Genomic resources for *Goniozus legneri*, *Aleochara bilineata* and *Paykullia maculata*, representing three independent origins of the parasitoid lifestyle in insects

Ken Kraaijeveld*, Peter Neleman*, Janine Mariën*, Emile de Meijer†, Jacintha Ellers*

*Department of Ecological Science, Faculty of Science, Vrije Universiteit, Amsterdam, The Netherlands

†Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

Author for correspondence: ken@kenkraaijeveld.nl

Data availability

The Whole Genome Shotgun projects have been deposited at DDBJ/ENA/GenBank under the accessions NCVS000000000 (G. legneri), NBZA000000000 (A. bilineata) and NDXZ000000000 (P. maculata). The versions described in this paper are versions NCVS010000000, NBZA010000000 and NDXZ010000000, respectively. Mapped reads and genome annotations are available through http://parasitoids.labs.vu.nl/parasitoids/. This website also includes genome browsers and viroblast instances for each genome.
Abstract

Parasitoid insects are important model systems for a multitude of biological research topics and widely used as biological control agents against insect pests. While the parasitoid lifestyle has evolved numerous times in different insect groups, research has focused almost exclusively on Hymenoptera from the parasitica clade. The genomes of several members of this group have been sequenced, but no genomic resources are available from any of the other, independent evolutionary origins of the parasitoid lifestyle. Our aim here was to develop genomic resources for three parasitoid insects outside the parasitica. We present draft genome assemblies for Goniozus legneri, a parasitoid Hymenopteran more closely related to the non-parasitoid wasps and bees than to the parasitica wasps, the Coleopteran parasitoid Aleochara bilineata and the Dipteran parasitoid Paykullia maculata. The genome assemblies are fragmented, but complete in terms of gene content. We also provide preliminary structural annotations. We anticipate that these genomic resources will be valuable for testing the generality of findings obtained from parasitica wasps in future comparative studies.

Key words

parasitoid, insect, genome, whole genome sequencing, genome report
Introduction

Parasitoid insects have long been used as model systems for the study of a wide variety of topics in biology, including life history, chemical ecology and population dynamics (Godfray 1994; Hawkins and Sheehan 1994; Wajnberg and Colazza 2013). Parasitoids are also widely employed as agents of biological control against insect pests (Heimpel and Mills 2017). In recent years, the study of parasitoid insects has received new impetus through the availability of a steadily growing number of available genome sequences. Genomes of 13 parasitoid insects have recently become available, all from within one monophyletic clade of parasitoid wasps (Branstetter et al. 2018). These genomes are rapidly becoming a rich source of information on many aspects of parasitoid biology, e.g. (Werren et al. 2010).

The vast literature on insect parasitoids deals almost exclusively with Hymenopteran parasitoids, which all share a single evolutionary origin (Eggleton and Belshaw 2013). The stinging wasps (Aculeata) diverged from this group and lost the parasitoid life style. However, recent phylogenetic studies place the parasitoid Chrysidoidea within the Aculeata, suggesting that they may have re-evolved the parasitoid life style secondarily (Carr et al. 2010; Peters et al. 2017). Outside the Hymenoptera, parasitoid lifestyles have evolved in multiple insect groups, including Diptera, Coleoptera, Lepidoptera, and Neuroptera (Eggleton and Belshaw 2013). Dozens, or even hundreds, of evolutionarily independent parasitoid lineages are found within the Diptera and Coleoptera (Feener and Brown 1997; Eggleton and Belshaw 2013). It has been estimated that 20% of all parasitoid insect species are Dipterans (Feener and Brown 1997). Some of these are economically important, either as biological control agent (Grenier 1988) or as pest (Prabhakar et al. 2017). The study of such non-hymenopteran parasitoid systems would benefit from genomic resources, as it is unknown to what extent insights from hymenopteran parasitoids can be extrapolated to non-hymenopteran parasitoids. Unfortunately, no sequenced genomes are available for any of these groups as yet.

Here, we present draft genome assemblies for three parasitoid insect species that each represent an evolutionary independent acquisition of the parasitoid lifestyle (Figure 1A).
**Goniozus legneri** is a parasitoid wasp from the superfamily Chrysidoidea (family Bethylidae) that is not part of the species-rich and well-studied parasitica clade. *G. legneri* can therefore function as an outgroup in comparative analyses of parasitica wasps and may have re-evolved the parasitoid lifestyle after it was lost at the base of the Aculeata. *G. legneri* is a gregarious parasitoid of Lepidopteran larvae that stings and paralyzes its prey before ovipositing on it externally (Khidr et al. 2013). The female then guards the host against utilization by other females (Khidr et al. 2013). *Aleochara bilineata* is a Coleopteran parasitoid of Dipteran pupae that represents another evolutionary independent acquisition of the parasitoid lifestyle. Females lay their eggs in the proximity of hosts (Bili et al. 2016). After hatching, the first instar larvae search for, enter and parasitize host pupae (Bili et al. 2016). *Paykullia maculata* is a parasitoid fly from the family of the Rhinophoridae, representing one of the many independent acquisitions of the parasitoid mode within the Diptera. Like all Rhinophoridae, *P. maculata* parasitizes isopods. Females lay their eggs in the vicinity of isopod aggregations. The larvae latch on to a passing isopod and enter it through an intersegmental membrane (Wijnhoven 2001). They feed on the isopod’s hemolymph and later also on non-vital tissues, like the female ovaria (Wijnhoven 2001). The isopod continues to feed and moult normally, until the parasitoid kills it and pupates within the host exoskeleton (Wijnhoven 2001). Here, we present draft genome sequences for these three species.

**Materials and Methods**

**Samples**

Specimens of *G. legneri* were obtained from a long-term culture at the University of Nottingham (UK), which is descended from founders originally collected in Uruguay (I. Hardy, pers. comm.). Specimens of *A. bilineata* were obtained from a culture maintained at the University of Rennes (France). Specimens of *P. maculata* were cultured from *Porcellio scaber*, collected in the field in The Netherlands (Wageningen and Amsterdam) during 2015. For each species, DNA was extracted from a single adult female using different spin column protocols. For *G. legneri* and *A. bilineata*, DNA was obtained by crushing the insect in 100-
200 µl PBS and adding 100 µl nuclei lysis buffer, 4 µl Proteinase K, 5 µl RNase (all Promega) and incubating at 60°C for 15 min. A further 340 µl lysis buffer was then added and the sample was centrifuged at full speed for 5 min. The supernatant was then transferred to a spin column (Promega), rinsed with 500 µl wash buffer (Promega) three times and eluted in 100 µl H2O. *P. maculata* was crushed in liquid nitrogen and DNA was extracted using QIAamp DNA Mini Kit (Qiagen) following the manufacturer’s protocol. The quantity and quality of the DNA samples was assessed using Nanodrop and Qubit.

### Whole genome sequencing

Two Truseq fragment libraries with slightly different insert sizes were constructed from each DNA sample. The two libraries for each species were barcoded, pooled and sequenced using 2x100 bp paired-end sequencing on Illumina HiSeq 2000 (*G. legneri* and *A. bilineata*), or 2x125 bp paired-end on Illumina HiSeq 2500 (*P. maculata*).

### De novo genome assembly

Prior to de novo assembly, we characterized the raw read data using SGA (Simpson 2014) and estimated genome size using KmerGenie (Chikhi and Medvedev 2014). Furthermore, we removed co-sequenced genomes to reduce complexity of the read set. To this end, reads were mapped to the mitochondrial DNA of well-sequenced related species (*Apis mellifera, Tribolium castaneum* and *Drosophila melanogaster* for *G. legneri*, *A. bilineata* and *P. maculata*, respectively) and to a panel of 12 *Wolbachia* strains (*wAu GCA_000953315.1; wBm GCA_000008385.1; wCle GCA_000829315.1; wPip_Pel GCA_000073005.1; wHa GCA_000376605.1; wNo GCA_000376585.1; wRi GCA_000022285.1; wMel GCA_000008025.1; wFol CP015510.1; wOo GCA_000306885.1; wOv GFA_000530755.1; wPip GCA_000208785.1) using Bowtie2 (Langmead and Salzberg 2012) with default settings. Unmapped reads were extracted from the Sam file using samtools view with -f 4 (Li et al. 2009) and reverted back to fastq using bedtools bamToFastq (Quinlan and Hall 2010). Reads were trimmed using platanus_trim (Kajitani et al. 2014). Given that the results from the SGA analysis indicated high levels of heterozygosity for *A. bilineata* and *P. maculata*, we chose to...
perform the *de novo* assembly in Platanus, which is specifically geared to deal with short-read data from heterozygous genomes (Kajitani *et al.* 2014). Assembly was followed by scaffold and gap_close steps as implemented in the Platanus pipeline. To assess coverage, reads were mapped back to the assembly using Bowtie2 with default settings and coverage was calculated using bedtools.

**Annotation**

Structural annotation was performed using Maker2 using default settings (Holt and Yandell 2011). Annotation statistics were obtained using GAG (Hall *et al.* 2014). The completeness of the genomes and gene sets was assessed by identifying the number of insect Benchmark Universal Single-Copy Orthologs (BUSCOs) (Simao *et al.* 2015). BUSCO v3.0.2 was run on both the genome assembly (using -m geno) and the Maker gene set at the peptide level (using -m prot) with the insecta_odb9 lineage dataset as reference. We compared the draft genome of each species to that of its closest available relative with a chromosome- or linkage group-level genome sequence available (*Apis mellifera* Amel_4.5, *Tribolium castaneum* Tcas5.2 and *Drosophila melanogaster* release 6 for *G. legneri*, *A. bilineata* and *P. maculata*, respectively) using SyMap v4.2 (Soderlund *et al.* 2011).

**Co-sequenced genomes**

To remove any co-sequenced genomes (in addition to *Wolbachia* and mitochondria, which were removed prior to assembly), we employed the Blobology pipeline (Kumar *et al.* 2013).

**Results and Discussion**

We generated 18.6-43.6 Gb data per species, covering each genome >100x (Table 1). These data were assembled into draft genomes that were reasonably close to the predicted size for each species (Table 2). The genome size of *G. legneri* appears small compared to other sequenced genomes of Hymenoptera, but is within the range for parasitoid wasps (e.g.
Macrocentrus cingulum: 128 Mb, Fopius arisanus: 153 Mb). The estimated genome size of A. bilineata is smaller than that of other sequenced Coleoptera (smallest to date is Hypothenemus hampeii: 151 Mb). The genome size of P. maculata is within the range observed for Diptera (e.g. Zaprionus indianus 124 Mb; Rhagoletis zephyria: 795 Mb). Further work is required to establish the accuracy of our genome size estimates.

The genome assemblies presented here were obtained from outbred, heterozygous individuals. The Platanus assembler is specifically designed to handle such data and creates a mosaic of the two haplotypes (Kajitani et al. 2014). We tested this by mapping the sequence reads back to the genome assembly. Bowtie2 in default settings reports only the best alignment for each read and haplotypes that had assembled as separate contigs should have half the coverage as collapsed haplotypes, as reads would only map to one of the haplotypes. In our case, the coverage histograms were unimodal, indicating that haplotypes were successfully collapsed.

The assembly for G. legneri yielded hits to Rhabditida nematodes, some of which are known parasites of insects. A Blast search of the entomopathogenic Rhabditidid Oscheius sp. (Lephoto et al. 2016) against the G. legneri genome assembly revealed two hits, upon which we removed one contig and one partial contig from the assembly. A. bilineata contained Wolbachia, but no other co-sequenced genomes. P. maculata contained no co-sequenced genomes.

The genome assemblies are very complete, with 97-99% of BUSCOs present and only 0.4-4.7% fragmented (Figure 1B). Lacking transcriptome data and other genomic resources for these or closely related species, the structural annotation is less complete. Maker2 annotated 5588-7463 genes per genome (Table 3), which is below the expected value for eukaryotes. BUSCO analysis indicated the gene sets to be 63-85% complete (Figure 1bB). It is thus imperative for future studies to interrogate the genome assembly for genes missing from the gene set.

The level of synteny to well-characterized genome sequences of related model species varied (Figure 1C). The draft genome of G. legneri shows many collinear regions with the
genome of *A. mellifera*, while the similarity between *P. maculata* and *D. melanogaster* is limited, with intermediate collinearity between *A. bilineata* and *T. castaneum* (Figure 1C).

These differences are probably caused by a combination of factors, including quality of the draft genome assemblies, levels of relatedness to the selected model organism and rates of genome evolution.

In summary, we present fragmented, but relatively complete genome assemblies of three parasitoid insects, representing three independent evolutionary origins of the parasitoid lifestyle. These genomes will be valuable for comparisons to the widely studied parasitoid Hymenoptera, for which numerous genomes are available (Branstetter et al. 2018). Our study highlights that useful genomic resources can now be obtained from highly heterozygous individual insects collected from outbred lab cultures or even from the field, relieving the need for labor-intensive inbreeding procedures.

**Acknowledgements**

We thank Ian Hardy, Denis Poinsot, Mikael Bili, Mark Lammers and Rik Zoomer for providing insect samples.

**Literature Cited**

Bili, M., A. M. Cortesero, Y. Outreman, and D. Poinsot, 2016 Host specialisation and competition asymmetry in coleopteran parasitoids. Oecologia 182: 111–118.

Branstetter, M. G., A. K. Childers, D. Cox-Foster, K. R. Hopper, K. M. Kapheim et al., 2018 Genomes of the Hymenoptera. Curr. Opin. Insect Sci. 25: 65–75.

Carr, M., P. Mayhew, and J. P. Young, 2010 Phylogeny of bethylid wasps (Hymenoptera: Bethylidae) inferred from 28S and 16S rRNA genes. Insect Syst. Evol. 41: 55–73.

Chikhi, R., and P. Medvedev, 2014 Informed and automated k-mer size selection for genome assembly. Bioinformatics 30: 31–37.

Eggleton, P., and R. Belshaw, 2013 Insect parasitoids: an evolutionary overview. Philos.
Feener, D. H., and B. V. Brown, 1997 Diptera as parasitoids. Annu. Rev. Entomol. 42: 73–97.

Godfray, H., 1994 Parasitoids: behavioral and evolutionary ecology. Princeton University Press, Princeton.

Grenier, S., 1988 Applied biological control with Tachinid flies (Diptera, Tachinidae): A review. Anzeiger für Schädlingskd. Pflanzenschutz Umweltschutz 61: 49–56.

Hall, B., T. DeReggo, and S. Geib, 2014 GAG: the Genome Annotation Generator (Version 1.0) [Software].

Hawkins, B., and W. Sheehan, 1994 Parasitoid community ecology. Oxford Univ Press, Oxford.

Heimpel, G., and N. Mills, 2017 Biological control: ecology and applications. Cambridge Univ Press, Cambridge.

Holt, C., and M. Yandell, 2011 MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12: 491.

Kajitani, R., K. Toshimoto, H. Noguchi, A. Toyoda, Y. Ogura et al., 2014 Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 24: 1384–1395.

Khidr, S. K., S. Mayes, and I. C. W. Hardy, 2013 Primary and secondary sex ratios in a gregarious parasitoid with local mate competition. Behav. Ecol. 24: 435–443.

Kumar, S., M. Jones, G. Koutsovoulos, M. Clarke, and M. Blaxter, 2013 Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. Front. Genet. 4: 237.

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat Meth 9: 357–359.

Lephoto, T. E., P. T. Mpangase, S. Aron, and V. M. Gray, 2016 Whole genome sequence of Oscheius sp. TEL-2014 entomopathogenic nematodes isolated from South Africa. Genomics Data 7: 259–261.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan et al., 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–9.

Peters, R. S., L. Krogmann, C. Mayer, A. Donath, S. Gunkel et al., 2017 Evolutionary History of the Hymenoptera. Curr. Biol. 27: 1013–1018.
Prabhakar, K. V., G. V. Vishaka, and T. K. Narayanaswamy, 2017 Mass Production of Silkworm (Bombyx mori L.) uzi fly Exorista bombycis (Louis). Int. J. Curr. Microbiol. App. Sci. 6: 5412–5415.

Quinlan, A. R., and I. M. Hall, 2010 BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.

Simao, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov, 2015 BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212.

Simpson, J. T., 2014 Exploring Genome Characteristics and Sequence Quality Without a Reference. Bioinformatics 1–9.

Soderlund, C., M. Bomhoff, and W. M. Nelson, 2011 SyMAP v3.4: A turnkey synteny system with application to plant genomes. Nucleic Acids Res. 39:.

Visser, B., C. Le Lann, F. J. den Blanken, J. a Harvey, J. J. M. van Alphen et al., 2010 Loss of lipid synthesis as an evolutionary consequence of a parasitic lifestyle. Proc. Natl. Acad. Sci. U. S. A. 107: 8677–82.

Wajnberg, E., and S. Colazza (Eds.), 2013 Chemical ecology of insect parasitoids. Wiley, Chichester.

Werren, J. H., S. Richards, C. a Desjardins, O. Niehuis, J. Gadau et al., 2010 Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science 327: 343–8.

Wijnhoven, H., 2001 Biologie en ecologie van de Nederlandse pissebedvliegen (Diptera: Rhinophoridae). Ned. Faun. Meded. 15: 91–110.
## Tables

### Table 1 Raw reads generated for assembly

| Species                  | Read pairs | Base pairs aligned | Coverage |
|--------------------------|------------|--------------------|----------|
| *Goniozus legneri*       | 140.0 M    | 22.52 Gb           | 160.8x   |
| *Aleochara bilineata*    | 170.3 M    | 18.55 Gb           | 217.8x   |
| *Paykullia maculata*     | 211.9 M    | 43.56 Gb           | 103.4x   |

### Table 2 Assembly summary statistics

| Metric                        | *G. legneri* | *A. bilineata* | *P. maculata* |
|-------------------------------|--------------|----------------|--------------|
| GC (%)                        | 40.6         | 39.9           | 28.4         |
| Scaffold count                | 7,863        | 33,003         | 147,656      |
| Contig count                  | 13,705       | 40,228         | 169,825      |
| Total length (Mb)             | 140.1        | 85.9           | 422.4        |
| Predicted length KmerGenie (Mb) | 142         | 112            | 536          |
| Gap (%)                       | 0.18         | 1.1            | 0.29         |
| Scaffold N50 (kb)             | 167.3        | 54.1           | 7.7          |
| Contig N50 (kb)               | 37.8         | 12.1           | 5.9          |
| Max. scaffold length (Mb)     | 1.5          | 0.876          | 0.17         |
| Max. contig length (kb)       | 371.7        | 316.8          | 86.6         |
| Number of scaffolds > 50 kb   | 684          | 445            | 108          |
| % of genome in scaffolds > 50 kb | 82.4        | 52.1           | 1.8          |
Table 3 Gene annotation summary statistics

| Metric                  | A. bilineata | G. legneri | P. maculata |
|-------------------------|--------------|------------|-------------|
| Number of genes         | 7220         | 7463       | 5588        |
| Number of exons         | 28720        | 46322      | 20856       |
| Total gene length (Mb)  | 17.2         | 25.5       | 18.1        |
| Mean gene length        | 2380         | 3418       | 3232        |
| % of genome covered by genes | 20.0        | 18.2       | 4.3         |
| mean exons per mRNA     | 4            | 6          | 4           |
Figure 1 Phylogenetic context and genome features. (A) Phylogeny of selected insect species from (Visser et al. 2010). Parasitoid species are indicated in black font and non-parasitoid species in grey font. The three species for which we provide draft genome assemblies are indicated by arrows. (B) Completeness of the genome assemblies and annotation. Orthologs of 1658 genes that are present as single copies in at least 90% of insects were retrieved using Busco v3. From top to bottom: Goniozus legneri genome, G. legneri gene set, Aleochara bilineata genome, A. bilineata gene set, Paykullia maculata genome and P. maculata gene set. (C) Circle plots illustrating the synteny between the 50 largest scaffolds of the genome assemblies of G. legneri, A. bilineata and P. maculata (top half of each circle) and the genomes of the closest available relative with chromosome- or linkage group- level assemblies (A. mellifera, T. castaneum and D. melanogaster, respectively; bottom half of each circle). Links are colored to match the chromosome or linkage group of the model species.
Figure 1

A. Phylogenetic relationships among different insect species, showing the distribution of BUSCO genes.

B. Distribution of BUSCO genes in different insect species, with the x-axis showing the percentage of BUSCO genes and the y-axis showing the various species.

C. Comparison of BUSCO gene completeness between different insect species, with circular plots representing the completeness of BUSCO genes in each species.

Legend:
- Blue: Complete and single-copy
- Green: Complete and duplicated
- Yellow: Fragmented
- Red: Missing

Gene set: Assembly and Gene set

% of BUSCO genes