However, balancing selection can be difficult to detect and distinguish from other types of selection that produce similar genomic patterns, for example, incomplete sweeps caused by positive selection (Fijarczyk and Babik 2015). Similarly, some immune genes also overlap with previously documented crossing over events in *B. terrestris* (Liu et al. 2017), suggesting that recombination also contributes to or may be the primary cause of the elevated genetic diversity that we observe. Although we detected signatures of recent positive selection acting on pathogen recognition genes, such genes are also often highly polymorphic as they are under selection to detect existing, reemerging and novel variants of co-evolving pathogens (Morlais et al. 2004; Sommer 2005; Early et al. 2017). Similarly, not only do recognition receptors need to differentiate between self and nonself to prevent attacks on an individual's own tissue, but they also need to identify beneficial or symbiotic microbes from harmful ones (Pan et al. 2018). Furthermore, we find certain autophagy genes in these regions, which may also be polymorphic to allow for the targeting of diverse and foreign compounds intended for degradation

(Ying and Feng 2019). Future studies on determining the extent of balancing selection acting on immune genes will benefit from intersexual (haploid vs. diploid), cross-population, and cross-species comparisons to identify conserved regions of elevated diversity, such as trans-species polymorphism, consistent with genomic signatures of long-standing balancing selection.

### Evolutionary Constraints on Immune Gene Evolution

Comparative genomic studies on genes of the immune system have routinely identified such genes to evolve under positive selection to respond to evolutionary changes in co-evolving pathogens. Although in the present study, we identify genes with signatures of positive selection, suggestive of recent adaptive responses, immune genes may also face other evolutionary pressures that place constraints on the adaptation of functional diversity. Purifying selection is a powerful force that acts to preserve gene function through the removal of deleterious alleles from a population (Nielsen et al. 2005). This form of selection is predicted to act particularly strongly on bumblebee males, given their haploid nature. In our analysis, we find several canonical immune genes that lack genetic variation and may be the products of purifying selection. This is supported by our finding of a number of these genes exhibiting signatures of episodic purifying selection (supplementary table S6, Supplementary Material online). Our sampling of wild-caught adult males across the island of Ireland suggests that such low diversity in our contemporary population does not hinder development or survival in the natural landscape. Such low diversity may also be the result of low recombination, which is particularly true for two genes, *FAS-associated factor 1* and *ubiquitin-like-conjugating enzyme ATG3*, found on chromosome 1, which were previously reported by Colgan et al. (2022) as located in an extended region of extremely low genetic diversity. The location of genes in such regions may place evolutionary constraints on the ability of selection to act upon these genes to respond to ongoing environmental challenges highlighting potential limitations on the ability of immune components to adapt to future pathogenic threats.

### What is Driving Selection in *Bombus* Canonical Immune Genes?

Given the strong selective pressures placed by pathogens on host immune systems, recent selection signatures in many of the canonical immune genes may be in response to increasing pathogen threats and their detrimental effects on bumblebee condition and fitness (Otti and Schmid-Hempel 2007; Yourth et al. 2008; Colgan et al. 2020). Indeed, these threats are becoming increasingly serious and, as mentioned previously, considered a key factor contributing to pollinator decline (Goulson 2010). In

particular, the increasing use of commercial pollinators has been shown to cause pathogen spillover to wild populations which may have been previously naive to particular pathogens and/or specific strains (Colla et al. 2006; Otterstatter and Thomson 2008). Although adaptation to pathogens may be the most parsimonious reason for our identified patterns, such selection signatures may also be the result of other processes or targets of selection. Many canonical immune genes have pleiotropic effects with their products being expressed at different points in the insect life cycle, contributing to essential roles in development, reproduction, and metabolism such as Toll receptors, which play central roles in both embryonic pattern formation and immunity (Zou et al. 2006; Valanne et al. 2011). Such genes may therefore be selected due to a function unrelated to immunity. Similarly, immune genes may not be the direct targets of selection but hitchhike through physical linkage with neighboring sites under selection (Nielsen et al. 2005). For several canonical immune genes, we find strong signatures of extended haplotype homozygosity, which is predicted to break down with physical distance from the target of selection, especially within bees, which have high recombination rates (Beye et al. 2006), suggesting that these genes are at least located in regions exposed to recent selection. Similarly, given the documented role of many selected genes in immune function (e.g., SCRs; Zhang et al. 2021), we may predict that such genes are evolving in response to pathogen-driven selection but further experimental work is required to determine this link.

#### Potential Redundancy in Certain Immune Genes of a Wild-caught *B. terrestris* Population

Despite the importance of the immune system, studies in other taxa have suggested that certain components of the immune system may be nonessential and evolve under relaxed selection (Nish and Medzhitov 2011). In our study, we found potential loss of function SNPs in three CLIP serine proteases and one serine protease inhibitor (serpin) present in the genomes of wild-caught males. Similar to canonical immune genes that lack polymorphism, the presence of loss of function alleles in wild-caught haploid adults would suggest that these alleles may not negatively impact development or survival in carriers. CLIP serine proteases perform diverse immune roles, including the triggering of immune pathways, such as the Toll pathway and the prophenoloxidase cascade (Kanost and Gorman 2008; An et al. 2013), whereas, in contrast, serpins act as regulators of such protease cascades (Shakeel et al. 2019). Although bees, in general, are documented to have a depleted canonical immune gene set compared with other insects (Evans et al. 2006), *B. terrestris* has an expanded gene set of both serine proteases and serpins compared with *A. mellifera* with many such genes evolving through tandem gene

duplications (Barribeau et al. 2015). Duplication events are common but are usually quickly purged from the genome. However, retained copies may also diverge through adaptive or neutral processes (Hastings et al. 2009). In the latter scenario, random accumulation of mutations through relaxed selection may occur, disrupting function, and potentially leading to pseudogenization. Although we found no evidence of copy number variation for such genes in our population data set nor did we specifically test for signatures of relaxed episodic selection, we do find evidence of greater divergence in terms of predicted protein sequence similarity for such genes compared with other immune genes shared with *A. mellifera* (fig. 3). This result in combination with the presence of putative loss of function alleles, suggests that the additional copies may be evolving under relaxed selection, resulting in the genes becoming functionally divergent and potentially nonessential from an immune perspective, which may explain why purifying selection has not removed these potential loss of function SNPs from our wild-caught population. However, functional studies on this population of bumblebees are needed to evaluate if and how potential loss of function SNPs impact protein function and to what extent. Furthermore, it should be explored whether possible loss of function SNPs in immune genes have divergent consequences on fitness across bumblebee sexes and castes, especially considering there are observed differences in immunity between males and females (Baer and Schmid-Hempel 2006).

In the current analysis, we investigated genomic patterns of selection through nucleotide diversity and extended haplotype homozygosity, which can provide insight into targets of recent positive and purifying selection. The use of haploid males for population genomic analyses, which has been used in previous studies (Colgan et al. 2022), likely provides higher resolution in patterns of selection due to the lack of requirement for phasing. However, to fully understand how much genetic variation is maintained in bumblebee populations, and how selection is shaping their genome, we should also examine variation in females, which given their diploid nature will provide additional insights into the genetic variation found within individuals in wild populations. Similarly, although bees have been characterized as having a depauperate canonical immune gene set (Evans et al. 2006; Barribeau et al. 2015), more recent studies suggest additional genes function in the hymenopteran immune system (e.g., Doublet et al. 2017; Möllmann and Colgan 2022); such genes were not examined in our present study given our restriction to the canonical immune gene set and associated gene groups detailed by Sadd et al. (2015) but warrant investigation. Furthermore, our analysis was restricted to canonical immune genes located on linkage groups, representative of chromosomes, accounting for ~90% of the original genes described by Sadd et al. (2015). The generation of new

bumblebee assemblies and associated annotations, such as those generated by the Darwin Tree of Life (Darwin Tree of Life Project Consortium 2022) or European Reference Genome Atlas (Formenti et al. 2022), will allow for the assessment of the entire expanded immune gene set in future studies.

## Conclusions

Advances in population genomic approaches are providing us with valuable new insights into many aspects of population biology and evolution, including the adaptive abilities of organisms to respond to their natural environments. Genomics is a powerful tool that can assess a huge amount of genetic markers and yields the resolution required to understand how the fundamental process of natural selection works in wild populations of ecologically important species. Evaluating how various kinds of selective pressures are acting in a population will help biologists decipher the evolutionary history, local adaptation, and resilience to threats, such as climate change and habitat fragmentation, of important keystone species. In this study, we used population genomics to investigate how evolution is shaping the immune system of a key bumblebee species. Bumblebees, like many other insects, are experiencing dramatic global declines driven by the intensification of agriculture, habitat loss, use of pesticides, climate change, and pathogen spread, with the latter threat placing the immune systems of their hosts under strong selective pressures. However, we know little about the kinds of selection that the immune systems of wild pollinators are experiencing. Considering this, we aimed to shed light on the selective processes acting on canonical immune genes in a wild-caught population of the earth bumblebee, *B. terrestris* and expect that our findings will inform understanding of pollinator–pathogen dynamics, an essential feat if we are to move toward conserving these vital species from potentially one of the contributing causes of their decline.

Overall, we found that canonical genes of a wild-caught bumblebee immune system are experiencing different kinds of selective pressures, which reflects the functional complexity of these important physiological systems and indicates how certain genes may be responding to changing environmental challenges. We found evidence that certain elements of the bumblebee immune gene repertoire are undergoing adaptive change that may increase this population's ability to defend against parasites and pathogens that they are exposed to. Furthermore, we found other canonical immune genes that appear to be under constraints, presumably reflecting their roles in essential cellular, developmental, and physiological processes, and, additionally, genes that may be redundant in bumblebee immunity. To conclude, we show the utility of population genomics to inform us about the adaptive abilities of a species in a natural

environment, which may prove a highly valuable tool to identify pollinator populations vulnerable or robust to increasing pathogenic threats. We anticipate our findings will also motivate future characterizations of the immune system of wild bumblebees as well as assessments of the genetic variation found in wild bumblebees and other insect pollinators. Additionally, we believe these approaches could be used in the future to evaluate the genomes of wild and commercially imported *B. terrestris* bumblebees and their capabilities to deal with immune threats. This information would inform changes to the management and policy surrounding bumblebee imports, especially in regions where commercial imports derive from non-native stocks. Overall, this study helps cultivate a deeper understanding of the dynamics between bumblebees and their pathogens, an area that requires thorough investigation if we are to fully comprehend, and potentially mitigate, the negative impacts that pathogen spread is thought to have on wild pollinating insect populations.

## Materials and Methods

### Sample Collections

To assess signatures of selection acting on canonical immune genes, we sampled 33 *B. terrestris* males foraging in natural settings across 27 sites from across the island of Ireland between June and September 2018 (supplementary table S1, Supplementary Material online). Field-based identifications of *B. terrestris* males were based on the standard morphology of this species (Falk 2019) and the lack of an aculeus, which is found only in females. All collected samples were frozen on the same day and stored at  $-20^{\circ}\text{C}$  until DNA extractions.

### DNA Extraction, Library Preparation and Sequencing

Genomic DNA was extracted from the heads of each bee using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA), following the protocol for animal tissues. Some modifications to the protocol were applied and included the following: 40  $\mu\text{l}$  of Qiagen Proteinase K was added to each sample and samples were vortexed prior to and every hour during a 3-h incubation step at  $56^{\circ}\text{C}$ . DNA samples were eluted in 150  $\mu\text{l}$  Qiagen Buffer AE and assessed for quality and quantity using agarose gel electrophoresis and a NanoDrop™ Spectrophotometer (Thermo Scientific). Extracted DNA samples were then sent for further quality assessment, genomic library preparation, and subsequent sequencing by a commercial company (Novogene, Cambridge, UK). Polymerase chain reaction-free genomic libraries were produced using the NEBNext Ultra I DNA library kit and paired-end (PE) sequencing (150 bp) was performed on an Illumina NovaSeq6000 sequencing platform. Sequencing generated an average of 13.4 million PE reads

per individual (supplementary table S1, Supplementary Material online), ranging from 10.6 to 19.8 million PE reads and a standard deviation of 1.64 million PE reads. Based on an estimated genome size of 248 Mb, the mean predicted raw coverage of each sample was 16.2× (min: 12.76×; max: 23.95×; supplementary table S1, Supplementary Material online).

### Quality Assessment, Filtering, and Data Alignment

Using FastQC (v.0.11.9; Andrews 2010), we assessed raw reads for overall base quality. Raw sequencing quality was high as the mean Phred score across all samples was >30 per base position. Using fastp (v.0.22.0; Chen et al. 2018), we removed sequences with over 10% ambiguous ("N") bases and those with >40% of bases with Phred scores <20. In addition, we used fastp to identify and remove Illumina adapter sequences. Post-filtering, we retained sequences that were ≥50 bases in length.

### Variant Calling and Filtering of Single Nucleotide Polymorphisms

Our variant calling and filtering approach followed that of Colgan et al. (2022). In brief, to identify polymorphic sites within our population, we aligned filtered reads against the *B. terrestris* reference genome assembly (Bter\_1.0, GCF\_000214255.1) using bowtie2 (v.2.4.4; Langmead and Salzberg 2012). Using these alignment files, variants were called using FreeBayes (v.1.3.5; Garrison and Marth 2012) with the following parameters: `-ploidy 2 -report-genotype-likelihood-max -use-mapping-quality -genotype-qualities -use-best-n-alleles 4 -haplotype-length 0 -min-base-quality 3 -min-mapping-quality 1 -min-alternate-frac 0.25 -min-coverage 1`. We analyzed the samples as potential diploids to allow for the identification of low-complexity sites, putative copy number variants, and misalignments, which may be called as heterozygous and can be removed from downstream analyses. We used VCFtools (v.0.1.17; Danecek et al. 2011) to filter out low-quality variant sites with a quality (QUAL) score of <20, variants in low complexity, repetitive regions with a max mean depth of >100 bp, insertions and deletions ("indels") and sites with missing data (`-max-missing 1.0`). As the sequenced bees were haploid, and therefore hemizygous, any sites called heterozygous were removed. To reduce the complexity of the data sets, sites with >2 alleles (i.e., multiallelic sites) and rare alleles occurring only once across all 33 individuals were removed using VCFtools. Lastly, we subsetted SNPs found only on the 18 linkage groups (representative of chromosomes) of the *B. terrestris* Bter\_1.0 reference genome assembly. We annotated all retained high-quality SNPs with variant-type data using snpEff (v.5.0e; Cingolani et al. 2012).

### Assessment of Population Structure

To determine evidence of population substructure in our data set, we use three approaches. First, we performed a principal component analysis using a pruned data set from called genotypes using the R package SNPRelate (v.1.20.1; Zheng 2013). Second, using this pruned data set, we performed an identity-by-state analysis using SNPRelate to determine relatedness among individuals. Third, we used a cluster-based approach to estimate co-ancestry among sampled individuals using ADMIXTURE (v.1.3.0; Alexander et al. 2009) with  $K = 1-20$  and cross-validation (`-cv`) to identify the most probable  $K$  populations in our data set. All three analyses suggested that our sampled bumblebees originate from a panmictic, single population, displaying only weak substructure (supplementary fig. S1, Supplementary Material online).

### Sliding Windows-based Estimation of Nucleotide Diversity and Tajima's D

We used PopGenome (v.2.7.5; Pfeifer et al. 2014) on the high-quality SNPs to calculate mean nucleotide diversity ( $\pi$ ) and Tajima's D using 100 kb sliding windows to identify genomic regions with significantly elevated or reduced measures of diversity in comparison with the rest of the genome. As the reference genome assembly contains regions of ambiguous bases, which can falsely present as regions of low diversity, we first identified and removed such regions. For this approach, we used bedtools (v.2.30.0; Quinlan 2014) to generate window-specific FASTA files containing the nucleotide sequences present in each window. We then calculated the percentage of ambiguous bases in each window using seqtk comp (v1.3-r106; available from: <https://github.com/lh3/seqtk>). We removed windows that contained over 10% of ambiguous bases ( $n = 417$ ), and therefore, represent likely gaps in the genome assembly. Post-filtering, we estimated nucleotide diversity and Tajima's D for each of the remaining windows ( $n = 3,894$ ). We then performed a z-score transformation on nucleotide diversity estimates and calculated corresponding  $P$ -values to determine regions of significantly elevated or reduced nucleotide diversity, which may represent signatures of positive and balancing selection, respectively. Lastly, we extracted the gene coordinates from the *B. terrestris* gene feature file (gff3) obtained from Ensembl Metazoa (Howe et al. 2021) for the bumblebee canonical immune genes (Sadd et al. 2015) and used bedtools intersect (v.2.30.0; Quinlan 2014) to identify the presence of such genes in windows of elevated or reduced nucleotide diversity.

As a complementary measure, we performed a gene-level analysis using PopGenome. We calculated the number of segregating sites and length-corrected nucleotide diversity for each gene in the *B. terrestris* genome. Based on these measures, we extracted canonical immune genes with zero diversity (i.e., where there was no evidence of



SNPs). We further assessed genes that lacked diversity by calculating the number of reads covering each gene per sample using samtools depth (v.1.10; Li et al. 2009) finding no significant difference in read coverage for each gene of interest compared with the genome-wide mean ( $t$ -test;  $P > 0.05$ ). Lastly, for canonical immune genes that lacked genetic variation, we examined individual alignment (BAM) files using the Integrated Genome Viewer (IGV; v.2.13.2; Robinson et al. 2011) to confirm the lack of called SNPs.

### Extended Haplotype Homozygosity Tests of Recent Positive Selection

To complement the nucleotide diversity analyses, we performed a haplotype-based analysis to identify regions of extended haplotype homozygosity (EHH), which may represent products of “hard” or “soft” selective sweeps. We used selscan (v.2.0; Szpiech 2021) to calculate  $nS_L$  (number of segregating sites by length) scores for each polymorphic site in our data set.  $nS_L$  is calculated using the number of segregating sites and the length of regions of EHH (Ferrer-Admetlla et al. 2014). We used selscan “norm” under default settings to normalize our  $nS_L$  results against the background genome. We calculated absolute  $nS_L$  ( $|nS_L|$ ) scores to explore patterns attributed to positive selection in line with approaches used in other studies (Colgan et al. 2022). For each gene on the *B. terrestris* linkage groups, we identified the SNP with the highest  $|nS_L|$  score. We then searched for canonical immune genes that had an  $|nS_L|$  score  $\geq 2$ , a threshold for positive selection (Ferrer-Admetlla et al. 2014).

### Copy Number Variation Analysis of Canonical Immune Genes

For the purposes of identifying putative duplicated or deleted regions of the genome, we followed an approach developed by Colgan et al. (2022). We first used lumpySV (v.0.2.13; Layer et al. 2014), a software tool that uses read pair information to determine evidence for the presence of copy number variation. Under the authors’ recommendations, we first realigned our filtered reads against the reference genome assembly using bwa (v.0.7.17-r1188; Li and Durbin 2009). From the resultant alignment file, we extracted discordant and split-paired reads using samtools (v.1.10; Li et al. 2009). We next extracted genomic coordinates for each putative duplicated site and retained only sites that were either  $>500$  bp or  $<100,000$  bp in length. Neighboring sites, which overlapped by at least 1 bp, were sorted and merged using bedtools (v.2.28.0; Quinlan 2014). For each putative duplicated site, we next calculated the percentage of ambiguous bases within each and removed sites where the percentage of ambiguous bases was above 10%. To determine if duplicated sites

had elevated read depth compared with other parts of the genome, we used bedtools (intersectBed) and samtools (depth) to calculate the number of reads aligned to each base within each putative duplicated region. We then calculated the median read depth for each genomic region within each individual using R (v.3.6.3). After this, we compared the median read depth per site against the genome-wide median read depth retaining sites whereby read depth was more than three standard deviations from the genome-wide median. For putative deleted sites, we examined the overlap of putative sites with canonical immune genes. We generated individual BAM files per site with bedtools intersectBed and calculated read depth per individual bee. Using these calculations, we examined loci with zero read depth, suggestive of a deletion in that individual. We only retained deletions found in at least two individuals. To identify copy number variation in immune genes, we parsed both the list of putative duplicated and deleted sites for canonical immune genes.

### Episodic Selection Analysis for Canonical Immune Genes

To identify macroevolutionary signatures of long-term (or episodic) selection, we performed a comparative genomic analysis using single-copy orthologues of all genes present in the reference genome assembly of *B. terrestris*, as well as ten other bee species from across three different bee families (Apidae: *A. mellifera*, *Habropoda laboriosa*, *Ceratina calcarata*, *Eufriesea mexicana*; Megachilidae: *Megachile rotundata*, *Osmia bicornis*, *Osmia lignaria*; and Halictidae: *Dufourea novaeangliae*, *Megalopta genalis*, *Nomia melanoderi*). We identified single-copy orthologues using OrthoFinder (v.2.5.2; Emms and Kelly 2019) using the longest predicted protein for each protein-coding gene present in each of our 11 bee species of interest. The branch-site models aBSREL (Smith et al. 2015) and BUSTED (Murrell et al. 2015), implemented in HyPhy (v.2.5.39; Pond and Muse 2005), were run to examine sites with signatures of episodic positive selection among the canonical immune genes. The sequence alignments and the phylogenetic trees required were inferred using MUSCLE (v.3.8.1551; Edgar 2004) and fastTree (v.2.1.11; Price et al. 2010), respectively. In addition, a modification of BUSTED (conservation model) was run to examine the presence of purifying or negative selection acting on canonical immune genes.

### Gene Ontology Term Enrichment Analysis

We performed GO term enrichment analysis on canonical immune genes with signatures of recent selection using topGO (v.2.38.1; Alexa and Rahnenführer 2009) using scripts previously developed by Colgan et al. (2019). Given the low-resolution GO terms assigned to *B. terrestris* genes, we assigned GO terms from *D. melanogaster* obtained from Ensembl Metazoa Biomart (Kinsella et al.

2011) to their bumblebee homologues. We performed a Fisher's exact test ( $P < 0.05$ ; node size = 50) for immune genes in genomic regions of significantly low nucleotide diversity. In this analysis, we used the "weight01" algorithm for topGO and ran the analysis separately for each GO category ("biological process," BP; "molecular function," MF; and "cellular component," CC).

## Supplementary material

Supplementary data are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

This work was supported by the Department of Agriculture, Food and the Marine under the Genetic Resources Grant Aid Scheme (Grant number 8/GR/12 GRGAS), the Irish Research Council (GOIPG/2018/2106: PhD scholarship awarded to S.J.L.), and Deutsche Bundesstiftung Umwelt (20021/743: PhD scholarship awarded to J.S.M.). The authors thank Randal Counihan for assistance in data processing and express our gratitude to all those who collected bumblebee samples for this project including Nadine Douglas, Dr Felipe Guapo de Melo, Maeve McCann, Dr Laura Russo, Yung-Yi Lin, and Anne Keane. Computational support and resources were provided by the Irish Centre for High-End Computing.

## Data availability

Raw sequence files are available from the NCBI Sequence Read Archive (BioProject ID: PRJNA870415). Scripts used for the current analysis are publicly available and hosted at [https://github.com/Joscolgan/bter\\_immune\\_gene\\_evolution](https://github.com/Joscolgan/bter_immune_gene_evolution).

## Literature Cited

- Aizen MA, et al. 2009. How much does agriculture depend on pollinators? Lessons from long-term trends in crop production. *Ann Bot.* 103(9):1579–1588.
- Alexa A, Rahnenführer J. 2009. Gene set enrichment analysis with topGO. *Bioconductor Improv.* 27:1–26.
- Alexander DH, et al. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.
- Ali Mohammadi Kojour M, et al. 2020. An overview of insect innate immunity. *Entomol Res.* 50(6):282–291.
- An C, et al. 2013. Serine protease MP2 activates prophenoloxidase in the melanization immune response of *Drosophila melanogaster*. *PLoS One* 8(11):e79533.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed on 24th March 2023.
- Baer B, Schmid-Hempel P. 2006. Phenotypic variation in male and worker encapsulation response in the bumblebee *Bombus terrestris*. *Ecol Entomol.* 31(6):591–596.
- Balloux F, van Dorp L. 2017. Q&A: what are pathogens, and what have they done to and for us?. *BMC Biol.* 15(1):91.
- Barreiro LB, Quintana-Murci L. 2010. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet.* 11(1):17–30.
- Barribeau SM, et al. 2015. A depauperate immune repertoire precedes evolution of sociality in bees. *Genome Biol.* 16(1):1–21.
- Beye M, et al. 2006. Exceptionally high levels of recombination across the honey bee genome. *Genome Res.* 16(11):1339–1344.
- Cameron SA, et al. 2011. Patterns of widespread decline in North American bumble bees. *Proc Natl Acad Sci USA.* 108(2):662–667.
- Chapman JR, et al. 2019. Balancing selection drives the maintenance of genetic variation in *Drosophila* antimicrobial peptides. *Genome Biol Evol.* 11(9):2691–2701.
- Chen S, et al. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34(17):i884–i890.
- Cingolani P, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6(2):80–92.
- Colgan TJ, et al. 2011. Polyphenism in social insects: insights from a transcriptome-wide analysis of gene expression in the life stages of the key pollinator, *Bombus terrestris*. *BMC Genomics.* 12(1):1–20.
- Colgan TJ, et al. 2019. Caste- and pesticide-specific effects of neonicotinoid pesticide exposure on gene expression in bumblebees. *Mol Ecol.* 28(8):1964–1974.
- Colgan TJ, et al. 2020. Infection by the castrating parasitic nematode *Sphaerularia bombi* changes gene expression in *Bombus terrestris* bumblebee queens. *Insect Mol Biol.* 29(2):170–182.
- Colgan TJ, et al. 2022. Genomic signatures of recent adaptation in a wild bumblebee. *Mol Biol Evol.* 39(2):msab366.
- Colla SR, et al. 2006. Plight of the bumble bee: pathogen spillover from commercial to wild populations. *Biol Conserv.* 129(4):461–467.
- Cook JM, Crozier RH. 1995. Sex determination and population biology in the Hymenoptera. *Trends Ecol Evol (Amst).* 10(7):281–286.
- Cooper D, Eleftherianos I. 2017. Memory and specificity in the insect immune system: current perspectives and future challenges. *Front Immunol.* 8:539.
- Danecek P, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.
- Darwin Tree of Life Project Consortium. 2022. Sequence locally, think globally: the darwin tree of life project. *Proc Natl Acad Sci USA.* 119(4):e2115642118.
- Doublet V, et al. 2017. Unity in defence: honeybee workers exhibit conserved molecular responses to diverse pathogens. *BMC Genomics.* 18(1):1–17.
- Early AM, et al. 2017. Survey of global genetic diversity within the *Drosophila* immune system. *Genetics* 205(1):353–366.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792–1797.
- Emms DM, Kelly S. 2019. Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20(1):1–14.
- Evans J, et al. 2006. Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect Mol Biol.* 15(5):645–656.
- Falk S. 2019. Field guide to the bees of Great Britain and Ireland. Oxford (UK): British Wildlife Publishing.
- Ferrer-Admetlla A, et al. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol Evol.* 31(5):1275–1291.
- Fijarczyk A, Babik W. 2015. Detecting balancing selection in genomes: limits and prospects. *Mol Ecol.* 24(14):3529–3545.
- Fitzpatrick Ú, et al. 2007. Rarity and decline in bumblebees—a test of causes and correlates in the Irish fauna. *Biol Conserv.* 136(2):185–194.

- Formenti G, et al. 2022. The era of reference genomes in conservation genomics. *Trends Ecol Evol (Amst)*. 37(3):197–202.
- Fürst MA, et al. 2014. Disease associations between honeybees and bumblebees as a threat to wild pollinators. *Nature* 506(7488): 364–366.
- Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv 1207:3907 [Preprint].
- Goulson D. 2010. Impacts of non-native bumblebees in Western Europe and North America. *Appl Entomol Zool (Jpn)*. 45(1): 7–12.
- Graystock P, et al. 2013. The Trojan hives: pollinator pathogens, imported and distributed in bumblebee colonies. *J Appl Ecol*. 50(5): 1207–1215.
- Harpur BA, Rehan SM. 2021. Connecting social polymorphism to single nucleotide polymorphism: population genomics of the small carpenter bee, *Ceratina australensis*. *Biol J Linn Soc*. 132(4): 945–954.
- Hartl DL, Brown SW. 1970. The origin of male haploid genetic systems and their expected sex ratio. *Theor Popul Biol*. 1(2):165–190.
- Hastings PJ, et al. 2009. Mechanisms of change in gene copy number. *Nat Rev Genetics*. 10(8):551–564.
- Heraghty SD, et al. 2022. Whole genome sequencing reveals the structure of environment-associated divergence in a broadly distributed montane bumble bee, *Bombus vancoverensis*. *Insect Syst Divers*. 6(5):5.
- Howe KL, et al. 2021. Ensemble 2021. *Nucleic Acids Res*. 49(D1): D884–D891.
- Huang Q, et al. 2016. Host-parasite interactions and purifying selection in a microsporidian parasite of honey bees. *PLoS One* 11(2): e0147549.
- Jiggins F, Kim K. 2007. A screen for immunity genes evolving under positive selection in *Drosophila*. *J Evol Biol*. 20(3): 965–970.
- Jin Z, Xie T. 2007. Dcr-1 maintains *Drosophila* ovarian stem cells. *Curr Biol*. 17(6):539–544.
- Joseph SB, Kirkpatrick M. 2004. Haploid selection in animals. *Trends Ecol Evol (Amst)*. 19(11):592–597.
- Kanost MR, Gorman MJ. 2008. Phenoloxidases in insect immunity. In: Beckage N, editor. *Insect immunology*. San Diego (CA): Academic Press. p. 69–96.
- Kardum Hjort C, et al. 2022. Genomic divergence and a lack of recent introgression between commercial and wild bumblebees (*Bombus terrestris*). *Evol Appl*. 15(3):365–382.
- Kent CF, et al. 2018. Conservation genomics of the declining North American bumblebee *Bombus terricola* reveals inbreeding and selection on immune genes. *Front Genet*. 9:316.
- Kinsella RJ, et al. 2011. Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database* 2011:bar030.
- Kosior A, et al. 2007. The decline of the bumble bees and cuckoo bees (Hymenoptera: Apidae: Bombini) of western and central Europe. *Oryx* 41(1):79–88.
- Krieger M. 1997. The other side of scavenger receptors: pattern recognition for host defense. *Curr Opin Lipidol*. 8(5):275–280.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9(4):357–359.
- Layer RM, et al. 2014. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 15(6):1–19.
- Lazzaro BP. 2005. Elevated polymorphism and divergence in the class C scavenger receptors of *Drosophila melanogaster* and *D. simulans*. *Genetics* 169(4):2023–2034.
- Levin TC, Malik HS. 2017. Rapidly evolving Toll-3/4 genes encode male-specific toll-like receptors in *Drosophila*. *Mol Biol Evol*. 34(9):2307–2323.
- Li H, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Liu H, et al. 2017. Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. *Mol Biol Evol*. 34(1):119–130.
- Manley R, et al. 2015. Emerging viral disease risk to pollinating insects: ecological, evolutionary and anthropogenic factors. *J Appl Ecol*. 52(2):331–340.
- Mavrouli MD, et al. 2005. MAP Kinases mediate phagocytosis and melanization via prophenoloxidase activation in medfly hemocytes. *Biochim Biophys Acta*. 1744(2):145–156.
- Möllmann JS, Colgan TJ. 2022. Genomic architecture and sexually dimorphic expression underlying immunity in the red mason bee, *Osmia bicornis*. *Insect Mol Biol*. 31(6):686–700.
- Morlais I, et al. 2004. Intraspecific nucleotide variation in *Anopheles gambiae*: new insights into the biology of malaria vectors. *Am J Trop Med Hyg*. 71(6):795–802.
- Murray TE, et al. 2013. Pathogen prevalence in commercially reared bumble bees and evidence of spillover in conspecific populations. *Biol Conserv*. 159:269–276.
- Murrell B, et al. 2015. Gene-wide identification of episodic selection. *Mol Biol Evol*. 32(5):1365–1371.
- Nielsen R, et al. 2005. Genomic scans for selective sweeps using SNP data. *Genome Res*. 15(11):1566–1575.
- Nish S, Medzhitov R. 2011. Host defense pathways: role of redundancy and compensation in infectious disease phenotypes. *Immunity* 34(5):629–636.
- Nyholm SV, Graf J. 2012. Knowing your friends: invertebrate innate immunity fosters beneficial bacterial symbioses. *Nat Rev Microbiol*. 10(12):815–827.
- Obbard DJ, et al. 2006. Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr Biol*. 16(6):580–585.
- Otterstatter MC, Thomson JD. 2008. Does pathogen spillover from commercially reared bumble bees threaten wild pollinators? *PLoS One* 3(7):e2771.
- Otti O, Schmid-Hempel P. 2007. *Nosema bombi*: a pollinator parasite with detrimental fitness effects. *J Invertebr Pathol*. 96(2):118–124.
- Palmer WH, Hadfield JD, Obbard DJ. 2018. RNA-interference pathways display high rates of adaptive protein evolution in multiple invertebrates. *Genetics* 208(4):1585–1599.
- Pan X, et al. 2018. The bacterium *Wolbachia* exploits host innate immunity to establish a symbiotic relationship with the dengue vector mosquito *Aedes aegypti*. *ISME J*. 12(1):277–288.
- Pfeifer B, et al. 2014. Popgenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol*. 31(7):1929–1936.
- Pizzorno MC, et al. 2021. Transcriptomic responses of the honey bee brain to infection with deformed wing virus. *Viruses* 13(2):287.
- Pond SLK, Muse SV. 2005. Hyphy: hypothesis testing using phylogenies. In: *Statistical methods in molecular evolution*. New York: Springer. p. 125–181.
- Price MN, et al. 2010. Fasttree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):e9490.
- Quinlan AR. 2014. BEDTools: the Swiss-army tool for genome feature analysis. *Curr Protoc Bioinformatics*. 47(1):11–12.
- Riddell CE, et al. 2014. Differential gene expression and alternative splicing in insect immune specificity. *BMC Genomics*. 15(1):1–15.
- Robinson JT, et al. 2011. Integrative genomics viewer. *Nat Biotechnol*. 29(1):24–26.
- Roh KB, et al. 2009. Proteolytic cascade for the activation of the insect toll pathway induced by the fungal cell wall component. *J Biol Chem*. 284(29):19474–19481.
- Roux J, et al. 2014. Patterns of positive selection in seven ant genomes. *Mol Biol Evol*. 31(7):1661–1685.

- Royet J, et al. 2011. Peptidoglycan recognition proteins: modulators of the microbiome and inflammation. *Nat Rev Immunol.* 11(12): 837–851.
- Rutrecht ST, Brown MJ. 2008. The life-history impact and implications of multiple parasites for bumble bee queens. *Int J Parasitol.* 38(7): 799–808.
- Rutschmann S, Kilinc A, Ferrandon D. 2002. Cutting edge: the toll pathway is required for resistance to gram-positive bacterial infections in *Drosophila*. *J Immunol.* 168(4):1542–1546.
- Sadd BM, et al. 2015. The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol.* 16(1):1–32.
- Schmid-Hempel P. 1998. Parasites in social insects. Princeton (NJ): Princeton University Press.
- Shakeel M, et al. 2019. Role of serine protease inhibitors in insect-host-pathogen interactions. *Insect Biochem Physiol.* 102(3):e21556.
- Shultz AJ, Sackton TB. 2019. Immune genes are hotspots of shared positive selection across birds and mammals. *Elife* 8:e41815.
- Shykoff JA, Schmid-Hempel P. 1991. Parasites and the advantage of genetic variability within social insect colonies. *Proc R Soc Lond B Biol Sci.* 243(1306):55–58.
- Siva-Jothy MT, Moret Y, Rolff J. 2005. Insect immunity: an evolutionary ecology perspective. *Adv Insect Phys.* 32:1–48.
- Smith MD, et al. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol.* 32(5):1342–1353.
- Sommer S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool.* 2(1):1–18.
- Stout JC, Murphy JT, Kavanagh S. 2019. Assessing market and non-market values of pollination services in Ireland (Pollival). Environmental Protection Agency Research Report, Wexford, Ireland: Johnstown Castle, Co. Wexford, Ireland. Available from: [https://www.epa.ie/publications/research/biodiversity/Research\\_Report\\_291.pdf](https://www.epa.ie/publications/research/biodiversity/Research_Report_291.pdf). Accessed on 24th March 2023.
- Strand MR. 2008. The insect cellular immune response. *Insect Sci* 15(1):1–14.
- Szpiech ZA. 2021. Selscan 2.0: scanning for sweeps in unphased data. *bioRxiv*. Available from: <https://www.biorxiv.org/content/10.1101/2021.10.22.465497v2>. Accessed on 24th March 2023.
- Tan W, et al. 2021. Population genomics reveals variable patterns of immune gene evolution in monarch butterflies (*Danaus plexippus*). *Mol Ecol.* 30(18):4381–4391.
- Tsakas S, Marmaras V. 2010. Insect immunity and its signalling: an overview. *Invertebrate Surviv J.* 7(2):228–238.
- Unckless RL, Lazzaro BP. 2016. The potential for adaptive maintenance of diversity in insect antimicrobial peptides. *Philos Trans R Soc B Biol Sci.* 371(1695):20150291.
- Valanne S, Wang J-H, Rämet M. 2011. The *Drosophila* toll signaling pathway. *J Immunol.* 186(2):649–656.
- Velthuis HH, Van Doorn A. 2006. A century of advances in bumblebee domestication and the economic and environmental aspects of its commercialization for pollination. *Apidologie (Celle).* 37(4): 421–451.
- Wagner DL. 2020. Insect declines in the Anthropocene. *Annu Rev Entomol.* 65:457–480.
- Wang Y, et al. 2022. Direct and indirect impacts of positive selection on genomic variation in *Drosophila serrata*. *bioRxiv*. Available from: <https://www.biorxiv.org/content/10.1101/2022.03.31.486660v1>. Accessed on 24th March 2023.
- Weaver DB, et al. 2021. Multi-tiered analyses of honey bees that resist or succumb to parasitic mites and viruses. *BMC Genomics.* 22(1):1–15.
- Xing Z, et al. 2019. The DDX5/dbp2 subfamily of DEAD-box RNA helicases. *Wiley Interdiscip Rev RNA.* 10(2):e1519.
- Ying SH, Feng MG. 2019. Insight into vital role of autophagy in sustaining biological control potential of fungal pathogens against pest insects and nematodes. *Virulence* 10(1):429–437.
- Yourth C, et al. 2008. Effects of natal and novel *Crithidia bombi* (Trypanosomatidae) infections on *Bombus terrestris* hosts. *Insectes Soc.* 55(1):86–90.
- Zhang K, et al. 2021. Scavenger receptor C regulates antimicrobial peptide expression by activating toll signaling in silkworm, *Bombyx mori*. *Int J Biol Macromol.* 191:396–404.
- Zheng X. 2013. A tutorial for the R Package SNPRelate. Washington: University of Washington [Preprint].
- Zou Z, et al. 2006. Comparative analysis of serine protease-related genes in the honey bee genome: possible involvement in embryonic development and innate immunity. *Insect Mol Biol.* 15(5):603–614.

**Associate editor:** Matthew Webster