Title
High-Quality Assemblies for Three Invasive Social Wasps from the Vespula Genus.

Permalink
https://escholarship.org/uc/item/8h8308gj

Journal
G3 (Bethesda, Md.), 10(10)

ISSN
2160-1836

Authors
Harrop, Thomas WR
Guhlin, Joseph
McLaughlin, Gemma M
et al.

Publication Date
2020-10-05

DOI
10.1534/g3.120.401579

Peer reviewed
High-Quality Assemblies for Three Invasive Social Wasps from the *Vespula* Genus

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,** Jens Van Eeckhoven,§ Brittany Graham,§ Rosemary A. Knapp,§ Emily J. Remnant,** Jens Van Eeckhoven,§ Brittany Graham,§ Rosemary A. Knapp,§ Emily J. Remnant,** Jens Van Eeckhoven,§

*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, New Zealand, §§School of Biology, Faculty of Biological Sciences, University of Leeds, LS2 9JT, United Kingdom, **School of Life and Environmental Sciences, University of Sydney, NSW 2006, Australia, ††Phase Genomics Inc., 1617 8th Ave N, Seattle, WA 98109, ††Department of Entomology, University of California, Riverside, CA 92521

ORCID IDs: 0000-0001-6557-0440 (T.W.R.H.); 0000-0003-3264-7178 (J.G.); 0000-0002-6724-9519 (M.F.L.L.); 0000-0001-6780-8838 (M.A. M.G.); 0000-0002-3808-8365 (E.J.R.); 0000-0002-1801-5687 (P.J.L.); 0000-0001-7790-9675 (P.K.D.)

**ABSTRACT** Social wasps of the genus *Vespula* have spread to nearly all landmasses worldwide and have become significant pests in their introduced ranges, affecting economies and biodiversity. Comprehensive genome assemblies and annotations for these species are required to develop the next generation of control strategies and monitor existing chemical control. We sequenced and annotated the genomes of the common wasp (*Vespula vulgaris*), 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 annotated gene sets relevant to the applied management of invasive wasp populations, including genes associated with spermatogenesis and development, pesticide resistance, olfactory receptors, immunity and venom. These genomes provide evidence for active DNA methylation in Vespidae and tandem duplications of venom genes. Our genomic resources will contribute to the development of next-generation control strategies, and monitoring potential resistance to chemical control.

**KEYWORDS** *Vespula germanica* *Vespula pensylvanica* *Vespula vulgaris* Hymenoptera social insects genomes

Social wasps (Hymenoptera: Vespidae) are remarkable because their highly eusocial lifestyle appears to have evolved independently of other eusocial Hymenoptera (Piekarski et al. 2018; Hines et al. 2007; Peters et al. 2017). The eusocial lifestyle is characterized by overlapping generations of adults living together, cooperative care of offspring and reproductive division of labor (Michener and Brothers 1974). Along with their foraging flexibility and predatory ability, eusociality appears to play a major role in the ecological success of social wasps (Grimaldi and Engel 2005; Lester and Beggs 2019).

Vespid wasps can be effective pollinators of plants including ivy or orchid species (Cheng et al. 2009; Jacobs et al. 2010). They are generalist predators and effect biological control of some pest species (Donovan 2003). Their ecological success can also be problematic. Invasive 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 (Lester and Beggs 2019). Colonies are smaller in the native ranges (Archer and Turner 2014 p. 61). In New Zealand’s native
beech forests, Vespid wasp populations can reach up to 40 nests per hectare and have a biomass similar to, or greater than, the combined biomasses of birds and mammals (Thomas et al. 1990; Lester et al. 2017). These invasive Vespid populations (Figure 1; Lester and Beggs 2019) have major impacts on ecosystems because of their large colony sizes, reproductive capacity and flexible predation.

Current control methods for vespid wasps are limited, with pesticides containing Fipronil being the most common and widespread chemical control method (Edwards et al. 2017). The use of this neurotoxic pesticide over large areas, and over consecutive years, may select for resistance. Fipronil resistance has been observed in other economically important insect pests (Matsumura et al. 2008; reviewed in Feyereisen et al. 2015). Next-generation pest control technologies, including gene drives, have been proposed as part of an alternative solution for controlling or eradicating invasive social wasps (Dearden et al. 2018). Targets for genetic modification include developmental genes associated with wasp fitness or fecundity (Lester and Beggs 2019; Dearden et al. 2018). Gene drives that have immune system targets have been proposed and developed in the laboratory for other pests such as mosquitoes (e.g., Gantz et al. 2015; Dong et al. 2018). 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.

Five genomes from the Polistes genus of the Vespidae sub-clade are available in the NCBI Genome database. Genomes for Polistes canadensis and Polistes dominula were assembled from short reads (Standage et al. 2016; Patalano et al. 2015). Genomes for Polistes dorsalis, Polistes fuscatus and Polistes metricus, based on Pacific Biosciences single-molecule sequencing, were recently deposited in the NCBI database (Miller et al. 2020). The three long-read genomes have improved contiguity compared to the short-read genomes produced for P. dominula and P. canadensis (Standage et al. 2016; Patalano et al. 2015; Miller et al. 2020). Here, we report chromosome-level assemblies for three Vespula species that are invasive pests in their introduced ranges (Figure 1; Lester and Beggs 2019). We undertake manual and computational annotation and phylogenomic analyses, with emphasis on sets of genes relevant to chemical control, olfaction, venom and DNA methylation. These resources will be useful for understanding the biology of vespid wasps, developing next-generation control strategies, and monitoring resistance to current chemical control.

MATERIALS AND METHODS

Genome assembly and scaffolding

Vespula pensylvanica samples were collected in Volcano, Hawaii in August 2017 (lat 19.43, lon -155.21). 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 adaptors 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 (Chapman et al. 2011, 2016). Chromosome-scale scaffolds were generated using Chicago and Dovetail Hi-C data from diploid females, and implemented using Dovetail’s proprietary HiRise software.

We collected V. vulgaris samples in May 2015, from Richmond Hill Forest Park, New Zealand (lat -41.3292, lon 173.4637). V. germanica samples were collected in March 2018, from Lincoln, New Zealand (lat -43.6384, lon 172.4784). We used libraries generated from a single, haploid male to produce 17.1 Gb of 150bp 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 (Bushnell 2014), we assembled draft genomes with Meraculous 2.2.6 (Chapman et al. 2011, 2016). We repeated assembly with a range of parameters, and used BUSCO analysis (Simão et al. 2015), 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/vvl-illumina. Draft genomes were scaffolded by Phase Genomics using Hi-C data generated from pools of 20 larvae. Chromatin conformation capture data were generated using Phase Genomics Proximo Hi-C 2.0 Kit, which is a commercially available version of the Hi-C protocol (Lieberman-Aiden et al. 2009). The Phase Genomics Proximo Hi-C genome scaffolding platform was used to create chromosome-scale scaffolds from the corrected assembly as described Bickhart et al. (2017).

Genome curation

Command-line arguments and scripts can be found at https://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. Our V. vulgaris and V. germanica assemblies were aligned to the V. pensylvanica genome using D-GENIES, which inserts contigs into syntenic locations, flanked by 100 Ns, to assign chromosome names to those most related to those in V. pensylvanica (Cabannettes and Klopp 2018). Scaffolds were numbered with four digits in order of size.

Repeat masking

Repeats were identified using RepeatModeler 2.0.1 (github.com/Dfam-consortium/RepeatModeler) and RepeatMasker 4.0.9 (repeatmasker.org/RMDownload.html) via the funannotate pipeline (Palmer and Stajich 2019).

RNA sequencing

Vespula vulgaris queens, workers, and larvae were sampled from mature nests in the native range of Belgium and the introduced range in New Zealand and total RNA transcriptome data were generated as described by Gruber et al. (2019).

Gene prediction

We performed iterative gene prediction using the Funannotate pipeline v1.6.0, manual annotation, and extrinsic protein evidence (Palmer and Stajich 2019). For V. vulgaris we used RNA-seq reads (described in the previous section) from V. vulgaris queens, workers, and larvae as additional evidence. The reads were trimmed with sickle (github.com/najoshi/sickle) and aligned to our assembly using STAR in two-pass mode (Dobin et al. 2012). Gene predictions were performed on the assembly using funannotate predict with the RNA alignments and extrinsic protein of all Vespula proteins from NCBI,Apis mellifera, Nasonia vitripennis, and the UniProt SWISS-PROT
database (Boutet et al. 2007; Pruitt et al. 2005; The Honeybee Genome Sequencing Consortium 2006; Werren et al. 2010).

Initial predictions were evaluated with GeneValidator with UniProt SWISS-PROT database as the high-quality targets. Genes whose protein predictions scored > 90 were used to train Augustus via the optimize_augustus.pl script (Drăgan et al. 2016). The prediction step of funannotate was then re-run as before using the trained vvulg AUGUSTUS species definition. This allowed retention of high-quality gene predictions from the target species to be used as a training set for AUGUSTUS gene prediction, a component of the funannotate pipeline. This process was repeated for *V. pensylvanica* and *V. germanica* using *V. vulgaris* species definition as the initial AUGUSTUS in the first iteration of gene prediction, generating a species-specific configuration in the following round. This two-step gene prediction with validation and training using high-confidence gene calls between the first and second round allowed for the creation of species-specific AUGUSTUS models.

**Manual curation**

Genes were manually curated in WebApollo (Lee et al. 2013). These manual annotations took precedence over intersecting computational 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 using AUGUSTUS-PPX for the LGIC and Olfactory families (Keller et al. 2011). Protein sequences of interest from external sources were clustered based on bitscore using BLAST+ and MCL (Dongen 2000). Clusters of protein sequences were converted to protein profiles via AUGUSTUS tool msa2prfl.pl. Assemblies were searched with fastBlockSearch and gene prediction was performed on matched regions with an additional flanking sequence of 1kbp. These predictions took precedence over intersecting computationally predicted genes.

**Annotation**

Further annotation was performed with funannotate using InterProScan 5.32-71.0 (Mitchell et al. 2019). Genes were renamed using custom scripts. 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

---

**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). D. Native and invaded ranges of the sequenced Vespula species. E. Rooted phylogenetic species tree of Vespula species and other sequenced Hymenoptera, with *Drosophila melanogaster* as an outgroup, derived from 415 orthogroups using OrthoFinder.
Hymenoptera base using OrthoFinder (Elsik et al. 2016; Emms and Kelly 2019).

**Methylation analysis**
Nucleotide and dinucleotide content of gene body sequences were calculated using a custom perl script. CpG[o/e] was calculated as the number of CpG dinucleotides divided by the number of C nucleotides times the number of G nucleotides. The number of components in CpG[o/e] distributions was estimated in R (R Core Team 2015) using mclust model-based clustering (Scrucca et al. 2016). The best fitting model was identified among several non-nested models using Bayesian information criteria (BIC).
RESULTS AND DISCUSSION

Genome assemblies and annotation

We used a combination of short-read Illumina 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; Supplementary table 1; N50 length 8.30–8.53 Mb), which likely represent the 25 chromosomes observed in *Vespula* karyotypes (Hoshiba et al. 1989). Each draft genome also contains 10–200 unanchored scaffolds (N50 lengths 1.77–2.28 kb; Supplementary table 2). These genomes are similar in size to the genomes of the closely related European paper wasp, *Polistes dominula* (Standage et al. 2016), and the red paper wasp, *Polistes canadensis* (Patalano et al. 2015). However, the contiguity in our *Vespula* assemblies is higher than Illumina-based assemblies of other Vespidae, and comparable to the latest-generation *Apis mellifera* assembly (Table 1). 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 re-arrangements (Figure 2B; Supplementary Figure 1). Repeat masking masked 17.86%, 18.75% and 18.71% of the *V. vulgaris*, *V. pensylvanica*, and *V. germanica* genomes, respectively. We predicted 16,751, 17,854, and 19,142 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 (Simão et al. 2015) on a published estimate for Hymenoptera (Peters et al. 2016). 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

Data availability

Raw sequence data are hosted in the NCBI Sequence Read Archive under accession PRJNA643352. Assembled genomes are available on GenBank under accessions JACSDY00000000 (Vespula pensylvanica), JACSDZ000000000 (Vespula germanica) and JACSEA000000000 (Vespula vulgaris). Supplemental material available at figshare: https://doi.org/10.25387/g3.12885599.
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 hymenopteran 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.

93.7–94.5% of predicted orthogroups in *P. canadensis* and 96.0–96.9% in *P. dominula* clustered into orthogroups that contained a *Vespula* gene. *Vespula* genes from our annotations were present in 58.9–62.8% of all orthogroups across Hymenoptera and *D. melanogaster*, compared to 40.7% and 41.2% for *P. canadensis* and *P. dominula*, respectively (Table 2). Other hymenopteran species had a member in 37.8–44.5% of orthogroups, indicating gene families may be missing in predictions among other species. Most of the genes we annotated were in shared orthogroups, with 100 genes (0.1–0.4% per species) in species-specific orthogroups. Other hymenopteran genomes had 0–3.3% (mean 1.7%) of genes in species-specific orthogroups. Our annotation results suggest that Vespid wasps have more genes than other Hymenoptera and/or gene annotation in other Hymenoptera is incomplete. This could be resolved by re-annotation of other hymenopteran genes using a comparative approach.

**Evidence for active DNA methylation**

DNA methylation has been functionally linked to caste specification in honeybees and ants and division of labor in honeybees (Herb et al. 2012; Kucharski et al. 2008; Bonasio et al. 2012; Guan et al. 2013). DNA methylation may be integral for aspects of eusociality (reviewed by Li-Byarlay et al. 2013), although recent studies have found no consistent link (Bewick et al. 2017; Glastad et al. 2017). In mice, Dnmt3 enzymes catalyze *de novo* DNA methylation (reviewed by Goll and Bestor 2005). The genomes of *P. canadensis* and *P. dominula* do not encode a Dnmt3 homolog (Standage et al. 2016; Patalano et al. 2015; Bewick et al. 2017). While this manuscript was under revision, three new *Polistes* genomes based on

---

**Table 2 Gene content and orthogroup representation for selected hymenopteran genomes**

| Species            | Genes | Orthogroups represented | Orthogroups represented (%) | Species-specific orthogroups | Genes in species-specific orthogroups (%) |
|--------------------|-------|-------------------------|----------------------------|-----------------------------|------------------------------------------|
| *Vespula vulgaris* | 16,751 | 13,141                  | 58.9%                      | 5                           | 0.1%                                     |
| *Vespula germanica* | 17,854 | 13,739                  | 61.6%                      | 23                          | 0.4%                                     |
| *Vespula pensylvanica* | 19,142 | 14,022                  | 62.8%                      | 11                          | 0.1%                                     |
| *Polistes canadensis* | 10,518 | 9,086                   | 40.7%                      | 12                          | 0.2%                                     |
| *Polistes dominula* | 11,069 | 9,193                   | 41.2%                      | 27                          | 0.5%                                     |
| *Apis mellifera*   | 14,064 | 8,949                   | 40.1%                      | 35                          | 0.6%                                     |
| *Nasonia vitripennis* | 14,647 | 8,901                   | 39.9%                      | 616                         | 11.1%                                     |

---

**Figure 3** A. Whole-genome alignments of five *Polistes* genomes and genomes of *V. germanica* and *V. pensylvanica* against Chr11 of the *V. vulgaris* genome, showing the Dnmt3 locus (Vvulg11g01820) and the surrounding 20 kb. None of the *Polistes* assemblies have regions that are homologous to the Dnmt3 locus, and annotations for *P. canadensis* and *P. dominula* do not include a Dnmt3 gene (Standage et al. 2016; Patalano et al. 2015; Bewick et al. 2017). All three *Vespula* genomes encode an ortholog of Dnmt3. B. Frequency histograms of CpG [o/e] observed in coding sequences of *Drosophila melanogaster*, *Apis mellifera*, *Polistes canadensis*, *Polistes dominula*, *Vespula germanica*, *Vespula vulgaris* and *Vespula pensylvanica*. The y-axis depicts the number of genes (Frequency) with CpG[0/e] values given on the x-axis. The distribution of CpG [o/e] in Vespid wasps is a trimodal distribution, with a low-CpG[0/e] peak consistent with the presence of historical DNA methylation in all five Vespid species.
Figure 4 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 Polistes, one of which is missing from the V. vulgaris genome.
Pacific Biosciences single-molecule sequencing were deposited in the NCBI database (Miller et al. 2020). Genome annotations have not been reported for the three long-read *Polistes* genomes. We aligned all five *Polistes* genomes and our *V. germanica* and *V. pensylvanica* genomes against Chr1 of the *V. vulgaris* genome, which encodes Dnmt3. None of the five *Polistes* genomes have a region that aligns to the *V. vulgaris* Dnmt3 locus (Figure 3A). All three *Vespa* genomes encode an ortholog of Dnmt3, indicating the lack of this gene in *Polistes* is due to gene loss following the divergence from the *Vespa* lineage approximately 50 million years ago (Peters et al. 2017). *Vespa* genomes all contain a single ortholog of Dnmt1 and have evidence of active DNA methylation (Figure 3B; Supplementary Figure 3).

**Gene content and duplications**
We found more olfactory receptor (OR) genes in *Vespa* genomes than in the genomes of *P. dominula* and *P. canadensis*. We predicted 120 OR genes in *V. vulgaris*, 133 in *V. germanica*, and 102 in *V. pensylvanica*. Annotations for *P. dominula* and *P. canadensis* contain 94 and 72 OR genes, respectively (Supplementary table 3). In contrast, honeybee and *Nasonia* genomes encode more OR genes (170 and 301, respectively; Robertson and Wanner 2006; Robertson et al. 2010). *Vespa* OR genes cluster into 28 orthogroups. The co-receptor Orco is present in all genomes and forms a stable orthogroup (orthogroup 7148; Larsson et al. 2004; Jones et al. 2005). In contrast, there are significant expansions of particular OR orthogroups in the *Vespid* wasps, and these differ from the groups expanded in *Nasonia* and honeybee (Robertson et al. 2010; Robertson and Wanner 2006). Orthogroup members are arranged in expanded tandem arrays on chromosome 3 (orthogroup 51), chromosome 13 (orthogroup 2434), and chromosome 25 (orthogroup 232) of *Vespa* genomes. In these clusters, numbers of genes vary between species, implying that duplications and deletions are recent and ongoing. 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 signaling systems present within these species.

Some of the genes encoding venom components also have variable copy numbers in *Vespa* genomes. The major allergens in *Vespa* venoms are phospholipase A1, hyaluronidase, and antigen 5 (Bílò et al. 2005; Kolarich et al. 2007). Phospholipase A1 is found in three tandem copies in the *P. dominula* and *V. germanica* genomes (chromosome 9 in *Vespa*), 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 4) implies that these are independent amplifications. The hyaluronidase gene is duplicated in our three *Vespid* genomes, but not *P. dominula*. These are tandem duplications that appear to be present in the last common ancestor of the three *Vespid* species. *P. dominula* and *V. germanica* and *V. pensylvanica* also have two Antigen 5 genes, but these duplications appear ancient before the common ancestor of Hymenoptera (Figure 4). In *Vespa* species, one copy is on chromosome 6 and one is on chromosome 7. In *V. vulgaris*, the chromosome 6 copy is absent. Duplication of venom genes in Vespid is no surprise, given the importance of venom to their life cycle.

Primary chemical control of *Vespa* populations is through the use of baits containing a low concentration of the phenylpyrazole insecticide, Fipronil (Lester and Beggs 2019; Edwards et al. 2017). We used targeted prediction to identify ligand-gated ion channel (LGIC), olfactory receptor, and spermatogenesis genes using Augustus protein profiles (Keller et al. 2011). Our annotation of the Fipronil target site, the GABA receptor Resistant to dieldrin (Rdl), did not suggest the presence of the classical Ala301 mutation that confers high resistance. *Vespula* LGICs are highly conserved, with one-to-one orthology in *Apis* and *Bombus* (Supplementary Figure 4). This suggests that any chemicals targeting *Vespa* LGICs will also affect bees, as is the case with Fipronil.

**CONCLUSIONS**
We have produced chromosome-level genome assemblies for three invasive social wasps in the *Vespa* 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 $N_{50}$ length > 300 kb; Richards and Murali 2015). Using manual curation and computational prediction, we have identified genes that may encode specific biology suitable for targeting with next-generation control technologies, and 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 independently of other Hymenoptera (PiekarSKI et al. 2018; Hines et al. 2007; Peters et al. 2017). 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 *Vespa*, in programs using both chemical 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. The genomic resources we have developed will also be essential for monitoring the effects of next-generation control methods and measuring genomic variation across natural populations.

**ACKNOWLEDGMENTS**
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.

**LITERATURE CITED**

Archer, M. E., and J. Turner, 2014 Handbooks for the identification of British insects. Vol. 6, Pt. 6: The vespid wasps (Tiphidae, Mutillidae, Sapygidae, Scoilidae and Vespidae) of the British Isles. Field Studies Council, Telford. ISBN:9780901546982

Bewick, A. J., K. J. Vogel, A. J. Moore, and R. J. Schmitz, 2017 Evolution of DNA Methylation across Insects. Mol. Biol. Evol. 34: 654–665. https://doi.org/10.1093/molbev/msw264

Bickhart, D. M., B. D. Rosen, S. Koren, B. L. Sayre, A. R. Hastie et al., 2017 Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat. Genet. 49: 643–650. https://doi.org/10.1038/ng.3802

Bílò, B. M., F. Rueff, H. Mosbach, F. Bonifazi, and J. N. G. Oude-Elberink, 2005 Diagnosis of Hymenoptera venom allergy. Allergy 60: 1339–1349. https://doi.org/10.1111/j.1398-9995.2005.00963.x
