Genomic signatures of recent adaptation in a wild bumblebee

Thomas J. Colgan*[1,2], Andres N. Arce[3,4], Richard J. Gill[3], Ana Ramos Rodrigues[3], Abdoulie Kanteh[1], Elizabeth J. Duncan[5], Li Li[1], Lars Chittka[1] and Yannick Wurm*[1,6]

Author affiliations:

1. School of Biological and Chemical Sciences, Queen Mary University of London, London, E1 4NS, United Kingdom.
2. Current Address: Institute of Organismic and Molecular Evolution, Johannes Gutenberg University Mainz, 55128 Mainz, Germany.
3. Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, SL5 7QN, United Kingdom.
4. Current Address: School of Engineering, Arts, Science & Technology, University of Suffolk, Ipswich, IP3 0FS, United Kingdom.
5. School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds, LS2 9JT, United Kingdom.
6. Alan Turing Institute, London, NW1 2DB, United Kingdom.

***Corresponding authors:**

Thomas J. Colgan: tcolgan@uni-mainz.de

Yannick Wurm: y.wurm@qmul.ac.uk

Keywords: Ecological genomics, population genomics, selective sweeps, bees, horizontal gene transfer, genetic health

Abstract

Environmental changes threaten insect pollinators, creating risks for agriculture and ecosystem stability. Despite their importance, we know little about how wild insects respond to environmental pressures. To understand the genomic bases of adaptation in an ecologically important pollinator, we analyzed genomes of *Bombus terrestris* bumblebees collected across Great Britain. We reveal extensive genetic diversity within this population, and strong signatures of recent adaptation throughout the genome affecting key processes including neurobiology and wing development. We also discover unusual features of the genome, including a region containing 53 genes that lacks genetic diversity in many bee species, and a horizontal gene transfer from a *Wolbachia* bacteria. Overall, the genetic diversity we observe and how it is distributed throughout the genome and the population should support the resilience of this important pollinator species to ongoing and future selective pressures. Applying our approach to more species should help understand how they can differ in their adaptive potential, and to develop conservation strategies for those most at risk.

Introduction

Behavioral experiments and analyses of observation records have shown that pesticide use, habitat fragmentation, emerging diseases and climatic change threaten insect pollinators including bees (Brown and Paxton 2009; Vanbergen 2013; Ollerton et al. 2014; Goulson et al. 2015). Despite the resulting risks for agricultural yields and for ecosystem stability, we know little about how wild insects may adapt to such environmental pressures. Similarly, we understand relatively little about how genomes are shaped in the wild (Harrisson et al. 2014). If a species has adapted in response to a detrimental environmental pressure, then we should see changes in the alleles of genes involved (Hohenlohe et al. 2010; Ellegren 2014) (Fig. 1A). Analyzing genomes of many individuals can identify such changes and reveal the constraints and adaptive potential of species (Chen et al. 2018; van Klink et al. 2020). The resulting knowledge should support conservation efforts and practices (Supple and Shapiro 2018).

The annual reproductive bumblebee *Bombus terrestris* is ideal for understanding how coping with recent rates of environmental change is possible because, unlike many other pollinators, it has shown little evidence of population decline (Ollerton et al. 2014). Furthermore, because male bumblebees are haploid, their genome sequences are unambiguous and intrinsically phased, providing more analytical power than the diploid genomes of female bumblebees and of many other insects. To understand which bumblebee genes and molecular processes underlie responses to recent selective pressures, we sequenced the genomes of male *B. terrestris* bumblebees from across Great Britain. We subsequently identified and characterized the genomic regions showing the strongest signs of recent adaptive evolution in this population. Our characterization of the amount and distribution of genetic diversity throughout the bumblebee genome provides measures of the genetic health of this species and highlights genes and processes through which it has recently adapted.

Results

Weak population substructure among *B. terrestris* in mainland Britain

We collected 46 unrelated male *B. terrestris* from across Great Britain and sequenced their genomes (411-fold total coverage; Figure 1B; Table S1). We found 1,227,312 single nucleotide polymorphisms (SNPs), with an average nucleotide diversity $\boldsymbol{\pi}$ of 1.51×10^{-3}.

To understand whether population structure constrains adaptation in this species, we performed identity-by-state and co-ancestry-based analyses. These analyses indicate that our dataset

represents one population (Figs 1C and S1-2). Similarly, while the second principal component correlates with latitude (Pearson's r=0.8, Figs 1C and S3), individual principal components explained at most 2.46% of genetic variation. Thus, despite the geographic heterogeneity of Great Britain, there is sufficient gene flow for these bees to be considered as one panmictic population, implying that new alleles have the potential to readily spread. The weak substructure of British *B. terrestris* is supported by studies using fewer markers (Schmid-Hempel et al. 2007; Lye et al. 2011; Moreira et al. 2015). Our result also indicates that no subset of our samples has the type of large-scale differentiation that could be expected from a cryptic subspecies.

Selection is fine-tuning functional regions throughout the bumblebee genome

We used two approaches to identify signatures of recent selection in the genome. First, we identified large "hard" sweep regions, where selection on an allele can lead to haplotype fixation (Berry et al. 1991). For this, we identified genomic segments longer than 100,000 nucleotides with significantly lower nucleotide diversity than the rest of the genome (z-score<-2σ). We found ninety such segments. Our second approach detected more localized signatures of selection, and "soft" sweeps, where two or more haplotypes are at high frequency. This can occur, for example, when strong selection on a new mutation occurs after a first allele reaches fixation, or when selection favors different alleles in different habitats (Hermisson and Pennings 2005). For this, we determined for each SNP the metric $|nS_L|$ ("number of segregating sites by length"), a measure of haplotype homozygosity that is robust to variation in recombination and mutation rates; an $|nS_L|$ score greater than 2 is considered evidence of recent selection (Ferrer-Admetlla et al. 2014; Szpiech and Hernandez 2014). The 10,132 SNPs with the highest 1% of $|nS_L|$ scores have particularly strong signatures of recent selection ($|nS_L|>2.56$, *i.e.*, ≳3 standard deviations from the average; Figure 2A; Table S2) and are typically within 300 nucleotides of SNPs with low $|nS_L|$ scores. This indicates that recombination rapidly breaks down haplotypes in this species and that selection has acted in fine-tuned manners. The SNPs with high $|nS_L|$ scores are mostly in genic regions (79%; $p<10^{-15}$) and, within these regions, are similarly represented in coding and non-coding sequence ($p>0.05$), indicating that selection recently acted on protein function and on gene regulation.

To understand which types of biological functions were under the strongest recent selection pressures, we inspected annotations of genes with the strongest $|nS_L|$ scores and performed rank-based analyses of Gene Ontology and InterPro descriptions of all bumblebee genes (Bonferroni adjusted $p<0.05$; Figs 2C and S4; Tables S3-4). The overviews of loci under recent selection, and the biological and molecular processes they affect, represent valuable resources for future

phenotypic work on bees and on adaptation in natural insect populations (Figs 2C and S4; Tables S3-4). Below, we highlight five particularly striking patterns regarding genes and regions with the strongest signatures of selective sweeps.

Strong selection on transcription factors

Genes related to transcriptional regulation were overrepresented among genes under selection (Figs 2C and S4; Tables S3-4). In particular, the gene with the strongest evidence of recent selection is the *B. terrestris* ortholog to the *schnurri* gene ($|nS_L|$=5.14). In *Drosophila*, this transcription factor regulates embryonic patterning and wing patterning through the Decapentaplegic pathway (Torres-Vazquez et al. 2000). Another transcription factor, the ortholog to the *ventral veins lacking* gene (*vvl*), has the 9th highest $|nS_L|$ score ($|nS_L|$=4.54). In *Drosophila*, *vvl* is involved in steroid biosynthesis and embryonic brain development (Meier et al. 2006; Danielsen et al. 2014), and intriguingly also interacts with the Decapentaplegic pathway to affect wing imaginal disc development (Certel et al. 2000) and vein patterning (de Celis et al. 1995). These results, together with "wing morphogenesis" being the eighth most overrepresented Gene Ontology description among genes under selection, suggest that there was strong recent selection on wing structure. Such selection could be linked to recent changes to foraging or flight patterns (Memmott et al. 2007; Miller-Struttmann et al. 2015), because climatic changes modified the physical constraints of flying (Corbet et al. 1993), or perhaps in response to pathogens, such as the deformed wing virus, which can cause extensive wing abnormalities in infected individuals (Genersch et al. 2006).

Strong selection acting on genes involved in bumblebee neurobiology

Neurological genes were overrepresented among genes with the highest $|nS_L|$ scores (Figure 2C). In line with this, four of the 30 genes with the highest $|nS_L|$ scores have potential roles in neurotransmission (*gamma-aminobutyric acid receptor alpha-like*; $|nS_L|$=4.44, *glutamate receptor ionotropic kainate 2*; $|nS_L|$=4.39), axon guidance (*Down Syndrome cell adhesion molecule 2*; $|nS_L|$=4.62), and memory formation (*neurotrimin*; $|nS_L|$=4.1). Furthermore, selection on G protein-coupled receptor signaling in *B. terrestris* mirrors previous analyses on honeybees (Harpur et al. 2014; Wallberg et al. 2014; Avalos et al. 2017). These receptor targets of hormones, pheromones and neurotransmitters are thus long-term targets of selection in social bees, potentially for roles responding to social or to environmental cues (Hauser et al. 2006). Selection on neurological genes could, for example, be linked to the need to improve complex cognitive and social behaviors of bumblebees (Loukola et al. 2017), for remembering increasingly complex foraging routes due to patchier habitats, or for neurological changes because of exposure to neurotoxins.

Positive selection on a gene horizontally transferred from *Wolbachia*

We used similarity searches to understand the potential functions of uncharacterized genes among the genes with the twenty highest $|nS_L|$ scores. This showed that the gene with the 16th-strongest signature of selection (LOC105666162; $|nS_L|$=4.33; Table S2; Figure S5) was horizontally transferred to *Bombus* from *Wolbachia*, a genus of bacterial endosymbionts that infect many invertebrates. Indeed, LOC105666162 has strong similarity to a gene in *Wolbachia* (BLASTP e-values<10^{-16}) but not to most other insects (Figure S6; Table S5; Supplementary Information). The data overwhelmingly suggest that this gene is integrated in the *Bombus* genome and not an artifact of potential contamination (Supplementary Information). Indeed, the same two orthologs flank LOC105666162 in *B. terrestris* and in *B. impatiens,* and this gene is present in genomic sequences of 14 other *Bombus* species (Sadd et al. 2015, Lin et al. 2019), suggesting that horizontal transfer occurred 40 million years ago (Peters et al. 2017). Additionally, sequencing depth across LOC105666162 is similar to the rest of the genome (paired t-test, $t_{df=95}$=-0.73, p=0.47). Crucially, we find no other *Wolbachia*-related sequences in our samples, consistent with the absence of evidence that *Wolbachia* could infect *Bombus.* So, what might LOC105666162 do? Unfortunately, functional work on this gene and its *Wolbachia* homolog WD0147 are lacking, but we do have some clues. Horizontally transferred genes are often inactive, yet LOC105666162 is expressed in germlines and other tissues of both sexes and castes (Harrison et al. 2015; Lewis et al. 2018; Colgan et al. 2019), consistent with it being functional (Figure S5; Table S6). The *Wolbachia* homolog is expressed in infected *Drosophila* gonads (Papafotiou et al. 2011); based on its amino acid sequence and expression profiles, this gene is a strong candidate for driving mechanisms of cytoplasmic incompatibility (Metcalf et al. 2014). One could speculate that LOC105666162 contributes to the lack of *Wolbachia* in bumblebees.

An evolutionary conserved region of extremely low genetic diversity

A 200,000 nucleotide-long region stood out in our analysis because it contains only 56 SNPs and thus has 21-fold lower nucleotide diversity ($\boldsymbol{\pi}$~7×10^{-5}) than the genome-wide average ($\boldsymbol{\pi}$=1.5×10^{-3}; $t_{df=46}$=90.4, p<10^{-15}; Figure 3). Furthermore, the low-diversity region has particularly high gene density (53 genes; z-score<-2σ) and represents a solid haplotype, with a population-derived recombination rate 298x lower ($\rho=1.2\times10^{-4}$) than the genome-wide average (ρ=0.035, $t_{df=1226700}$=-460, p<10^{-15}).

To test whether the characteristics of this region are specific to *B. terrestris*, we identified orthologous regions in another bumblebee species *Bombus impatiens* and in the honeybee *Apis mellifera*. The orthologous region in *B. impatiens* contains 52 of the 53 genes and similarly has

lower diversity (π~1.7×10^{-4}) than other regions (genome-wide average $\pi=1.2\times10^{-3}$; $t_{df=18}$=-22.8, $p<10^{-15}$). Orthology to the honeybee is split between two regions separated by 4.4Mb, indicating that rearrangements have occurred since the common ancestor of bumblebees and honeybees existed 78 million year ago. For both regions, honeybee populations had at least 13-fold lower nucleotide diversity than the rest of the genome (region 1: $t_{df=19.328}$=-98.9, $p<10^{-15}$; region 2: $t_{df=15.8}$=-65.2, $p<10^{-15}$; Figure 3). These patterns indicate that an intrinsic long-standing process is responsible for the low genetic diversity of these regions in bees. While the regions include genes that are likely under strong purifying selection (Table S7), no particular gene annotation was overrepresented which could help interpretation. Unlike the rest of the genome, the lack of genetic diversity in this large region suggests that bees will have limited ability to adapt to selection pressures involving the genes it contains.

Selection on potential insecticide susceptibility genes

Selection for resistance to neurotoxic pesticides can lead to changes in expression or sequence of target receptors or detoxification enzymes in insect pests (Ffrench-Constant 2013). Because bumblebees can be exposed to pesticides when foraging on crops, and given the extensively documented detrimental effects of pesticide exposure on bumblebee health (Vanbergen 2013; Goulson et al. 2015), we preliminarily examined signatures of recent selection in genes for which orthologs in other species are known to be involved in responses to insecticide exposure. Four target receptors of insecticides were among the 10% of genes with the highest $|nS_L|$ scores: three nicotinic acetylcholine receptor subunits which are targets of neonicotinoid insecticides (*nAChR1a*, *nAChR6a*, *nAChR7a*; $|nS_L|$>3.12 for all; Table S8), and *metabotropic glutamate receptor 2* ($|nS_L|$=3.59), a target of the natural plant toxin L-canavanine (Mitri et al. 2009). Mutations in orthologs of two of the nicotinic acetylcholine receptor subunit genes confer resistance to neonicotinoid exposure in other species (Puinean et al. 2013; Shimada et al. 2020). Among genes putatively involved in detoxification, five cytochrome P450s and four carboxylesterases had strong signatures of selection (top 10% of genes with high $|nS_L|$ scores ($|nS_L|$>2.89); Table S8; Figure S7). We also found strong signatures of selection for 42 genes that are differentially expressed in bumblebees after exposure to neonicotinoid pesticides (Bebane et al. 2019; Colgan et al. 2019; Table S9). There was no overlap between these 42 genes and those previously identified as having a role in insecticide resistance. Future research will help pinpoint the reasons for the patterns we observe, and whether some of the recent changes may reduce susceptibility to toxins naturally present in pollen and nectar, or to synthetic pesticides.

Discussion

Human-induced environmental changes add to long-standing ecological and evolutionary challenges faced by wild animals. Identifying potential causes of pollinator declines has to date relied on inferences from laboratory experiments or on correlative associations in the field. Our study takes an important step towards understanding the bases of resilience of an important pollinator species by uncovering signatures left in the organism's genetic "blueprint" in response to selective pressures. The strong signatures of selection we find at loci distributed throughout the *B. terrestris* genome are consistent with the view that insect pollinators face many different pressures. Environmental pressures likely contributed to recent changes we found to affect *B. terrestris* genes underlying physiology, neurology, and wing development.

Bumblebees also face intra- and interspecies pressures including, competition for food, habitat, and mates, and pressures from predators, pathogens and parasites. Large-scale gene expression and functional genomic datasets are only beginning to be produced for bumblebees (López-Osorio and Wurm 2020) and will be crucial for disentangling how the specific changes we observed affect phenotypes and fitness. Similarly, historical sampling of museum specimens could help characterize changes over time in morphology, population structure and allele frequencies.

The fine-tuning of adaptive responses in *B. terrestris* is highlighted by our finding strong signatures of selective sweeps within few nucleotides of neutrally evolving loci. Several characteristics of this species likely facilitate this. Crucially, the high recombination rate (Liu et al. 2017) and social lifestyle of *B. terrestris* mean that one queen can produce hundreds of haploid males, encompassing a broad diversity of allelic combinations. These males are fully exposed to the environment as they spend weeks foraging and trying to attract a mate (Wolf et al. 2012). Male bees are also subject to haploid selection, which should lead to faster adaptation than in diploid species (Meisel and Connallon 2013; Pracana et al. 2021). Furthermore, the broad gene flow and large population size of *B. terrestris* enables the maintenance of large amounts of genetic diversity and the rapid spread of adaptive alleles. While our data would be unable to detect slight changes in population size over the past century, our data and analyses support the absence of major recent population bottlenecks in this species. Future comparisons with sister species including those that are declining will clarify whether *B. terrestris* may have additionally harbored a generalist genetic toolkit further pre-disposing it to resilience.

We show that locating recent signatures of selection throughout the genome can indicate which genes and pathways changed for *B. terrestris* to adapt in response to the pressures it has faced.

Furthermore, the amount and distribution of genetic diversity we observed throughout its genome suggest that this bumblebee species maintains an ability to respond to future pressures. Our work in this bumblebee species complements recent efforts in vertebrates and model systems. Future comparative genomic studies with other social and solitary pollinators will improve our ability to disentangle why species differ in their resilience to recent environmental changes. Additionally, scaling up our approach will enable the creation of frameworks for predicting detailed responses to environmental challenges for entire ecological networks. Overall, while insect declines are worrying, we show how at least one common pollinator is adapting.

Materials and Methods

Bumblebee collection, DNA extraction and sequencing

In the summer of 2014 we collected up to two males from each of 28 sites, with each site being >20 km from the nearest neighboring site (Figure 1B). Male *Bombus terrestris* (large earth or buff-tailed bumblebee) were caught using butterfly nets and transferred into individual 100 ml pots after morphological confirmation of sex and species. Pots were placed into a bag at 4-10°C. Within two hours, males were rapidly transferred to 2 ml cryotubes and then snap-frozen in liquid nitrogen. Subsequent storage was at -80°C.

From each bee, dissected tissue was homogenized in 200µl of phenol in a 2 ml screw-cap tube (Table S1). Subsequently, DNA was extracted using phenol-chloroform followed by purification with the Sigma GenElute Mammalian Genomic DNA miniprep kit. DNA purity was initially assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific, UK) followed by quantification with a Qubit v3 fluorometer (Thermo Fisher Scientific, UK). DNA from each male was fragmented to ~550 bp using a Covaris M220 ultrasonicator and fragment size distribution assessed using a TapeStation 2200 (Agilent Technologies, UK). From each sample, we prepared an individually indexed Illumina TruSeq PCR-free DNA library, which was quantified using qPCR MasterMix (ABI Prism) and primer premix (Kapa Biosystems, UK). Libraries were pooled in equimolar concentrations and pairs of 125 bp sequences were produced on two lanes of Illumina HiSeq 2500 at Biomedical Research Centre Genomics, London, UK. Five samples were additionally sequenced on one lane of Illumina HiSeq 2500 at Oxford Genomics, Oxford, UK.

Quality assessment and filtering of raw Illumina sequences

We obtained 616 million paired-end reads from the 51 samples we initially collected. Using bowtie2 (v.2.2.5; (Langmead and Salzberg 2012)) with the parameter '-X 1000', we aligned raw reads to the *B. terrestris* reference genome (GCF_000214255.1; (Sadd et al. 2015)). The 422× cumulative genome coverage provided strong power to detect sites with nucleotide sequence polymorphism. Four males were removed from all biological analyses due to low coverage. A fifth male was only used for the analysis of contaminant sequences (SI Appendix) because >58.1% of reads from this male lacked similarity to the reference genome. The mean mapped coverage for each of the remaining 46 samples was 11.8× (min: 7×; max: 26.7×). Quality of raw reads was assessed using FastQC (v.0.11.3; https://www.bioinformatics.babraham.ac.uk/projects/fastqc). Illumina adapters were detected

and removed using Trimmomatic (v.0.33; Bolger et al. 2014). Using Khmer (v.2.1.1; Crusoe et al. 2015) to first interleave pairs of reads, we removed sequences of low quality (where >25% of the read has a Phred quality score of strictly <20) using the fastx toolkit (v.0.0.14; http://hannonlab.cshl.edu/fastx_toolkit). We used Khmer to remove 31-mers present three or fewer times across the entire dataset, as they likely represent technical artefacts or particularly rare variants that we would be unable to analyze. Sequences shorter than 50 bp were removed using seqtk (v.1.0-r82-dirty; https://github.com/lh3/seqtk). The final cleaned dataset thus comprised 46 males with a mean coverage of 11.1× (min 6.7×; max 24.4×). This cleaned dataset provides sufficient power to genotype the majority of polymorphic sites because 96.2% of the genome had >1X coverage in each of the 46 males. Overall, ~99% of the reference genome had at least 20× coverage.

Identification of polymorphic sites and genotyping of individuals

After mapping cleaned reads to the reference assembly using bowtie2, we called variants using freebayes (v.1.0.2-29-g41c1313; Garrison and Marth 2012) with the following parameters: --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-fraction 0.25 --min-coverage 1 --use-reference allele. We first removed the aforementioned five low-coverage individuals as they were each missing >10% of genotype calls, thus retaining data from 46 males. We then removed entire SNPs with low genotype quality scores (--minQ 20) and variants in collapsed repetitive regions (--max-mean-DP 100) using VCFtools (v.0.1.15; Danecek et al. 2011). We removed sites that appeared to be heterozygous, which is impossible in haploids, and all sites with more than two alleles as they also likely represent collapsed regions in the reference genome. To further reduce dataset complexity, we used --remove-indels to only consider single nucleotide polymorphisms (SNPs). We calculated allele frequencies and retained genotypes only where the rare allele was present in at least two males. Finally, we only considered those SNPs in regions of the genome that are mapped to the 18 linkage groups (representative of chromosomes). Mean nucleotide diversity $\boldsymbol{\pi}$ was calculated using 10 kb sliding windows with 5 kb overlap using PopGenome (v.2.2.4; Pfeifer et al. 2014).

Assessment population structure

We investigated among collected bumblebees by performing identity-by-state (IBS) analyses on a pruned set of SNPs generated by SNPRelate (v.1.8.0; Zheng et al. 2012) using parameters that are similar to those previously used for *Drosophila* (Schmidt et al. 2017) (--ld-threshold=0.2 --slide.max.n=500). We further investigated population structure using three

approaches with unpruned SNPs: principal component analysis using SNPRelate, ADMIXTURE (v.1.3.0; (Alexander et al. 2009)) with K=1-40 using cross-validation (--cv) as a measure to identify the best K value, and the linkage-aware approach fineSTRUCTURE (v.0.1.0; Lawson et al. 2012).

Evidence of recent selective sweeps

First, we identified regions of the genome with particularly low nucleotide diversity, indicative of "hard" sweeps. Second, to identify potential "soft" selective sweeps, we calculated nS_L (Ferrer-Admetlla et al. 2014) for all high confidence SNPs using selscan (v.1.1.0b; Szpiech and Hernandez 2014). This metric is a measure of extended haplotype homozygosity. We normalized all nS_L scores against the empirical genome-wide distribution using selscan "norm", using default settings. We used the top 1% ($|nS_L| \geq 2.56$) absolute score of the nS_L metric ($|nS_L|$) for all downstream analyses because $|nS_L| \geq 2$ indicates a selective sweep. Normalized $|nS_L|$ scores per gene, as well as NCBI RefSeq gene symbol and description, are provided in Table S2.

Data availability

Sequencing data are available from NCBI BioProject (PRJNA628944).

Code availability

The scripts underpinning our analysis are available from: https://github.com/Joscolgan/bter_population_genomics.

Acknowledgements

We thank Richard Nichols, Rodrigo Pracana, Eckart Stolle, Carlos Martinez Ruiz and Gino Brignoli for advice on analysis and wording. We thank Monika Struebig and Christopher Durrant for help with molecular work. We also thank BRC Genomics, London, and Oxford Genomics, Oxford, UK, for the sequencing of genomic libraries. This work was supported by the Biotechnology and Biological Sciences Research Council (grant BB/K004204/1), the Natural Environment Research Council (grants NE/L00626X/1 and NE/L00755X/1), NERC EOS Cloud and QMUL Research-IT and MidPlus computational facilities (The Engineering and Physical Sciences Research Council; grant EP/K000128/1).

Author Contributions

YW, RJG, LC, TJC and ANA conceived, designed the project and coordinated collections. ARR helped collect wild bumblebees. LL performed dissections. TJC performed most genome-level analyses, with assistance from YW and AK. EJD helped with interpretation of results. TJC and YW drafted the manuscript. All authors wrote and revised the final manuscript.

Figures

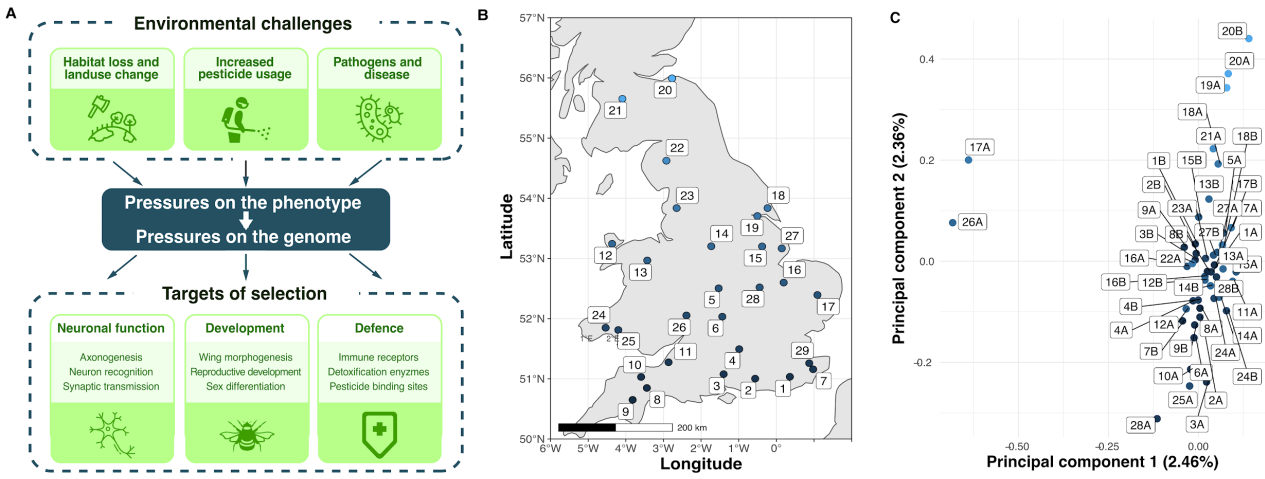

Figure 1 | Environmental pressures affecting insect pollinators and population structure of wild-caught British *B. terrestris*. (A) Overview of key environmental selective pressures on wild bumblebee populations, and some of the biological pathways and processes expected to be under selection in response. **(B)** Twenty-eight collection sites across Great Britain, colored according to latitude. **(C)** Population structure of 46 males according to the first two principal components (PC1 and PC2). Each point refers to one male, with up to two males (A and B) per site, colored according to collection site latitude.

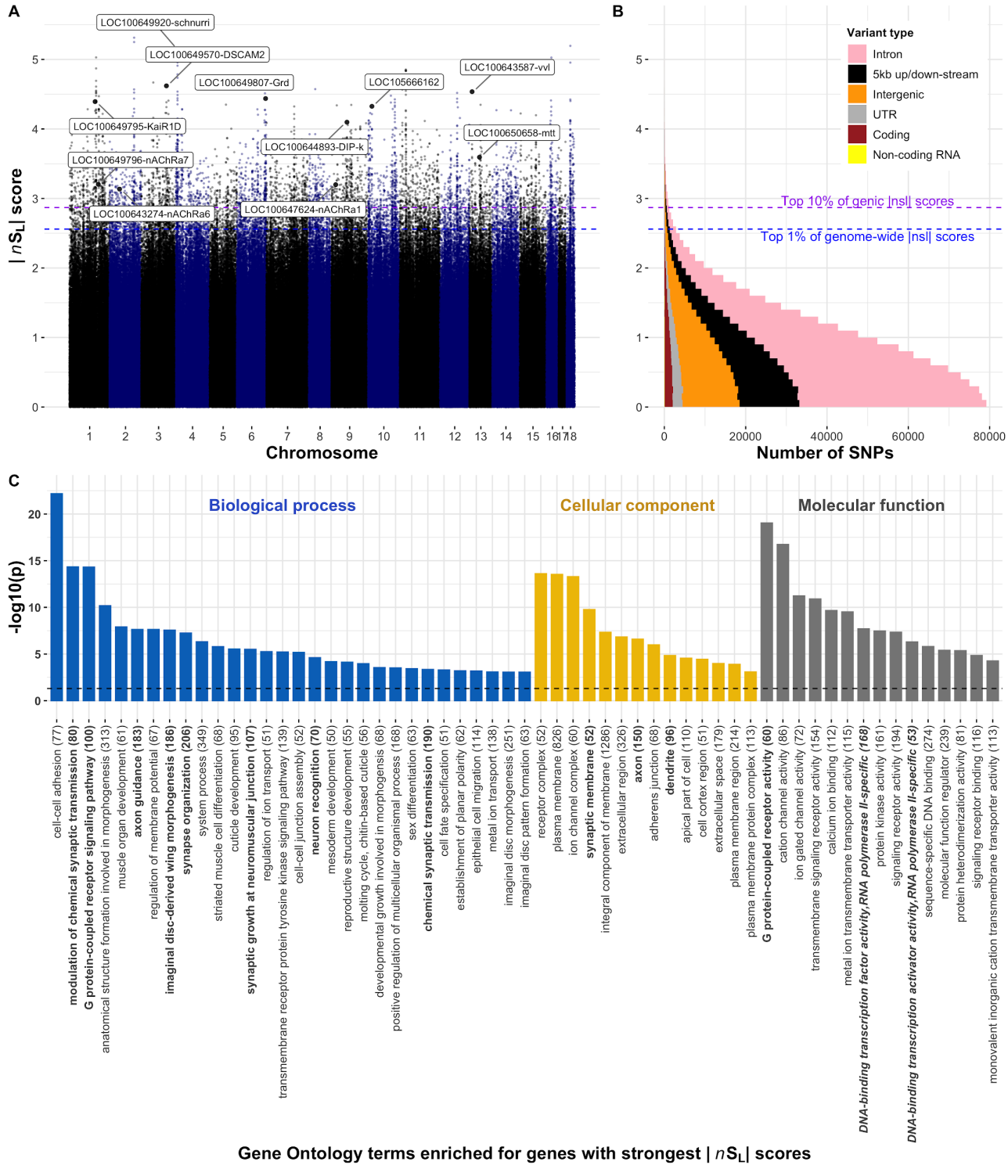

Figure 2 | Genome-wide signatures of selective sweeps in British *Bombus terrestris* bumblebees. (A) $|nS_L|$ measures of selection for all SNPs in the bumblebee genome. Each dot represents one SNP; labelled dots represent the SNP with highest $|nS_L|$ score for genes of interest, including transcription factors, insecticide susceptibility genes and a *Wolbachia*-like gene, with high $|nS_L|$ scores. Labels indicate Flybase gene symbols when clear *Drosophila* orthology exists, otherwise, the NCBI gene symbol is provided. Blue and purple horizontal dashed lines respectively indicate the 1st percentile of overall $|nS_L|$ scores and 10th percentile

of genic $|nS_L|$ scores. **(B)** Distributions of $|nS_L|$ scores show that most SNPs are in genic regions, and that most $|nS_L|$ scores are consistent with neutral or purifying rather than directional evolution as 96% of SNPs have $|nS_L| < 2$. **(C)** Diverse Gene Ontology terms are enriched in genes with high $|nS_L|$ scores (-log10 transformed Bonferroni-adjusted *p* values). Terms associated with roles in neurology and transcription factor activity are respectively highlighted in bold and bold italics. The total number of annotated genes for each term is in parentheses.

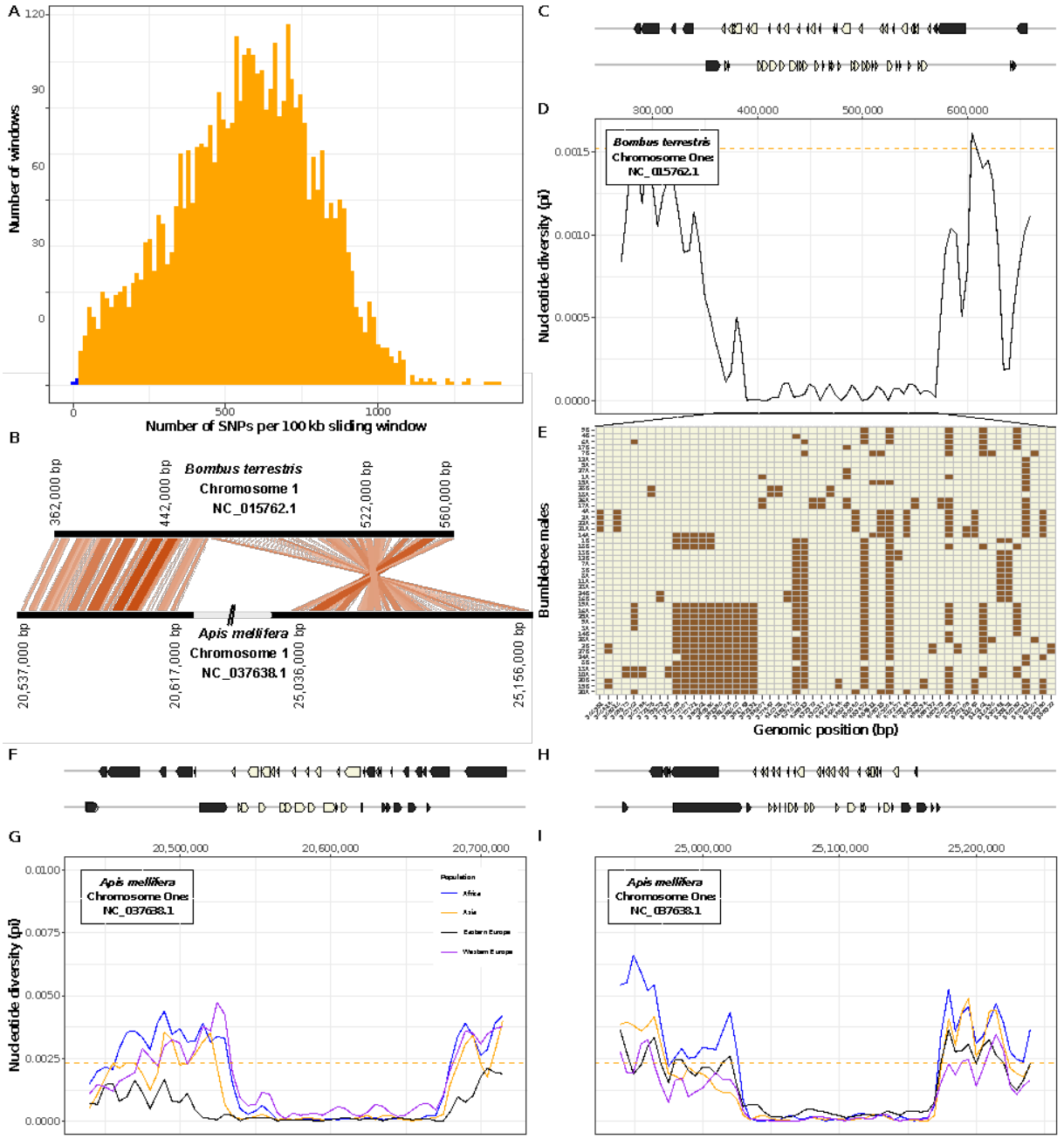

Figure 3 | Conserved gene-rich region of low nucleotide diversity in bumblebee and honeybee. (A) Numbers of SNPs identified in 100 kb sliding windows across the bumblebee genome. The two windows with the lowest numbers of SNPs (in blue) are adjacent to each other on chromosome one. **(B)** Relative genomic positions of homologous regions of low diversity on chromosomes one of *Bombus terrestris* and *Apis mellifera*. **(C)** Genomic coordinates of 53 genes present in (beige) and flanking (grey) the region of low diversity. **(D)** Nucleotide diversity (**π**, calculated in 10 kb sliding windows) is low in this region in comparison to flanking regions and to the genome-wide mean (dashed line). **(E)** Genotypes for each of 46 *B. terrestris* males (rows) at each SNP (columns; chromosomal coordinate shown in the x axis). Colors

indicate reference allele (beige) or alternative allele (brown). **(F-I)** In the honeybee *Apis mellifera,* homology with the region of low diversity in *B. terrestris* is split between two regions. For both regions, we show genomic positions of genes (F, H). In four populations, both regions have lower nucleotide diversity (**π**, calculated in 10 kb sliding windows) than flanking regions or the rest of the genome (dashed line; G, I).

References

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664.

Avalos A, Pan H, Li C, Acevedo-Gonzalez JP, Rendon G, Fields CJ, Brown PJ, Giray T, Robinson GE, Hudson ME, et al. 2017. A soft selective sweep during rapid evolution of gentle behaviour in an Africanized honeybee. *Nat. Commun.* 8:1550.

Bebane PSA, Hunt BJ, Pegoraro M, Jones ARC, Marshall H, Rosato E, Mallon EB. 2019. The effects of the neonicotinoid imidacloprid on gene expression and DNA methylation in the buff-tailed bumblebee *Bombus terrestris*. *Proc. Biol. Sci.* 286:20190718.

Berry AJ, Ajioka JW, Kreitman M. 1991. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* 129:1111–1117.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Brown MJF, Paxton RJ. 2009. The conservation of bees: a global perspective. *Apidologie* 40:410–416.

de Celis JF, Llimargas M, Casanova J. 1995. *Ventral veinless*, the gene encoding the Cf1a transcription factor, links positional information and cell differentiation during embryonic and imaginal development in *Drosophila melanogaster*. *Development* 121:3405–3416.

Certel K, Hudson A, Carroll SB, Johnson WA. 2000. Restricted patterning of vestigial expression in *Drosophila* wing imaginal discs requires synergistic activation by both Mad and the drifter POU domain transcription factor. *Development* 127:3173–3183.

Chen C, Wang H, Liu Z, Chen X, Tang J, Meng F, Shi W. 2018. Population genomics provide insights into the evolution and adaptation of the eastern honey bee (*Apis cerana*). *Mol. Biol. Evol.* 35:2260–2271.

Colgan TJ, Fletcher IK, Arce AN, Gill RJ, Ramos Rodrigues A, Stolle E, Chittka L, Wurm Y. 2019. Caste- and pesticide-specific effects of neonicotinoid pesticide exposure on gene expression in bumblebees. *Mol. Ecol.* 28:1964–1974.

Corbet SA, Fussell M, Ake R, Fraser A, Gunson C, Savage A, Smith K. 1993. Temperature and the pollinating activity of social bees. *Ecol. Entomol.* 18:17–30.

Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, Charbonneau A, Constantinides B, Edvenson G, Fay S, et al. 2015. The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Res.* 4:900.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.

Danielsen ET, Moeller ME, Dorry E, Komura-Kawa T, Fujimoto Y, Troelsen JT, Herder R, O'Connor MB, Niwa R, Rewitz KF. 2014. Transcriptional control of steroid biosynthesis genes in the *Drosophila* prothoracic gland by ventral veins lacking and knirps. *PLoS Genet.* 10:e1004343.

Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* 29:51–63.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol. Biol. Evol.* 31:1275–1291.

Ffrench-Constant RH. 2013. The molecular genetics of insecticide resistance. *Genetics* 194:807–815.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv [q-bio.GN]* [Internet]. Available from: http://arxiv.org/abs/1207.3907

Genersch E, Yue C, Fries I, de Miranda JR. 2006. Detection of Deformed wing virus, a honey bee viral pathogen, in bumble bees (*Bombus terrestris* and *Bombus pascuorum*) with wing deformities. *J. Invertebr. Pathol.* 91:61–63.

Goulson D, Nicholls E, Botías C, Rotheray EL. 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* 347:1255957.

Harpur BA, Kent CF, Molodtsova D, Lebon JMD, Alqarni AS, Owayss AA, Zayed A. 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc. Natl. Acad. Sci. U. S. A.* 111:2614–2619.

Harrison MC, Hammond RL, Mallon EB. 2015. Reproductive workers show queenlike gene expression in an intermediately eusocial insect, the buff-tailed bumble bee *Bombus terrestris*. *Mol. Ecol.* 24:3043–3063.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P. 2014. Using genomics to characterize evolutionary potential for conservation of wild populations. *Evol. Appl.* 7:1008–1025.

Hauser F, Cazzamali G, Williamson M, Blenau W, Grimmelikhuijzen CJP. 2006. A review of neurohormone GPCRs present in the fruitfly *Drosophila melanogaster* and the honey bee *Apis mellifera*. *Prog. Neurobiol.* 80:1–19.

Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169:2335–2352.

Hohenlohe PA, Phillips PC, Cresko WA. 2010. Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. *Int. J. Plant Sci.* 171:1059–1071.

van Klink R, Bowler DE, Gongalsky KB, Swengel AB, Gentile A, Chase JM. 2020. Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science* 368:417–420.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357–359.

Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. *PLoS Genet.* 8:e1002453.

Lewis SH, Quarles KA, Yang Y, Tanguy M, Frézal L, Smith SA, Sharma PP, Cordaux R, Gilbert C, Giraud I, et al. 2018. Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat Ecol Evol* 2:174–181.

Lin G, Huang Z, Wang L, Chen Z, Zhang T, Gillman LN, Zhao F. 2019. Evolutionary rates of bumblebee genomes are faster at lower elevations. *Mol. Biol. Evol.* 36:1215–1219.

Liu H, Jia Y, Sun X, Tian D, Hurst LD, Yang S. 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.* 34:119–130.

López-Osorio F, Wurm Y. 2020. Healthy pollinators: Evaluating pesticides with molecular medicine approaches. *Trends Ecol. Evol.* 35:380–383.

Loukola OJ, Solvi C, Coscos L, Chittka L. 2017. Bumblebees show cognitive flexibility by improving on an observed complex behavior. *Science* 355:833–836.

Lye GC, Lepais O, Goulson D. 2011. Reconstructing demographic events from population genetic data: the introduction of bumblebees to New Zealand. *Mol. Ecol.* 20:2888–2900.

Meier S, Sprecher SG, Reichert H, Hirth F. 2006. *ventral veins lacking* is required for specification of the tritocerebrum in embryonic brain development of *Drosophila*. *Mech. Dev.* 123:76–83.

Meisel RP, Connallon T. 2013. The faster-X effect: integrating theory and data. *Trends Genet.* 29:537–544.

Memmott J, Craze PG, Waser NM, Price MV. 2007. Global warming and the disruption of plant-pollinator interactions. *Ecol. Lett.* 10:710–717.

Metcalf JA, Jo M, Bordenstein SR, Jaenike J, Bordenstein SR. 2014. Recent genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism. *PeerJ* 2:e529.

Miller-Struttmann NE, Geib JC, Franklin JD, Kevan PG, Holdo RM, Ebert-May D, Lynn AM, Kettenbach JA, Hedrick E, Galen C. 2015. Functional mismatch in a bumble bee pollination mutualism under climate change. *Science* 349:1541–1544.

Mitri C, Soustelle L, Framery B, Bockaert J, Parmentier M-L, Grau Y. 2009. Plant insecticide L-canavanine repels *Drosophila* via the insect orphan GPCR DmX. *PLoS Biol.* 7:e1000147.

Moreira AS, Horgan FG, Murray TE, Kakouli-Duarte T. 2015. Population genetic structure of *Bombus terrestris* in Europe: Isolation and genetic differentiation of Irish and British populations. *Mol. Ecol.* 24:3257–3268.

Ollerton J, Erenler H, Edwards M, Crockett R. 2014. Pollinator declines. Extinctions of aculeate pollinators in Britain and the role of large-scale agricultural changes. *Science* 346:1360–1362.

Papafotiou G, Oehler S, Savakis C, Bourtzis K. 2011. Regulation of *Wolbachia* ankyrin domain encoding genes in *Drosophila* gonads. *Res. Microbiol.* 162:764–772.

Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A, Podsiadlowski L, Petersen M, Lanfear R, et al. 2017. Evolutionary history of the Hymenoptera. *Curr. Biol.* 27:1013–1018.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* 31:1929–1936.

Puinean AM, Lansdell SJ, Collins T, Bielza P, Millar NS. 2013. A nicotinic acetylcholine receptor transmembrane point mutation (G275E) associated with resistance to spinosad in *Frankliniella occidentalis*. *J. Neurochem.* 124:590–601.

Pracana R, Burns R, Hammond RL, Haller BC, Wurm Y. 2021. Individual-based modeling of genome evolution in haplodiploid organisms. *bioRxiv*. 2021.10.25.465450.

Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, Elsik CG, Gadau J, Grimmelikhuijzen CJP, Hasselmann M, Lozier JD, et al. 2015. The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol.* 16:76.

Schmid-Hempel P, Schmid-Hempel R, Brunner PC, Seeman OD, Allen GR. 2007. Invasion success of the bumblebee, Bombus terrestris, despite a drastic genetic bottleneck. *Heredity* 99:414–422.

Schmidt JM, Battlay P, Gledhill-Smith RS, Good RT, Lumb C, Fournier-Level A, Robin C. 2017. Insights into DDT Resistance from the *Drosophila melanogaster* Genetic Reference Panel. *Genetics* 207:1181–1193.

Shimada S, Kamiya M, Shigetou S, Tomiyama K, Komori Y, Magara L, Ihara M, Matsuda K. 2020. The mechanism of loop C-neonicotinoid interactions at insect nicotinic acetylcholine receptor α1 subunit predicts resistance emergence in pests. *Sci. Rep.* 10:7529.

Supple MA, Shapiro B. 2018. Conservation of biodiversity in the genomics era. *Genome Biol.* 19:131.

Szpiech ZA, Hernandez RD. 2014. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* 31:2824–2827.

Torres-Vazquez J, Warrior R, Arora K. 2000. *schnurri* is required for dpp-dependent patterning of the *Drosophila* wing. *Dev. Biol.* 227:388–402.

Vanbergen AJ. 2013. Threats to an ecosystem service: pressures on pollinators. *Front. Ecol. Environ.* 11:251–259.

Wallberg A, Han F, Wellhagen G, Dahle B, Kawata M, Haddad N, Simões ZLP, Allsopp MH, Kandemir I, De la Rúa P, et al. 2014. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nat. Genet.* 46:1081–1088.

Wolf S, Toev T, Moritz RLV, Moritz RFA. 2012. Spatial and temporal dynamics of the male effective population size in bumblebees (Hymenoptera: Apidae). *Popul. Ecol.* 54:115–124.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28:3326–3328.