# Genomes of three invasive social wasps

Thomas W.R. Harrop\*,1, Joseph Guhlin\*,1, Gemma M. McLaughlin\*, Elizabeth Permina\*, Peter Stockwell†, Josh Gilligan\*, Marissa F. Le Lec\*, Monica A. M. Gruber‡, Oliver Quinn‡, Mackenzie Lovegrove\*, Elizabeth J. Duncan§, Emily J. Remnant\*\*, Erin E. Wilson-Rankin††, Jessica Purcell††, Philip J. Lester‡, Peter K. Dearden\*

\*, †, ‡, §

1 These authors contributed equally to the work.

\* Genomics Aotearoa and Department of Biochemistry, University of Otago, Dunedin 9054, Aotearoa-New Zealand.

† Pathology Department, University of Otago, Dunedin, Aotearoa-New Zealand.

‡ Centre for Biodiversity and Restoration Ecology, School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington, New Zealand.

§ School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, United Kingdom.

\*\* School of Life and Environmental Sciences, University of Sydney, Sydney NSW 2006, Australia.

†† Department of Entomology, University of California, Riverside, CA 92521, USA.

**reference numbers for data available in public repositories**

## 

**Running head**: Genomes of invasive social wasps

**Keywords**: Hymenoptera, Genomics, Venom, Immune system, Development, Eusociality, Gene drive, Pesticide resistance, Gene silencing, Viruses.

**Correspondence**:

Peter K. Dearden   
PO Box 56   
Dunedin 9054   
New Zealand  
  
+64 3 479 7863   
peter.dearden@otago.ac.nz

## Abstract

**Background**: Social wasps of the genus *Vespula* have spread to nearly all landmasses worldwide and have become significant pests, affecting economies and biodiversity and causing problems for humans due to venomous stings. Comprehensive genome assemblies and annotations for these species are required to develop the next generation of control strategies and monitor existing chemical control.

**Results**: We sequenced and annotated the genomes of the common wasp (*Vespula vuglaris*), German wasp (*Vespula germanica*), and the western yellowjacket (*Vespula pensylvanica*). Our chromosome-level *Vespula* assemblies each contain 176–179 Mb of total sequence assembled into 25 scaffolds, with 10–200 unanchored scaffolds, and 16,566 – 18,948 genes. We analysed gene sets relevant to the applied management of invasive populations, including genes associated with spermatogenesis and development, pesticide resistance, olfactory receptors, immunity and viruses and venom.

**Conclusions**: These genomes provide evidence for active DNA methylation as well as tandem duplications of venom genes. The lack of unique target sites and the relatively rich complement of cytochrome P450 genes may limit the success of traditional chemical control of *Vespula* species. We have identified targets for next-generation pest control methods, including gene drives and gene silencing.

## Background

Social wasps (Hymenoptera: Vespidae) are remarkable for many reasons, but especially because their highly eusocial lifestyle appears to have evolved independently of other eusocial Hymenoptera [1-3]. This lifestyle, which is characterized by overlapping generations of adults living together, cooperative care of brood and reproductive division of labor [4], along with their foraging flexibility and predatory ability [5] play major roles in their ecological success [6]. Colonies of common wasps (*Vespula vulgaris*) can contain up to 230,000 workers, while nests of the western yellowjacket (*Vespula pensylvanica*) containing up to half a million individuals have been observed [7]. Vespid wasps can play beneficial roles in their communities. For example, in some situations, they may be more effective pollinators than honeybees or bumblebees [8]. The ecological success of Vespid wasps can also be problematic. Invasive Vespid species (Figure 1; [7]) have major impacts on ecosystems because of their large colony sizes, reproductive capacity and flexible predation. Single wasp colonies harvest up to 100 kg of insect prey each year [9]. In New Zealand’s native beech forests, invasive Vespid wasp populations can reach up to 40 nests per hectare [10] and have a biomass similar to, or greater than, the combined biomasses of birds and mammals [11]. With such extreme wasp densities and their opportunistic foraging strategies [5], populations of invertebrate prey such as caterpillars or spiders are unlikely to survive predation [12, 13].

Apart from their ecological impacts, the combined venom, aggressive behaviour, and abundance of these wasps frequently result in human hospitalisation or deaths around the world. Anaphylaxis from stinging insects has been estimated to occur in ~3% of adults and can be fatal even on the first sting [14]. Despite the fearsome reputation of snakes and spiders, wasps and bees are the leading cause of hospital admission due to venom in Australia [15], and hospital admissions due to venom anaphylaxis are increasing in the UK at a rate of 11.5% per year [16]. Vespid and honeybee (*Apis mellifera*) stings are prevalent in Northern and Central Europe. In the Mediterranean region, more people are stung by *Vespula* and *Polistes* spp. than by honey bees [17]. Venom immunotherapy can be an effective treatment for systemic allergic reaction to wasp venom but relies on venom extracted directly from insects [18]. Knowledge of genes associated with venom production in *Vespula* could enable the use of recombinant technologies to produce large quantities of venom for medical purposes (e.g., [16]).

Social wasps are significant pest species because of their impacts on biodiversity and human health [7]. Current control methods are limited, with pesticides containing Fipronil being the most common and widespread chemical control method [19]. The use of this neurotoxic pesticide over large areas, and over consecutive years, inevitably selects for resistance, and Fipronil resistance has been observed in other economically important insect pests ([20]; reviewed in [21]). Next-generation pest control technologies, including gene drives, have been proposed as part of the solution for controlling or eradicating invasive social wasps [22]. Targets for genetic modification include developmental genes associated with wasp fitness or fecundity [7, 22]. Gene drives that have immune system targets have been proposed and developed in the laboratory for other pests such as mosquitoes (e.g., [23, 24]). However, detailed knowledge of *Vespula* genomes is required to ensure that targeted control of Vespids does not affect other beneficial Hymenoptera, such as honeybees or biocontrol agents.

We report the genomes of three *Vespula* species that are invasive pests in their introduced ranges (Figure 1; [7]). We undertake manual and computational annotation and phylogenomic analyses, with emphasis on sets of genes relevant to applied management of invasive populations, including genes associated with spermatogenesis and development, pesticide resistance, olfactory receptors, immunity and viruses, and venom. These resources will be useful for developing next-generation, genetic control strategies, and for monitoring resistance to current chemical control.

## Materials and methods

**Genome assembly and scaffolding:** The *V. pensylvanica* assembly was produced by Dovetail Genomics (https://dovetailgenomics.com/), starting with an Illumina library generated from a single haploid male to produce 115.5 Gb of 150bp paired-end reads. After trimming adapters and removing bases or truncating reads with low quality scores, the draft genome was assembled with Meraculous 2.2.6 using a kmer size of 55, which produced the best fit with a constrained heterozygous model [127, 128]. Chromosome-scale scaffolds were generated using Chicago and Dovetail Hi-C data from diploid females, and implemented using Dovetail’s proprietary HiRise software.

For *V. germanica* and *V. vulgaris*, we used libraries generated from a single, haploid male to produce 17.1 Gb of 150b paired-end reads and 27.3 Gb of 125b paired-end reads, respectively. After trimming adaptor sequences, removing contaminants and verifying pairing with BBmap 38.00 (https://sourceforge.net/projects/bbmap/), we assembled draft genomes with Meraculous 2.2.6 [127, 128]. We repeated assembly with a range of parameters, and used BUSCO analysis [27], assembly size and contiguity statistics to choose the best set of parameters for each dataset., Code for running and assessing the assemblies is hosted at https://github.com/tomharrop/vger-illumina and https://github.com/tomharrop/vvul-illumina. Draft genomes were scaffolded by Phase Genomics using Hi-C data generated from single, haploid drones. Chromatin conformation capture data was generated using the Proximo Hi-C platform, which is a commercially available version of the Hi-C protocol [129]. These data were used to generate chromosome-scale scaffolds from the corrected assembly, as described in [130].

**Genome curation**: Command-line arguments and scripts can be found at ([https://github.com/jguhlin/vespula\_paper)[https://github.com/jguhlin/vespula\_paper](https://github.com/jguhlin/vespula_paper)%5Bhttps://github.com/jguhlin/vespula\_paper)]. Assembled genomes were cleared of contamination by removing contigs which had BLAST taxonomy results which did not include *Polistes*, *Vespula*, or the word "wasp." Contigs without BLAST results were kept if they contained predicted genes found in a hymenoptera orthogroup from an initial gene prediction. Remaining contigs were only kept if they fell within 2 standard deviations of mean GC% of our kept contigs.  
The largest 25 chromosomes of *V. pensylvanica* were renamed to chromosomes and ordered according to size.  
*V. vulgaris* and *V. germanica* were aligned to the *V. pensylvanica* genome using D-GENIES, which inserts contigs into syntenic locations, flanked by 100 N's, to assign chromosome names to those most related to those in *V. pensylvanica* [131]. Scaffolds were numbered with four digits in order of size.

**Repeat Masking.** Repeats were identified using RepeatModeler and RepeatMasker via the funannotate pipeline [132-134].

**Gene Prediction.** We performed gene prediction by using the Funannotate pipeline v1.6.0, manual annotation, and extrinsic protein evidence [132]. For *V. vulgaris* we also used RNA-seq as additional evidence. Specifically, RNA-seq reads for *V. vulgaris* from queen, worker, and larvae were trimmed with sickle, then aligned to our assembly using STAR to identify splice junctions, and then re-aligned , as specified in the STAR manual [135]. Gene predictions were performed on the assembly using funannotate predict with the RNA alignments and extrinsic protein evidence from public genomes of all *Vespula* proteins from NCBI, *Apis mellifera, Nasonia vitripennis*, and the UniProt SWISS-PROT database [41, 53, 136, 137]. Initial predictions were run through GeneValidator against the UniProt SWISS-PROT database and genes whose protein predictions with scores > 90 were kept and used to train Augustus via the optimize\_augustus command line [138]. The prediction step was re-run as before except with the trained vvulg AUGUSTUS species definition. This process was repeated for *V. pensylvanica* and *V. germanica* using *V. vulgaris* species definition as the AUGUSTUS species in the first round, generating a species-specific configuration in the following round.

**Manual Curation.** Genes were manually curated in WebApollo [139]. These manual annotations took precedence over any intersecting gene predictions. Manual annotation was performed on *V. vulgaris* and lifted over to *V. germanica* and *V. pensylvanica* where possible.

**Gene Family Specific Predictions.** Gene-family specific predictions were enhanced via AUGUSTUS-PPX using a custom pipeline for the LGIC and Olfactory families [28]. Protein sequences of interest from external sources were clustered based on bitscore using BLAST+ and MCL [140]. Protein profiles were generated via AUGUSTUS tools msa2prfl.pl. Assemblies were searched with fastBlockSearch and gene prediction was performed on matching regions with an additional flanking sequence of 1kbp. These predictions took precedent over any intersecting genes.

**Annotation.** Further annotation was performed with funannotate using InterProScan 5.32-71.0 and genes were renamed using custom scripts [141]. Protein predictions were compared with GeneValidator to both our Hymenoptera + *Drosophila* Protein Set and UniProt-SwissProt to generate GV scores and statistics. Proteins were compared with the publicly available genomes from Hymenoptera base [30] using OrthoFinder [29].

**Identification of novel virus sequences:** Total RNA transcriptome data was generated as described in [103]. RNA sequencing reads were assembled de novo using Trinity [142]. Assembled contigs were compared to reference protein sequences of all previously characterised viruses downloaded from GenBank using BLASTx, with an e-value of 1-E5, to maximise sensitivity while minimizing false-positive results [143, 144]. Resulting virus-like contigs were then BLASTed to a non-redundant database to remove non-viral hits, such as host contigs that show similarity to viral sequences.

Novel virus protein sequences were aligned to the current databases of homologous viral proteins using Muscle [145]. Alignments were viewed in Geneious [146] and manually trimmed to remove large gaps and non-conserved regions. Maximum likelihood phylogenetic trees of the novel *Vespula* Luteo-like virus and the *Vespula* Moku-like variant were inferred using PhyML.

Nucleotide alignments of the *Vespula* Luteo-like virus variants observed in each of the 6 colonies were performed using Muscle. A phylogenetic tree was generated using maximum likelihood in MEGA [147], with the TN93+G substitution model, which had the optimal BIC score, and 100 bootstrap replicates.

**CpG[o/e] calculation and phylogenetic analyses:** Nucleotide and dinucleotide content of CDS sequences were calculated using a custom perl script (https://github.com/ejduncan/CpGoe). The number of components in these distributions was estimated using mclust model-based clustering [148]. The best fitting model was identified among several non-nested models using Bayesian information criteria (BIC).

Orthologs were identified by reciprocal Blast, sequences were aligned using Clustal Omega (v 1.2.0) [149], GBlocks v0.91b was used to eliminate poorly aligned positions [150]. Phylogeny was reconstructed using MrBayes [151] using the Jones model of amino acid substitution. The Monte Carlo Markov Chain was run with four chains until convergence (1,000,000 generations) with trees sampled every 1,000 generations, the first 25% of trees were discarded as ‘burn-in’.

## Results and discussion

### Genome assemblies and gene content

We used a combination of second-generation sequencing and Hi-C scaffolding to assemble draft genomes for *V. germanica*, *V. pensylvanica,* and *V. vulgaris*. The genomes each contain 176–179 Mb of total sequence assembled into 25 superscaffolds (Figure 2A; N50 length 8.30–8.53 Mb), which likely represent the 25 chromosomes observed in *Vespula* karyotypes ([25]). Each draft genome also contains 10–200 unanchored scaffolds (N50 lengths 1.77–2.28 kb; Supplementary Data 1). These genomes are similar in size to the genome of the closely related European paper wasp *Polistes dominula* [26]. We ordered and named scaffolds in the *Vespula* assemblies based on scaffold length in *V. pensylvanica*. The three genomes are highly syntenic, with evidence of some structural rearrangements (Fig 2B, S1). Repeat masking masked 17.86%, 18.75% and 18.71% the *V. vulgaris, V. pensylvanica,* and *V. germanica* genomes, respectively. We predicted 16,566, 17,644, and 18,948 genes for *V. vulgaris*, *V. germanica*, and *V. pensylvanica*, respectively. We found between 92.2% and 96.0% of expected single-copy orthologs using BUSCO with the Hymenoptera lineage dataset [27]. The contiguity of our assemblies and completeness of our annotations indicates that the combination of short-read sequencing and Hi-C scaffolding on haploid material is a very effective strategy for assembling high-quality Hymenopteran genomes.

To predict divergence time of *Vespula* species from *Polistes*, we reconstructed a phylogeny using other Hymenopteran genomes. Based on a previously published estimate for Hymenoptera [1], the divergence time of Vespid wasps from their last common ancestor with *P. dominula* is estimated to be 51 million years (34-71 mya 95% CI). Scaling our ultrametric phylogenetic tree to this estimate places *V. vulgaris* split from *V. pensylvanica* and *V. germanica* at ~6 mya, and *V. pensylvanica* split from *V. germanica* 4.5 mya (Supplementary Figure 2).

We manually curated 361 gene models in *V. vulgaris* and used these curations to improve automated prediction steps for the other two species. During manual curation, we focused on a range of gene sets relevant to the evolution and applied management of invasive *Vespula* spp., including olfactory receptors, pesticide resistance, immunity and viruses, venom, and spermatogenesis and development. We also used targeted prediction to identify ligand-gated ion channel (LGIC), olfactory receptor, and spermatogenesis genes using Augustus protein profiles [28]. To investigate relationships between *Vespula* genes, we clustered predicted proteins into orthogroups with predicted proteins from the Hymenoptera Genome Database, using *Drosophila melanogaster* as the outgroup [29-31]. Each orthogroup contains a set of genes putatively descended from a single gene in the last common ancestor of the species represented in the orthogroup. Between 82.6% and 88.4% of our predicted *Vespula* proteins belonged to orthogroups. *V. vulgaris* shares 12,560 and 12,084 orthogroups with *V. pensylvanica* and *V. germanica*, respectively, and *V. pensylvanica* shares 13,209 orthogroups with *V. germanica* (Figure 2C)*.* Orthogroups including other Hymenoptera species allowed us to predict the core- and pan-genomes for Hymenoptera (Figure 2D). This analysis suggests that Hymenoptera have a closed pan-genome, because as we include more genomes the rate of discovery of new orthogroups decreases. We also observed more orthogroups in our Vespid genomes than in other Hymenoptera, which could indicate over-prediction resulting from our annotation. However, most of the genes we annotated were in shared orthogroups across the three *Vespula* species, with only 11 genes in species-specific orthogroups (0.023%), suggesting that either Vespid wasps have more genes than other Hymenoptera, or gene annotation in other Hymenoptera is incomplete. This could be resolved by re-annotation of other Hymenopteran genes using a comparative approach.

### Spermatogenesis, sex determination and development

Hymenopteran insects, including Vespid wasps, use haplodiploidy as a sex determination system. Diploid females develop from fertilized eggs, while haploid males develop from unfertilized eggs. We manually annotated nine genes associated with sperm development in Hymenoptera, including an MCM8-like protein and the genes *Sdic* and *oncus*, which are expressed in testis in the jewel wasp *Nasonia vitripennis* [26, 32, 33], and *boule*, which is linked to spermatid differentiation in the male germ line including in the hymenopteran sawfly, *Athalia rosae* [33, 34]. Population suppression by a gene drive targeting the sex determination factor *doublesex* has been demonstrated in mosquitoes [35]. In honeybees, sex is determined by a gene network triggered by embryo ploidy (reviewed in [36]). The *complementary sex determiner* (*Csd*) gene is at the top of this cascade [37]. Two different *Csd* alleles are required for CSD protein to splice *feminizer* (*fem*) RNA into an active form ([37]; reviewed in [38]). An active feminizer protein acts with transformer-2 to alter splicing of mRNA for a transcription factor, *doublesex*, produce a female-specific isoform, leading to the activation of female-specifying genes [39]. Inactive CSD leads to a male-specific splice variant of the doublesex transcription factor [40]. We did not find any sequences similar to honeybee *Csd* in our three *Vespula* genomes, but each genome contains a copy of the *fem* gene. All species have *transformer-2* and *doublesex* orthologs, suggesting that the end of the sex determination cascade is conserved. It is unclear how the ploidy of wasp embryos is detected, or if this triggers alternative splicing of *doublesex* mRNA. Our annotation of the components of the pathway will allow the mechanism of sex determination in Vespid wasps to be studied, potentially leading to targets for genetic control.

Axis formation mechanisms present another potential target for control, as data from genome sequences [41, 42] and developmental studies [43, 44] show that early-acting developmental pathways evolve rapidly. Anterior-posterior patterning in Hymenoptera has been described in honeybees [44] and *N. vitripennis* [45]. In these species, anterior patterning genes include *orthodenticle* and *hunchback*, which are present in the genomes of all three *Vespula* spp. Posterior development is specified differently between honeybees and *N. vitripennis*. In *N. vitripennis*, the oosome (a germ-line specifying organelle) carries posterior specifying RNAs to the posterior of the oocyte [46]. These RNAs are nucleated by the *oskar* protein [46], which is absent from the honeybee genome [42]. In honeybees, the posterior of the embryo appears to be specified by the relocation of maternal RNAs for *caudal* and *tailless* [47, 48]. We found orthologs of *oskar* in all three *Vespula* genomes, implying that these wasps may specify posterior fate and germ cells using a similar mechanism to *N. vitripennis*, involving either pole-plasm or an oosome. Embryological studies on Vespid wasps would clarify the role of these genes in patterning and their potential use as targets for control.

Terminal patterning mechanisms differ between hymenopteran species and other insects. In *Nasonia* and honeybees, terminal patterning seems to be linked to anterior-posterior patterning, albeit by different mechanisms [43, 45, 48, 49]. In other holometabolous insects, terminal patterning involves the activation of the torso cell signalling pathway, via a ligand, trunk, which is localised by torso-like (reviewed in [50][51]). Torso signalling is also critical for moulting, as prothoracicotropic hormone (PTTH) also binds torso to trigger ecdysone release [52]. In honeybees, none of *trunk*, *torso,* and *ptth* are present in the genome [41]. *torso* and *ptth* genes, but not *trunk,* are present in *Nasonia* [53]. A *torso-like* gene is present in all these species [54]. We found *torso* and *ptth* genes, but not *trunk*, in all three *Vespula* genomes. This implies that the mechanism of terminal patterning in *Vespula* is different to the mechanisms in *Tribolium* and *Drosophila*, but that the role of PTTH in triggering moulting may be conserved. Targeting the PTTH-torso interaction may interfere with moulting in Vespids in a way that does not affect bees.

Later acting developmental genes, such as those involved in segmentation are well conserved between Vespid wasps and Honeybees, and some at least, are expressed in similar patterns (for example see Figure 3).

### Genes associated with chemical control

Primary chemical control of *Vespula* populations is through the use of baits containing a low concentration of the phenylpyrazole insecticide Fipronil, which attracts foraging worker wasps [7, 19]. The workers return to the nest and distribute the bait. These toxic baits can effectively control wasp numbers in the field [19]. Fipronil resistance has been identified in economically important pests including the diamondback moth (*Plutella xylostella*; [55]), planthoppers [56-59] , cattle ticks [60] and rice leaf beetle (*Oulema oryzae*, [61]; reviewed in [21]), but resistance has not yet been reported in *Vespula* spp or other social wasps. Fipronil resistance can arise due to mutations in one of the target-site genes, the GABA receptor *Resistant to dieldrin* (*Rdl*). *Rdl* resistance mutations include the Ala301 mutation that confers high resistance to cyclodiene insecticides [62], other amino acid substitutions associated with phenylpyrazole resistance [63, 64], and copy-number expansions of the *Rdl* locus in some species [65-67]. Our annotation of *Rdl* in *Vespula* genomes did not suggest the presence of the classical Ala301 mutation that confers high resistance to cyclodienes or copy-number variation, but a comprehensive population survey would be required to understand variation and resistance potential at the *Rdl* locus in Vespid populations.

While *Rdl* is the only known target site of Fipronil, phenylpyrazoles are likely to bind to and inhibit other receptors [68]. *Rdl* is a member of the cys-loop Ligand-gated ion channels (LGIC), a family of receptor subunits that predominantly mediate inhibitory and excitatory neurotransmission [62]. LGIC family subunits are common targets of insecticides. For example, neonicotinoid and spinosyn insecticide classes target nicotinic acetylcholine receptor subunits [69, 70]. Insect LGICs are highly conserved, but there is some copy-number variation in more divergent branches [66, 71, 72], which may provide unique and effective pest-specific receptors to target with novel selective chemicals. We examined receptor diversity of *Vespula* LGICs compared to other Hymenopteran genomes (*Apis mellifera, Bombus impatiens, Nasonia vitripennis*) and *Drosophila melanogaster* (Supplementary figure 3). *Vespula* receptors are highly conserved, with one-to-one orthology in *Apis* and *Bombus*. This suggests that chemicals targeting *Vespula* LGICs will also affect bees.

Metabolism of insecticides by cytochrome P450 enzymes (P450s) is a major resistance mechanism [73-75], and variation in P450 genes contributes to phenotypic variation in xenobiotic metabolism, which may aid adaptation to novel toxins including insecticides [76]. Although P450-mediated Fipronil resistance has not been reported in arthropods, P450s can metabolise Fipronil *in vitro* [77]. The number of P450 genes in insect genomes (*CYPomes*) varies from around 36–180 genes [75]. CYPomes typically expand by successive duplications in one or more subfamily [76]. To produce accurate CYPomes for *Vespula*, we manually annotated 60 P450 genes in *V. vulgaris* and used these to improve prediction of P450s in all three genomes. We found 61, 59, and 62 P450s in *V. germanica*, *V. pensylvanica* and *V. vulgaris*, respectively (Supplementary data 2). Generally, CYPome size of *Vespula* is intermediate between the small CYPome size of bees and the large CYPome size in ants (Supplementary figure 4A; [78]).Our orthogroup analysis identified an expansion in *Vespula* CYP3 clan genes related to the *Drosophila* CYP9 family (Supplementary figure 4B). We also found an expansion of CYP4 genes, which are homologous to *D. melanogaster* *Cyp4e1* (Supplementary figure 4C). *Cyp4e1* is associated with resistance to permethrin [79]. In contrast, we only found one homolog of the 7 mitochondrial CYP12 paralogs from *Drosophila* in each *Vespula* genome (Supplementary figure 4D). The *D. melanogaster* CYP12 gene *Cyp12d1* is involved in resistance to DDT, dicyclanil, malathion and chlorantraniliprole [80-84], and other CYP12 genes are linked to lufenuron and nitenpyram metabolism [85, 86]. Although the total size of *Vespula* CYPomes is intermediate relative to other Hymenoptera, they contain radial expansions of some families that may encode metabolic activity towards insecticides. CYPomes of naïve insects may encode P450 activity against novel insecticides that they have not evolved to metabolise [84, 87]. Based on the lack of potential *Vespula*-specific insecticide target genes and their rich CYPomes, *Vespula* spp. may be poor candidates for long-term traditional chemical control.

### Olfaction

Olfaction is vital for many aspects of life for insects. In particular it is involved in prey targeting and foraging [88] reviewed in [5]), and sexual behaviour [89][90]. Olfaction in insects functions via families of olfactory receptors (ORs) [91, 92]. OR genes encode g-coupled protein receptors with seven transmembrane domains [91, 92], and can evolve rapidly [93]. Insect ORs act with the conserved Odorant receptor co-receptor (Orco) protein, which has a similar sequence and structure to other ORs [94]. Olfaction is particularly important for eusocial species [95, 96], as it is an important way in which information is shared in the colony, such as repressive signals or pheromones produced by queen honeybees to repress worker reproduction. Orco mutants in ant species lose the ability to follow pheromone trails, lack social organization, and have a reduction in fitness [97, 98]. We found more OR genes in *Vespula* genomes than in the genome of *P. dominula*. *V. germanica* has 104 predicted olfactory receptor genes, *V. pensylvanica*, 92 and *V. vulgaris* 106. *P. dominula* has 78. In contrast, honeybee [99] and *Nasonia* [100] genomes encode more OR genes (170 and 301, respectively). *Vespula* OR genes cluster into 20 orthogroups. The co-receptor *Orco* is present in all genomes and forms a stable orthogroup (orthogroup 6719) [94, 101]. In contrast, there are significant expansions of particular OR orthogroups in the Vespid wasps, and these differ from the groups expanded in *Nasonia* [100] and honeybee [99]. Orthogroup members are arranged in expanded tandem arrays on chromosome 3 (orthogroup 49; Figure 4), chromosome 13 (orthogroup 908), and chromosome 25 (orthogroup 173) of *Vespula* genomes. In these clusters, numbers of genes vary between species implying that duplications and deletions are recent and ongoing (Figure 4). Variation in olfactory receptors between wasp species, and between these wasps and other hymenoptera may indicate species-specific olfactory biology. These may be key to understanding the social behavior and pheromone signalling systems present within these species. This could also be targeted for control. De-orphaning these receptors, especially those that are Vespid-wasp specific, may provide opportunities to develop novel chemical lures or chemical disruptors of behaviour [102]-such as those associated with the maintenance of social structure. This loss of group cohesion, or the luring of wasps could be used in control strategies.

### Insect viruses as novel wasp control measures

Control of *Vespula* wasp populations may be achievable through the discovery of novel viruses that infect them. We searched for viral sequences in the total RNA transcriptomes of larvae, adult workers and virgin queens from *V. vulgaris,* sampled from six mature nests in autumn. (Samples were taken from mature nests in the native range of Belgium and the invaded range in New Zealand [103]). We found evidence for 70 novel virus species and 8 previously identified virus species across all 18 transcriptomes (supplementary data 3, supplementary figures 5-7). Assembled viral contigs ranged in size from 1.4-12 kb, and the viral component of each library made up 0.04-1% of total RNA. The previously uncharacterised viral sequences have between 23-92% amino acid similarity compared to their most similar published sequence (supplementary data 3). The majority of virus sequences were present in one sample only (61 virus species; 78%), but 17 viruses were found in more than one sample. Of these, 10 were found in more than one colony, with four found in both the native and invasive range (supplementary data 3). We found one virus, a Luteo-like virus, present in all three life stages in all six colonies (Supplementary figure 5). This 3110 base viral contig was fully assembled and at relatively high abundance in all samples (0.02-0.43% of total RNA, at 46 to 7378-fold average coverage), with up to 10% nucleotide variation between samples (Supplementary figure 6). Luteoviruses are common plant pathogens, but a related group of Luteo-like viruses have been identified in a range of arthropods including mosquitoes and ticks [104-106]. The abundance and prevalence of this virus implies it is a genuine *Vespula*-infecting virus. We also identified a novel Iflavirus with 71% amino-acid identity to the known *Vespula*-infecting Moku virus, indicating that there are multiple wasp infecting Moku-like viruses (Supplementary figure 7).

More viral sequences were present in larvae (66 virus sequences, compared to 23 in workers and 4 in queens). Wasps are carnivorous and have polyphagous diet consisting of a diverse array of arthropods [107]. Many of the novel virus contigs are most similar to viruses identified in arthropods and belong to common arthropod virus families (Supplementary data 3; [105]), and may reflect the viral content of prey organisms present in the gut of feeding *Vespula* larvae. Our results indicate that viral metagenomics may inform us about the types of prey eaten by *Vespula*. Many insect viruses are multi-host pathogens [108-110], suggesting that wasps could be susceptible to viral infection sourced from their diet, a route that may provide a novel method of wasp control. We have observed viral replication or fitness effects of viruses in *Vespula*, suggesting these wasps are parasitized by viruses [108, 111, 112]. *Vespula* species may require enhanced anti-viral or other immune mechanisms to combat increased exposure to dietary pathogens.

### Venom

Different *Vespula* spp. have different venom compositions [113]. The major allergens in these venoms are phospholipase A1, hyaluronidase, and antigen 5 [113, 114]. Venom is used in wasp or colony defence, and typically not for hunting or subduing prey. We identified genes that encode these venom components in all three of our wasp genomes. Phospholipase A1 is found in three tandem copies in the *P. dominula* and *V. germanica* genomes (chromosome 9 in *Vespula*), and one copy in each of the *V. pensylvanica* and *V. vulgaris* genomes. The phylogenetic placement of these duplicates in *Polistes* and *V. germanica* (Figure 5) implies that these are independent amplifications. The hyluronidase gene is duplicated in our three Vespid genomes, but not *P. dominula*. These are tandem duplications that appear to have been present in the last common ancestor of the three *Vespula* species. *P. dominula*, *V. germanica* and *V. pensylvanica* also have two Antigen 5 genes, but these duplications appear ancient (Figure 5) before the common ancestor of hymenoptera. In *Vespula* species, one copy is on chromosome 6 and one is on chromosome 7. In *V. vulgaris*, the chromosome 6 copy is absent. That many of these venom genes are duplicated in Vespids in no surprise, given the importance of venom to their life cycle. These variant copies may be important in generating venom diversity and need to be considered in the generation of antivenoms. These different venom genes identified here will enable advances in venom immunotherapy, such as the use of recombinant technologies to produce venom for medical purposes.

### DNA Methylation

DNA methylation has been functionally linked to caste specification in honeybees [117] and ants [118] and division of labour in honeybees [119, 120]. These studies have provided support for the hypothesis that DNA methylation is integral for aspects of eusociality (reviewed in [121]) although recent studies have found no consistent link [122][123]. It is thought that *de novo* DNA methylation is catalysed by the Dnmt3 enzyme and that the maintenance methyltransferase Dnmt1 is responsible for maintaining these marks during cell division. The only other Vespid genome sequences available are *Polistes canadensis* and *Polistes dominula*. Neither of these genomes encode a Dnmt3 gene, yet do have measurable levels of DNA methylation, suggesting that Dnmt1 may be able to methylate DNA *de novo* in *Polistes* [26] as has been shown for some other species [122].

All three *Vespula* genomes encode an ortholog of Dnmt3, indicating the lack of this gene in Polistes is due to gene loss following the divergence from the *Vespula* lineage approximately 50 million years ago [1]. *Vespula* genomes all contain a single ortholog of *Dnmt1* (Supplementary figure 8)

Over evolutionary time, methylated cytosine residues in DNA become deaminated to thymines. This results in a gradual reduction of CpG dinucleotides in germline methylated genes, while regions with limited germline methylation retain a higher CpG content [124]. Analysis of the observed versus expected ratio of CpG dinucleotides (CpG[o/e]) results in a unimodal distribution in genomes of insects with no DNA methylation such as *Drosophila melanogaster* (Fig. 6a) and a bimodal distribution for insects with active DNA methylation systems such as the honeybee *Apis mellifera* (Fig. 6b) [125]. Measurement of the CpG[o/e] predicts extant DNA methylation levels measured experimentally [122]. Our analysis of CpG[o/e] distributions in *Polistes* (Fig. 6c,d) indicates the distribution is best described as three overlapping normal distributions, and the presence of a low-CpG[o/e] peak is consistent with the low-levels of DNA methylation that have been detected experimentally, and may also reflect the recent loss of the Dnmt3 gene in this lineage (~50 mya) as CpG[o/e] reflects historical DNA methylation. Consistent with our findings that the *Vespula* genomes encode a full-complement of genes thought to be required for DNA methylation we also find a low- CpG[o/e] peak consistent with active DNA methylation in these species (Fig. 6e-g).

### Conclusions

We have produced chromosome-level genome assemblies for three invasive social wasps in the *Vespula* genus. Our approach of short-read sequencing and Hi-C scaffolding using haploid material allowed us to produce assemblies that exceed the genome quality targets suggested by the i5k insect genome sequencing initiative (scaffold n50 length > 300 kb; [126]). Using manual curation and computational prediction, we have identified genes that may encode specific biology suitable for targeting with next-generation control technologies, genes that may be affected by selection by current chemical controls.

These are the first three genomes from this branch of the Aculeata subclade, which will be useful in phylogenetic comparisons of the remarkable life history characteristics of hymenoptera. In particular, these genomes will be valuable for understanding the evolution of eusociality, which has appeared twice in Vespid wasps [1-3] independently of other hymenoptera. Comparing the genomes of Vespid wasps, which are highly eusocial, with the closely related paper wasps, which are primitively eusocial, may also help our understanding of how evolution elaborates mechanisms of colonial living.

Vespid wasps are significant invasive pests in many parts of the world. These genomes will be of major importance for applied management of *Vespula*, in programs using both existing control methods and for next-generation applications. Our assemblies will provide species-specific targets for novel control methods, such as RNA interference, gene drives and the deployment of damaging viruses. They will also be essential for monitoring the effects of next-generation control methods and measuring genomic variation across natural populations.

#### Figure legends

**Figure 1**: *Vespula* species are major invasive pests. A–C. Examples of (A) *Vespula germanica,* (B) *Vespula pensylvanica* (Jon Sullivan, public domain), and (C) *Vespula vulgaris* (Sid Mosdell, CC BY 2.0 <https://creativecommons.org/licenses/by/2.0>). B. Native and invaded ranges of the sequenced *Vespula* species. C. Rooted phylogenetic species tree of *Vespula* species and other sequenced Hymenoptera, with *Drosophila melanogaster* as an outspecies, derived from 415 Orthogroups using OrthoFinder.

**Figure 2**: A. Assembled chromosome lengths of the three *Vespula* species. Chromosomes were named in order of scaffold length in *V. pensylvanica.* B. Synteny between selected chromosomes of *Vespula* species. Chromosome 1 has minor translocations and inversions but is syntenic overall. Chromosome 2 has a larger inversion between *V. pensylvanica* and *V. germanica* that is not present in *V. vulgaris*. Chromosome 5 has a large translocation and inversion between *V. pensylvanica* and *V. germanica*. Chromosomes lengths are scaled to chromosome 1. C. Shared orthogroups between the three *Vespula* species. D. The pan and core genomes of Hymenoptera. The core genome (the set of orthogroups present in all genomes sampled) decreases as more genomes are sampled, to a final core genome size of 3,106 orthogroups when all 23 Hymenopteran genomes are compared. The pan genome (the total set of orthogroups present in one or more of the sampled genomes) does not continue to grow as more genomes are sampled, indicating a closed Hymenopteran pan genome. We predicted more orthogroups (13,826–14,691 orthogroups) in *Vespula* species than in non-*Vespula* Hymenoptera (median 9204 orthogroups), which resulted in a larger pan genome in comparisons that included *Vespula* species. However, there is also a split in the pan genome size in comparisons that do not include *Vespula,* which highlights the effect of annotation on core and pan genome comparisons. Only genes present in an orthogroup were analysed, and we plotted a random subset of 1,000 comparisons for genome numbers that resulted in more than 1000 comparisons.

**Figure 3**: Common wasp (*Vespula vulgaris*) embryos stained for *wingless* RNA, showing expression in thin stripes of cells in each segment. A) Germband stage embryo wrapped around the yolk or the egg, anterior to the left, ventral view, damaged at the posterior end. B and C) Successively older embryo, oriented with anterior to the left, dorsal up. Scale bars represent 100 microns.

**Figure 4**: Comparison of a tandemly duplicated group of olfactory receptor genes between *V. vulgaris* and *V. germanica* found on chromosome three of Vespid genomes. The differences in phylogenetic history of the genes, and variation in gene numbers in this cluster between species implies ongoing evolution of this cluster after the divergence of these two Vespid wasps.

**Figure 5**: Ultrametric trees of orthogroups encoding Vespid wasp venom components in insect genomes. A) Venom hyaluronidase phylogram showing duplication of this gene in Vespid genomes but not *Polistes*. B) Venom phospholipase phylogram indicating independent duplications in *Polistes* and *V. germanica* from a single gene ancestor in related hymenoptera. C) Antigen 5 phylogram, indicating two genes encode this venom component in Vespid wasps and Polisties, one of which is missing from the V. vulgaris genome.

**Figure 6**: Frequency histogram of CpG[o/e] observed in coding sequences of A)*Drosophila melanogaster*, B) *Apis mellifera*, C) *Polistes canadensis* d)*Polistes dominula*, E) *Vespula germanica*, F) *Vespula vulgaris* and G)*Vespula pensylvanica* . The y-axis depicts the number of genes (Frequency) with CpG[o/e] values given on the x-axis. The distribution of CpG(o/e) in the Vespid wasps is a trimodal distribution, with a low- CpG[o/e] peak consistent with the presence of historical DNA methylation all five Vespid species.

**Supplementary information**

**Suppplementary data 1:** assembly statistics.

**Supplementary data 2**: Cytochrome P450 genes from *Vespula* species.

**Supplementary data 3:** Novel viruses identified in *V. vulgaris* transcriptomes

**Supplementary figure 1:** Synteny plot of 25 chromosomes of our *Vespid* wasps assemblies. Syntenic regions are at least 5kbp.

**Supplementary figure 2:** Ultrametric tree generated from OrthoFinder of our Hymenopteran entries including *Drosophila melanogaster* as an outgroup.

**Supplementary figure 3:** Ligand-Gated Ion Channel (LGIC) receptor subunits of *Vespula*. Maximum likelkihood phylogenetic tree of amino acid sequences (bootstrap values of 100 replicates are shown) indicates that subunit receptor diversity of *Vespula* LGICs is conserved with LGICs present in honey bee and bumblebee genomes (*Apis mellifera* and *Bombus impatiens*) with one-to-one orthology of all subunits .

**Supplementary figure 4:** The CYPomes of *Vespula* species. A. The number of cytochrome P450 genes found in annotated Hymenoptera genomes. After removing multiple isoforms of the same gene, we counted the number of P450 genes from each species that were placed into orthogroups. We were not able to resolve isoforms of the same gene in the annotations for species marked with asterisks. Scale bar shows number of amino acid substitutions per site. B and C. Blooms of *Vespula* CYP3 clan genes related to *Drosophila* CYP9 (B) and CYP4 clan genes related to *Drosophila* *Cyp4e1*. D. *Vespula* CYPomes contain one homolog of the CYP12 family that is expanded in *Drosophila*.

**Supplementary figure 5:** Luteo-like viruses. Maximum likelihood phylogenetic tree of amino acid sequences of Luteo-like viruses. The novel *V. vulgaris* luteo-like virus is indicated in red.

**Supplementary figure 6**: Comparison of the *V. vulgaris* Luteo-like virus nucleotide sequences obtained from the transcriptomes of all six colonies. Sequences from each sample type (queen, worker and larvae) form clades based on colony.

**Supplementary figure7**: Moku-like iflaviruses. Maximum likelihood phylogenetic tree of amino acid sequences of Luteo-like viruses. The novel *Vespula* luteo-like virus is indicated in red.

**Supplementary figure 8**: Bayesian phylogeny of Hymenopteran Dnmt1 protein sequences. Phylogeny indicates that a duplication event of an ancestral Dnmt1 gene gave rise to the two paralogs (Dnmt1a and 1b) observed in Apoidea, this duplication may have occurred in the lineage leading to Apiodea (bees) and Formicidae (ants), with the ants subsequently losing Dnmt1b. Alternatively the duplication may have occurred in the last common ancestor of Apidoea, Formicidae and Vespidae (~190 mya) with Dnmt1a being selectively lost in the Vespid wasps and Dnmt1b being lost in the Formicidae. Posterior probabilities are indicated by the relative size of the symbol at the nodes, all branches were supported by posterior probabilities of >0.5.

### Acknowledgements

The authors would like to thank P.M. Dearden for editing the manuscript. New Zealand Genomics Ltd for generation of the Common wasp genome data. The New Zealand National eScience Infrastructure (NeSI) for computational support. D. Hart and D.J. Champion for I.T. support and B.P. Dearden for critical discussions in the formulation of this work.

### Funding

This project was supported by Genomics Aotearoa (to PKD) and the Biological Heritage National Science Challenge (to PJL) both funded by the Ministry of Business Innovation and Employment (Hīkina Whakatutuki), Government of New Zealand, as well as a US National Science Foundation Grant and UC Riverside Seed Grant to JP and EWR; Dovetail Genomics Matching Funds Grant to JP.

### Availability of data and materials

**[SRA accessions TBC]**

### Authors’ contributions

PJL and PKD conceived and designed the project. All authors aided in obtaining and analysing the genomic data. PJL, PKD, TWRH, JG, EJR and EJD wrote the manuscript draft, and all authors participated in the revision of the final version. All authors read and approved the final manuscript.

### Ethics approval and consent to participate

Common wasps (*V. vulgaris*) were collected under the permit National Authorisation Number 38337-RES from the Department of Conservation in New Zealand. Samples of other wasps were collected from private land where no permit was required. No other ethical approval was required.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

## References

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

2. Hines HM, Hunt JH, O'Connor TK, Gillespie JJ, Cameron SA: **Multigene phylogeny reveals eusociality evolved twice in vespid wasps.** *Proc Natl Acad Sci U S A* 2007, **104:**3295-3299.

3. Piekarski PK, Carpenter JM, Lemmon AR, Moriarty Lemmon E, Sharanowski BJ: **Phylogenomic evidence overturns current conceptions of social evolution in wasps (Vespidae).** *Molecular biology and evolution* 2018, **35:**2097-2109.

4. Michener CD, Brothers DJ: **Were workers of eusocial hymenoptera initially altruistic or oppressed?** *Proc Natl Acad Sci U S A* 1974, **71:**671-674.

5. Richter MR: **Social wasp (Hymenoptera: Vespidae) foraging behavior.** *Annu Rev Entomol* 2000, **45:**121-150.

6. Grimaldi D, Engel MS, Engel MS: *Evolution of the Insects.* Cambridge University Press; 2005.

7. Lester PJ, Beggs JR: **Invasion Success and Management Strategies for Social Vespula Wasps.** *Annu Rev Entomol* 2019, **64:**51-71.

8. Thomson DM: **Effects of long‐term variation in pollinator abundance and diversity on reproduction of a generalist plant.** *Journal of Ecology* 2019, **107:**491-502.

9. Harris R: **Frequency of overwintered *Vespula germanica* (Hymenoptera: Vespidae) colonies in scrubland‐pasture habitat and their impact on prey.** *New Zealand Journal of Zoology* 1996, **23:**11-17.

10. Lester PJ, Haywood J, Archer ME, Shortall CR: **The long‐term population dynamics of common wasps in their native and invaded range.** *Journal of Animal Ecology* 2017, **86:**337-347.

11. Thomas CD, Moller H, Plunkett GM, Harris RJ: **The prevalence of introduced *Vespula vulgaris* wasps in a New Zealand beech forest community.** *New Zealand journal of ecology* 1990, **13:**63-72.

12. Toft RJ, Rees JS: **Reducing predation of orb‐web spiders by controlling common wasps (*Vespula vulgaris*) in a New Zealand beech forest.** *Ecological Entomology* 1998, **23:**90-95.

13. Beggs JR, Rees JS: **Restructuring of Lepidoptera communities by introduced *Vespula* wasps in a New Zealand beech forest.** *Oecologia* 1999, **119:**565-571.

14. Golden DB: **Insect sting anaphylaxis.** *Immunology and allergy clinics of North America* 2007, **27:**261-272.

15. Welton RE, Williams DJ, Liew D: **Injury trends from envenoming in Australia, 2000–2013.** *Internal medicine journal* 2017, **47:**170-176.

16. Turner PJ, Gowland MH, Sharma V, Ierodiakonou D, Harper N, Garcez T, Pumphrey R, Boyle RJ: **Increase in anaphylaxis-related hospitalizations but no increase in fatalities: an analysis of United Kingdom national anaphylaxis data, 1992-2012.** *Journal of Allergy and Clinical Immunology* 2015, **135:**956-963. e951.

17. Biló BM, Rueff F, Mosbech H, Bonifazi F, Oude‐Elberink JNG: **Diagnosis of Hymenoptera venom allergy.** *Allergy* 2005, **60:**1339-1349.

18. Ludman SW, Boyle RJ: **Stinging insect allergy: current perspectives on venom immunotherapy.** *J Asthma Allergy* 2015, **8:**75-86.

19. Edwards E, Toft R, Joice N, Westbrooke I: **The efficacy of Vespex® wasp bait to control *Vespula* species (Hymenoptera: Vespidae) in New Zealand.** *International Journal of Pest Management* 2017, **63:**266-272.

20. Matsumura M, Takeuchi H, Satoh M, Sanada-Morimura S, Otuka A, Watanabe T, Van Thanh D: **Species-specific insecticide resistance to imidacloprid and fipronil in the rice planthoppers *Nilaparvata lugens* and *Sogatella furcifera* in East and South-east Asia.** *Pest Management Science* 2008, **64:**1115-1121.

21. Feyereisen R, Dermauw W, Van Leeuwen T: **Genotype to phenotype, the molecular and physiological dimensions of resistance in arthropods.** *Pesticide Biochemistry and Physiology* 2015, **121:**61-77.

22. Dearden PK, Gemmell NJ, Mercier OR, Lester PJ, Scott MJ, Newcomb RD, Buckley TR, Jacobs JME, Goldson SG, Penman DR: **The potential for the use of gene drives for pest control in New Zealand: a perspective.** *Journal of the Royal Society of New Zealand* 2018, **48:**225-244.

23. Gantz VM, Jasinskiene N, Tatarenkova O, Fazekas A, Macias VM, Bier E, James AA: **Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*.** *Proc Natl Acad Sci U S A* 2015, **112:**E6736-6743.

24. Dong Y, Simoes ML, Marois E, Dimopoulos G: **CRISPR/Cas9 -mediated gene knockout of *Anopheles gambiae* FREP1 suppresses malaria parasite infection.** *PLoS Pathog* 2018, **14:**e1006898.

25. Hoshiba H, Matsuura M, Imai HT: **Karyotype evolution in the social wasps.** *The Japanese Journal of Genetics* 1989, **64:**209-222.

26. Standage DS, Berens AJ, Glastad KM, Severin AJ, Brendel VP, Toth AL: **Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect.** *Molecular ecology* 2016, **25:**1769-1784.

27. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics* 2015, **31:**3210-3212.

28. Keller O, Kollmar M, Stanke M, Waack S: **A novel hybrid gene prediction method employing protein multiple sequence alignments.** *Bioinformatics* 2011, **27:**757-763.

29. Emms DM, Kelly S: **OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences.** *BioRxiv* 2018**:**466201.

30. Elsik CG, Tayal A, Diesh CM, Unni DR, Emery ML, Nguyen HN, Hagen DE: **Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine.** *Nucleic acids research* 2015, **44:**D793-D800.

31. Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R: **The Release 6 reference sequence of the *Drosophila melanogaster* genome.** *Genome research* 2015, **25:**445-458.

32. Parsch J, Meiklejohn CD, Hauschteck-Jungen E, Hunziker P, Hartl DL: **Molecular evolution of the *ocnus* and *janus* genes in the *Drosophila melanogaster* species subgroup.** *Mol Biol Evol* 2001, **18:**801-811.

33. Mikhaylova LM, Boutanaev AM, Nurminsky DI: **Transcriptional regulation by Modulo integrates meiosis and spermatid differentiation in male germ line.** *Proc Natl Acad Sci U S A* 2006, **103:**11975-11980.

34. Sekine K, Furusawa T, Hatakeyama M: **The *boule* gene is essential for spermatogenesis of haploid insect male.** *Dev Biol* 2015, **399:**154-163.

35. Kyrou K, Hammond AM, Galizi R, Kranjc N, Burt A, Beaghton AK, Nolan T, Crisanti A: **A CRISPR–Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes.** *Nature Biotechnology* 2018, **36:**1062-1066.

36. Heimpel GE, De Boer JG: **Sex determination in the Hymenoptera.** *Annu Rev Entomol* 2008, **53:**209-230.

37. Beye M, Hasselmann M, Fondrk MK, Page Jr RE, Omholt SW: **The gene csd is the primary signal for sexual development in the honeybee and encodes an SR-type protein.** *Cell* 2003, **114:**419-429.

38. Beye M: **The dice of fate: the *csd* gene and how its allelic composition regulates sexual development in the honey bee, *Apis mellifera*.** *BioEssays* 2004, **26:**1131-1139.

39. Hasselmann M, Gempe T, Schiøtt M, Nunes-Silva CG, Otte M, Beye M: **Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees.** *Nature* 2008, **454:**519.

40. Gempe T, Hasselmann M, Schiøtt M, Hause G, Otte M, Beye M: **Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway.** *PLoS biology* 2009, **7:**e1000222.

41. Consortium HGS: **Insights into social insects from the genome of the honeybee *Apis mellifera*.** *Nature* 2006, **443:**931.

42. Dearden PK, Wilson MJ, Sablan L, Osborne PW, Havler M, McNaughton E, Kimura K, Milshina NV, Hasselmann M, Gempe T: **Patterns of conservation and change in honey bee developmental genes.** *Genome Research* 2006, **16:**1376-1384.

43. Duncan EJ, Benton MA, Dearden PK: **Canonical terminal patterning is an evolutionary novelty.** *Developmental biology* 2013, **377:**245-261.

44. Wilson MJ, Dearden PK: **Diversity in insect axis formation: two *orthodenticle* genes and *hunchback* act in anterior patterning and influence dorsoventral organization in the honeybee (*Apis mellifera*).** *Development* 2011, **138:**3497-3507.

45. Lynch JA, Brent AE, Leaf DS, Pultz MA, Desplan C: **Localized maternal orthodenticle patterns anterior and posterior in the long germ wasp *Nasonia*.** *Nature* 2006, **439:**728.

46. Lynch JA, Özüak O, Khila A, Abouheif E, Desplan C, Roth S: **The phylogenetic origin of oskar coincided with the origin of maternally provisioned germ plasm and pole cells at the base of the Holometabola.** *PLoS genetics* 2011, **7:**e1002029.

47. Wilson MJ, Havler M, Dearden PK: ***Giant*, *Krüppel*, and *caudal* act as gap genes with extensive roles in patterning the honeybee embryo.** *Developmental biology* 2010, **339:**200-211.

48. Wilson MJ, Dearden PK: ***Tailless* patterning functions are conserved in the honeybee even in the absence of Torso signaling.** *Developmental biology* 2009, **335:**276-287.

49. Lynch JA, Olesnicky EC, Desplan C: **Regulation and function of *tailless* in the long germ wasp *Nasonia vitripennis*.** *Development genes and evolution* 2006, **216:**493-498.

50. Duncan EJ, Johnson TK, Whisstock JC, Warr CG, Dearden PK: **Capturing embryonic development from metamorphosis: how did the terminal patterning signalling pathway of Drosophila evolve?** *Current Opinion in Insect Science* 2014, **1:**45-51.

51. Johnson TK, Henstridge MA, Warr CG: **MACPF/CDC proteins in development: Insights from *Drosophila* torso-like.** In *Seminars in cell & developmental biology*. Elsevier; 2017: 163-170.

52. Rewitz KF, Yamanaka N, Gilbert LI, O’Connor MB: **The insect neuropeptide PTTH activates receptor tyrosine kinase torso to initiate metamorphosis.** *Science* 2009, **326:**1403-1405.

53. Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Group NGW: **Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species.** *Science* 2010, **327:**343-348.

54. Skelly J, Pushparajan C, Duncan E, Dearden P: **Evolution of the Torso activation cassette, a pathway required for terminal patterning and moulting.** *Insect molecular biology* 2019.

55. Li A, Yang Y, Wu S, Li C, Wu Y: **Investigation of resistance mechanisms to fipronil in diamondback moth (Lepidoptera: Plutellidae).** *Journal of economic entomology* 2006, **99:**914-919.

56. Nakao T, Kawase A, Kinoshita A, Abe R, Hama M, Kawahara N, Hirase K: **The A2'N mutation of the RDL gamma-aminobutyric acid receptor conferring fipronil resistance in *Laodelphax striatellus* (Hemiptera: Delphacidae).** *Journal of Economic Entomology* 2011, **104:**646-652.

57. Nakao T, Naoi A, Kawahara N, Hirase K: **Mutation of the GABA receptor associated with fipronil resistance in the whitebacked planthopper, *Sogatella furcifera*.** *Pesticide Biochemistry and Physiology* 2010, **97:**262-266.

58. Nakao T, Hama M, Kawahara N, Hirase K: **Fipronil resistance in *Sogatella furcifera*: molecular cloning and functional expression of wild-type and mutant RDL GABA receptor subunits.** *Journal of Pesticide Science* 2012**:**1201310224-1201310224.

59. Wei Q, Mu XC, Wu SF, Wang LX, Gao CF: **Cross‐resistance to three phenylpyrazole insecticides and A2′ N mutation detection of GABA receptor subunit in fipronil‐resistant *Laodelphax striatellus* (Hemiptera: Delphacidae).** *Pest management science* 2017, **73:**1618-1624.

60. Janer EC, Klafke G, Fontes F, Capurro M, Schumaker T: **Mutations in *Rhipicephalus microplus* GABA gated chloride channel gene associated with fipronil resistance.** *Ticks and tick-borne diseases* 2019, **10:**761-765.

61. Nakao T, Naoi A, Hama M, Kawahara N, Hirase K: **Concentration-dependent effects of GABA on insensitivity to fipronil in the A2′ S mutant RDL GABA receptor from fipronil-resistant *Oulema oryzae* (Coleoptera: Chrysomelidae).** *Journal of economic entomology* 2012, **105:**1781-1788.

62. Rocheleau TA, Steichen JC, Chalmers AE: **A point mutation in a *Drosophila* GABA receptor confers insecticide resistance.** *Nature* 1993, **363:**449.

63. Le Goff G, Hamon A, Berge JB, Amichot M: **Resistance to fipronil in *Drosophila simulans*: influence of two point mutations in the RDL GABA receptor subunit.** *J Neurochem* 2005, **92:**1295-1305.

64. Remnant EJ, Morton CJ, Daborn PJ, Lumb C, Yang YT, Ng HL, Parker MW, Batterham P: **The role of *Rdl* in resistance to phenylpyrazoles in *Drosophila melanogaster*.** *Insect Biochem Mol Biol* 2014, **54:**11-21.

65. Anthony N, Unruh T, Ganser D, ffrench-Constant R: **Duplication of the Rdl GABA receptor subunit gene in an insecticide-resistant aphid, *Myzus persicae*.** *Mol Gen Genet* 1998, **260:**165-175.

66. Remnant EJ, Good RT, Schmidt JM, Lumb C, Robin C, Daborn PJ, Batterham P: **Gene duplication in the major insecticide target site, *Rdl*, in *Drosophila melanogaster*.** *Proc Natl Acad Sci U S A* 2013, **110:**14705-14710.

67. Yu LL, Cui YJ, Lang GJ, Zhang MY, Zhang CX: **The ionotropic g-aminobutyric acid receptor gene family of the silkworm, Bombyx mori.** *Genome* 2010, **53:**688-697.

68. Guest M, Goodchild JA, Bristow JA, Flemming AJ: **RDL A301S alone does not confer high levels of resistance to cyclodiene organochlorine or phenyl pyrazole insecticides in *Plutella xylostella*.** *Pestic Biochem Physiol* 2019, **158:**32-39.

69. Perry T, McKenzie JA, Batterham P: **A Da6 knockout strain of *Drosophila melanogaster* confers a high level of resistance to spinosad.** *Insect Biochem Mol Biol* 2007, **37:**184-188.

70. Perry T, Heckel DG, McKenzie JA, Batterham P: **Mutations in Da1 or Db2 nicotinic acetylcholine receptor subunits can confer resistance to neonicotinoids in Drosophila melanogaster.** *Insect Biochem Mol Biol* 2008, **38:**520-528.

71. Jones AK, Bera AN, Lees K, Sattelle DB: **The cys-loop ligand-gated ion channel gene superfamily of the parasitoid wasp, *Nasonia vitripennis*.** *Heredity* 2010, **104:**247-259.

72. Jones AK, Sattelle DB: **The cys-loop ligand-gated ion channel superfamily of the honeybee, *Apis mellifera*.** *Invert Neurosci* 2006, **6:**123-132.

73. Bass C, Denholm I, Williamson MS, Nauen R: **The global status of insect resistance to neonicotinoid insecticides.** *Pesticide Biochemistry and Physiology* 2015, **121:**78-87.

74. Perry T, Batterham P, Daborn PJ: **The biology of insecticidal activity and resistance.** *Insect Biochem Mol Biol* 2011, **41:**411-422.

75. Feyereisen R, Durst F: **Ecdysterone Biosynthesis: A Microsomal Cytochrome-P-450-Linked Ecdysone 20-Monooxygenase from Tissues of the African Migratory Locust.** *European Journal of Biochemistry* 1978, **88:**37-47.

76. Feyereisen R: **Arthropod CYPomes illustrate the tempo and mode in P450 evolution.** *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2011, **1814:**19-28.

77. Tang J, Amin Usmani K, Hodgson E, Rose RL: **In vitro metabolism of fipronil by human and rat cytochrome P450 and its interactions with testosterone and diazepam.** *Chem Biol Interact* 2004, **147:**319-329.

78. Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, Berenbaum MR, Feyereisen R, Oakeshott JG: **A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee.** *Insect Molecular Biology* 2006, **15:**615-636.

79. Terhzaz S, Cabrero P, Brinzer RA, Halberg KA, Dow JA, Davies SA: **A novel role of Drosophila cytochrome P450-4e3 in permethrin insecticide tolerance.** *Insect Biochem Mol Biol* 2015, **67:**38-46.

80. Brandt A, Scharf M, Pedra JHF, Holmes G, Dean A, Kreitman M, Pittendrigh BR: **Differential expression and induction of two *Drosophila* cytochrome P450 genes near the Rst(2)DDT locus.** *Insect Molecular Biology* 2002, **11:**337-341.

81. Le Goff G, Boundy S, Daborn PJ, Yen JL, Sofer L, Lind R, Sabourault C, Madi-Ravazzi L, ffrench-Constant RH: **Microarray analysis of cytochrome P450 mediated insecticide resistance in *Drosophila*.** *Insect Biochemistry and Molecular Biology* 2003, **33:**701-708.

82. Daborn PJ, Lumb C, Boey A, Wong W, ffrench-Constant RH, Batterham P: **Evaluating the insecticide resistance potential of eight *Drosophila melanogaster* cytochrome P450 genes by transgenic over-expression.** *Insect Biochemistry and Molecular Biology* 2007, **37:**512-519.

83. Battlay P, Leblanc PB, Green L, Garud NR, Schmidt JM, Fournier-Level A, Robin C: **Structural Variants and Selective Sweep Foci Contribute to Insecticide Resistance in the *Drosophila* Genetic Reference Panel.** *G3: Genes, Genomes, Genetics* 2018, **8:**3489-3497.

84. Green L, Battlay P, Fournier-Level A, Good RT, Robin C: **Cis- and trans-acting variants contribute to survivorship in a naïve *Drosophila melanogaster* population exposed to ryanoid insecticides.** *Proceedings of the National Academy of Sciences* 2019, **116:**10424-10429.

85. Bogwitz MR, Chung H, Magoc L, Rigby S, Wong W, O'Keefe M, McKenzie JA, Batterham P, Daborn PJ: **Cyp12a4 confers lufenuron resistance in a natural population of *Drosophila melanogaster*.** *Proceedings of the National Academy of Sciences* 2005, **102:**12807-12812.

86. Harrop TWR, Denecke S, Yang YT, Chan J, Daborn PJ, Perry T, Batterham P: **Evidence for activation of nitenpyram by a mitochondrial cytochrome P450 in *Drosophila melanogaster*.** *Pest Management Science* 2018, **74:**1616-1622.

87. Harrop TWR, Sztal T, Lumb C, Good RT, Daborn PJ, Batterham P, Chung H: **Evolutionary Changes in Gene Expression, Coding Sequence and Copy-Number at the Cyp6g1 Locus Contribute to Resistance to Multiple Insecticides in *Drosophila*.** *PLOS ONE* 2014, **9:**e84879.

88. Couto A, Monceau K, Bonnard O, Thiéry D, Sandoz J-C: **Olfactory attraction of the hornet *Vespa velutina* to honeybee colony odors and pheromones.** *Plos One* 2014, **9:**e115943.

89. Ruther J, Matschke M, Garbe L-A, Steiner S: **Quantity matters: male sex pheromone signals mate quality in the parasitic wasp Nasonia vitripennis.** *Proceedings of the Royal Society B: Biological Sciences* 2009, **276:**3303-3310.

90. Van Oystaeyen A, Oliveira RC, Holman L, van Zweden JS, Romero C, Oi CA, d'Ettorre P, Khalesi M, Billen J, Wäckers F: **Conserved class of queen pheromones stops social insect workers from reproducing.** *Science* 2014, **343:**287-290.

91. Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR: **A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*.** *Neuron* 1999, **22:**327-338.

92. Gao Q, Chess A: **Identification of candidate *Drosophila* olfactory receptors from genomic DNA sequence.** *Genomics* 1999, **60:**31-39.

93. Robertson HM, Warr CG, Carlson JR: **Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*.** *Proceedings of the National Academy of Sciences* 2003, **100:**14537-14542.

94. Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB: **Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction.** *Neuron* 2004, **43:**703-714.

95. Smith BH: **Recognition of female kin by male bees through olfactory signals.** *Proceedings of the national Academy of Sciences* 1983, **80:**4551-4553.

96. Breed MD, Diaz PH, Lucero KD: **Olfactory information processing in honeybee, *Apis mellifera*, nestmate recognition.** *Animal behaviour* 2004, **68:**921-928.

97. Trible W, Olivos-Cisneros L, McKenzie SK, Saragosti J, Chang NC, Matthews BJ, Oxley PR, Kronauer DJC: ***orco* Mutagenesis Causes Loss of Antennal Lobe Glomeruli and Impaired Social Behavior in Ants.** *Cell* 2017, **170:**727-735 e710.

98. Yan H, Opachaloemphan C, Mancini G, Yang H, Gallitto M, Mlejnek J, Leibholz A, Haight K, Ghaninia M, Huo L, et al: **An Engineered *orco* Mutation Produces Aberrant Social Behavior and Defective Neural Development in Ants.** *Cell* 2017, **170:**736-747 e739.

99. Robertson HM, Wanner KW: **The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family.** *Genome research* 2006, **16:**1395-1403.

100. Robertson HM, Gadau J, Wanner KW: **The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis*.** *Insect molecular biology* 2010, **19:**121-136.

101. Jones WD, Nguyen T-AT, Kloss B, Lee KJ, Vosshall LB: **Functional conservation of an insect odorant receptor gene across 250 million years of evolution.** *Current Biology* 2005, **15:**R119-R121.

102. Venthur H, Zhou J-J: **Odorant Receptors and Odorant-Binding Proteins as Insect Pest Control Targets: A Comparative Analysis.** *Frontiers in Physiology* 2018, **9**.

103. Gruber MAM, Quinn O, Baty JW, Dobelmann J, Haywood J, Wenseleers T, Lester PJ: **Fitness and microbial networks of the common wasp, *Vespula vulgaris* (Hymenoptera: Vespidae), in its native and introduced ranges.** *Ecological Entomology* 2019, **44:**512-523.

104. Pettersson JHO, Shi M, Bohlin J, Eldholm V, Brynildsrud OB, Paulsen KM, Andreassen Å, Holmes EC: **Characterizing the virome of *Ixodes ricinus* ticks from northern Europe.** *Scientific Reports* 2017, **7:**10870.

105. Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, Qin X-C, Li J, Cao J-P, Eden J-S, et al: **Redefining the invertebrate RNA virosphere.** *Nature* 2016, **540:**539–543.

106. Shi M, Neville P, Nicholson J, Eden J-S, Imrie A, Holmes EC: **High-Resolution Metatranscriptomics Reveals the Ecological Dynamics of Mosquito-Associated RNA Viruses in Western Australia.** *Journal of Virology* 2017, **91:**e00680-00617.

107. Harris RJ: **Diet of the wasps *Vespula vulgaris* and *V. germanica* in honeydew beech forest of the South Island, New Zealand.** *New Zealand Journal of Zoology* 1991, **18:**159-169.

108. Loope KJ, Baty JW, Lester PJ, Wilson Rankin EE: **Pathogen shifts in a honeybee predator following the arrival of the *Varroa* mite.** *Proc R Soc B* 2019, **286**.

109. Manley R, Boots M, Wilfert L: **Emerging viral disease risk to pollinating insects: ecological, evolutionary and anthropogenic factors.** *The Journal of Applied Ecology* 2015, **52:**331-340.

110. Sébastien A, Lester PJ, Hall RJ, Wang J, Moore NE, Gruber MAM: **Invasive ants carry novel viruses in their new range and form reservoirs for a honeybee pathogen.** *Biology Letters* 2015, **11**.

111. Brenton-Rule EC, Dobelmann J, Baty JW, Brown RL, Dvorak L, Grangier J, Masciocchi M, McGrannachan C, Shortall CR, Schmack J, et al: **The origins of global invasions of the German wasp (*Vespula germanica*) and its infection with four honey bee viruses.** *Biological Invasions* 2018, **20:**3445-3460.

112. Dobelmann J, Loope KJ, Wilson‐Rankin E, Quinn O, Baty JW, Gruber MAM, Lester PJ: **Fitness in invasive social wasps: the role of variation in viral load, immune response and paternity in predicting nest size and reproductive output.** *Oikos* 2017, **126:**1208-1218.

113. Kolarich D, Loos A, Léonard R, Mach L, Marzban G, Hemmer W, Altmann F: **A proteomic study of the major allergens from yellow jacket venoms.** *Proteomics* 2007, **7:**1615-1623.

114. Bilo B, Rueff F, Mosbech H, Bonifazi F, Oude‐Elberink J, Hypersensitivity EIGoIV: **Diagnosis of Hymenoptera venom allergy.** *Allergy* 2005, **60:**1339-1349.

115. Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, Clark AG, Robinson GE: **Genes involved in convergent evolution of eusociality in bees.** *Proceedings of the National Academy of Sciences* 2011, **108:**7472-7477.

116. Urich K: **Glycolysis.** In *Comparative Animal Biochemistry.* Springer; 1994: 514-561

117. Kucharski R, Maleszka J, Foret S, Maleszka R: **Nutritional control of reproductive status in honeybees via DNA methylation.** *Science* 2008, **319:**1827-1830.

118. Bonasio R, Li Q, Lian J, Mutti NS, Jin L, Zhao H, Zhang P, Wen P, Xiang H, Ding Y: **Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*.** *Current Biology* 2012, **22:**1755-1764.

119. Guan C, Barron AB, He XJ, Wang ZL, Yan WY, Zeng ZJ: **A comparison of digital gene expression profiling and methyl DNA immunoprecipitation as methods for gene discovery in honeybee (*Apis mellifera*) behavioural genomic analyses.** *PloS one* 2013, **8:**e73628.

120. Herb BR, Wolschin F, Hansen KD, Aryee MJ, Langmead B, Irizarry R, Amdam GV, Feinberg AP: **Reversible switching between epigenetic states in honeybee behavioral subcastes.** *Nature neuroscience* 2012, **15:**1371.

121. Li-Byarlay H, Li Y, Stroud H, Feng S, Newman TC, Kaneda M, Hou KK, Worley KC, Elsik CG, Wickline SA: **RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee.** *Proceedings of the National Academy of Sciences* 2013, **110:**12750-12755.

122. Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ: **Evolution of DNA methylation across insects.** *Molecular biology and evolution* 2016, **34:**654-665.

123. Glastad KM, Arsenault SV, Vertacnik KL, Geib SM, Kay S, Danforth BN, Rehan SM, Linnen CR, Kocher SD, Hunt BG: **Variation in DNA methylation is not consistently reflected by sociality in Hymenoptera.** *Genome biology and evolution* 2017, **9:**1687-1698.

124. Bird AP: **DNA methylation and the frequency of CpG in animal DNA.** *Nucleic acids research* 1980, **8:**1499-1504.

125. Elango N, Hunt BG, Goodisman MA, Soojin VY: **DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*.** *Proceedings of the National Academy of Sciences* 2009, **106:**11206-11211.

126. Richards S, Murali SC: **Best practices in insect genome sequencing: what works and what doesn’t.** *Current opinion in insect science* 2015, **7:**1-7.

127. Chapman JA, Ho I, Sunkara S, Luo S, Schroth GP, Rokhsar DS: **Meraculous: de novo genome assembly with short paired-end reads.** *PloS one* 2011, **6:**e23501.

128. Chapman JA, Ho IY, Goltsman E, Rokhsar DS: **Meraculous2: fast accurate short-read assembly of large polymorphic genomes.** *arXiv preprint arXiv:160801031* 2016.

129. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome.** *science* 2009, **326:**289-293.

130. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST: **Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome.** *Nature genetics* 2017, **49:**643.

131. Cabanettes F, Klopp C: **D-GENIES: dot plot large genomes in an interactive, efficient and simple way.** *PeerJ* 2018, **6:**e4958.

132. Palmer J, Stajich J: **nextgenusfs/funannotate: funannotate v1.6.0.** Zenodo; 2019.

133. Smit AFA, Hubley R: **RepeatModeler Open-1.0.** 2008.

134. Smit AFA, Hubley R, Green P: **RepeatMasker Open-4.0.** 2013.

135. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics* 2013, **29:**15-21.

136. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A: **UniProtKB/Swiss-Prot.** In *Plant Bioinformatics: Methods and Protocols.* Edited by Edwards D. Totowa, NJ: Humana Press; 2007: 89-112: *Methods in Molecular Biology™*].

137. Pruitt KD, Tatusova T, Maglott DR: **NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Research* 2005, **33:**D501-D504.

138. Drăgan M-A, Moghul I, Priyam A, Bustos C, Wurm Y: **GeneValidator: identify problems with protein-coding gene predictions.** *Bioinformatics* 2016, **32:**1559-1561.

139. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE: **Web Apollo: a web-based genomic annotation editing platform.** *Genome Biology* 2013, **14:**R93.

140. Dongen Sv, Hazewinkel M, Eijck DJNv, Universiteit U: **Graph clustering by flow simulation.** s.n.], 2000.

141. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, et al: **InterPro in 2019: improving coverage, classification and access to protein sequence annotations.** *Nucleic Acids Res* 2019, **47:**D351-D360.

142. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nat Biotech* 2011, **29:**644-652.

143. Shi M, Lin X-D, Vasilakis N, Tian J-H, Li C-X, Chen L-J, Eastwood G, Diao X-N, Chen M-H, Chen X: **Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the Flaviviridae and related viruses.** *Journal of virology* 2016, **90:**659-669.

144. Remnant EJ, Shi M, Buchmann G, Blacquière T, Holmes EC, Beekman M, Ashe A: **A diverse range of novel RNA viruses in geographically distinct honey bee populations.** *Journal of virology* 2017, **91:**e00158-00117.

145. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Research* 2004, **32:**1792-1797.

146. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.** *Bioinformatics* 2012, **28:**1647-1649.

147. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.** *Molecular Biology and Evolution* 2016, **33:**1870-1874.

148. Fraley C, Raftery AE: **Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST.** *Journal of Classification* 2003, **20:**263-286.

149. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J: **Fast, scalable generation of high‐quality protein multiple sequence alignments using Clustal Omega.** *Molecular systems biology* 2011, **7**.

150. Castresana J: **Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.** *Molecular biology and evolution* 2000, **17:**540-552.

151. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19:**1572-1574.