Hybrid genome assembly and evidence-based annotation of the egg parasitoid and biological control agent *Trichogramma brassicae*

K. B. Ferguson*, T. Kursch-Metz†,‡, E. C. Verhulst§, and B. A. Pannebakker*

* Wageningen University, Laboratory of Genetics, Wageningen, The Netherlands

† Technische Universität Darmstadt, Department of Biology, Darmstadt, Germany

‡ AMW Nützlinge GmbH, Pfungstadt, Germany

§ Wageningen University, Laboratory of Entomology, Wageningen, The Netherlands

DATA REFERENCE NUMBERS

ENA BioProject: PRJEB35413

ENA Assembly: CADCXV010000000.1

Additional GFF file, DOI: 10.6084/m9.figshare.12073833.v1

Supplementary data in DANS EASY repository, DOI: 10.17026/dans-23w-a9tn

1 **ABSTRACT**

2 *Trichogramma brassicae* (Bezdenko) are egg parasitoids that are used
3 throughout the world as biological control agents and in laboratories as model
4 species. Despite this ubiquity, few genetic resources exist beyond COI, ITS2, and
5 RAPD markers. Aided by a *Wolbachia* infection, a wild-caught strain from
6 Germany was reared for low heterozygosity and sequenced in a hybrid *de*
7 *novo* strategy, after which several assembling strategies were evaluated. The
8 best assembly, derived from a DBG2OLC-based pipeline, yielded a genome of
9 235 Mbp made up of 1,572 contigs with an N50 of 556,663 bp. Following a

rigorous *ab initio*-, homology-, and evidence-based annotation, 16,905 genes were annotated and functionally described. As an example of the utility of the genome, a simple ortholog cluster analysis was performed with sister species *T. pretiosum*, revealing over 6000 shared clusters and under 400 clusters unique to each species. The genome and transcriptome presented here provides an essential resource for comparative genomics of the commercially relevant genus *Trichogramma*, but also for research into molecular evolution, ecology, and breeding of *T. brassicae*.

INTRODUCTION

The chalcidoid *Trichogramma brassicae* (Bezdenko) (Hymenoptera: Trichogrammatidae) is a minute parasitoid wasp (~0.5 mm in length) that develops within the eggs of other insects (Smith, 1996). For over 50 years, it has been in use world-wide as a biological control agent as many lepidopteran pests of different crops are suitable hosts (Polaszek, 2009). The most common application of *T. brassicae* in Europe is against *Ostrinia nubilalis* (Hubner) (Lepidoptera: Pyralidae), the European corn borer. For example, in 2003 alone, over 11000 ha of maize in Germany was treated with *T. brassicae* (Zimmermann, 2004). It is also released against lepidopteran pests in spinach fields as well as in greenhouses (e.g. tomato, pepper, and cucumber) (Klug and Meyhöfer, 2009). With its wide application in biological control, *T. brassicae* is a well-studied species. Field trials have been conducted on several aspects, such as host location and dispersal behaviour (Suverkropp et al., 2010, 2009), overwintering ability (Babendreier et al., 2003), while other biological control related studies considered issues related to low temperature storage (Lessard

34 and Boivin, 2013), reaction to insecticides (Delpuech and Delahaye, 2013;
35 Ghorbani et al., 2016; Jamshidnia et al., 2018; Liu and Zhang, 2012; Thubru et
36 al., 2018), or risk assessment (Kuske et al., 2004).
37 Next to its application as a biological control agent, this tiny parasitoid has
38 been used in other research, both in genetic studies (Cruaud et al., 2018;
39 Laurent et al., 1998; Wajnberg, 1993) and ecological studies (Cusumano et al.,
40 2015; Fatouros and Huigens, 2012; Huigens et al., 2009). In addition, several
41 initiatives investigate the infection of *T. brassicae* with *Wolbachia* bacteria
42 (Ivezić et al., 2018; Poorjavad et al., 2012) and the consequences of such an
43 infection (Farrokhi et al., 2010; Poorjavad et al., 2018; Rahimi-Kaldeh et al.,
44 2018). As *T. brassicae* is a cryptic species with several other congenerics,
45 misidentification and misclassification is a known issue (Polaszek, 2009). In
46 response, molecular identification of trichogrammatids is well studied and
47 established (Ivezić et al., 2018; Rugman-Jones and Stouthamer, 2017;
48 Stouthamer et al., 1999; Sumer et al., 2009). Recently, several RADseq libraries
49 were constructed from single *T. brassicae* wasps to aide in resolving the
50 aforementioned phylogenetic issues within *Trichogramma* (Cruaud et al.,
51 2018). Otherwise, the genomics of *T. brassicae* have largely been neglected
52 even though a well annotated genome would allow researchers and
53 biological control practitioners access to a wealth of information and open
54 new avenues for comparative genomics and transcriptomics for evolutionary,
55 ecological, and applied research.
56 Here, we report the whole-genome sequencing and annotation of a *T.*
57 *brassicae* strain infected by *Wolbachia* that had thelytokous reproduction, in

58 which females arise from unfertilized eggs. A hybrid *de novo* sequencing
59 strategy was chosen to address two common issues: we used long PacBio
60 Sequel reads to bridge the large segments of repetitive sequences often found
61 in Hymenoptera, while countering the error bias of long read technology with
62 the accuracy of Illumina short reads. A similar strategy was recently applied to
63 improve the *Apis mellifera* genome, where the long PacBio reads were the
64 backbone that boosted the overall contiguity of the genome, alongside the
65 incorporation of repetitive regions (Wallberg et al., 2019).
66 In this report, we present the hybrid *de novo* genome of *T. brassicae*. Three
67 different assemblers were evaluated, and the most complete genome
68 assembly was used for decontamination and *ab initio*-, homology-, and
69 evidence-based annotation. The resulting annotation was functionally
70 described using gene ontology analysis. Finally, a heterozygosity comparison
71 and simple ortholog cluster analysis with the congeneric *T. pretiosum* was
72 performed, which can be considered a starting point for future comparative
73 genomics of the commercially important genus *Trichogramma*.

74 **METHODS**

75 **Species origin and description:**

76 Individuals of *Trichogramma brassicae* were acquired by AMW Nützlinge
77 GmbH (Pfungstadt, Germany). The strain was baited in May 2013 in an apple
78 orchard near Eberstadt, Germany. The orchard was surrounded by blackberry
79 hedges, forest, and other orchards. For baiting, the eggs of *Sitotroga cerealella*
80 (Olivier) (Lepidoptera: Gelechiidae) (Mega Corn Ltd., Bulgaria) were glued on
81 paper cards (AMW Nützlinge GmbH, Germany), usually used for releasing

82 *Trichogramma* sp. in corn fields and households. These cards were placed
83 directly into the trees, approximately two meters above ground. After five days
84 in the field, baiting cards were collected and incubated together at 25°C.
85 Following emergence, individuals were kept together, offered *S. cerealella*
86 eggs, and reared in a climate chamber (27±2°C, L:D=24:0h for four days, then
87 transferred to16±2°C, L:D=0:24h until emergence).
88 In 2016, the offspring of twenty isolated females were transferred to
89 Wageningen University (The Netherlands) to be reared for low heterozygosity.
90 The resulting offspring were reared in a single general population on irradiated
91 *Ephestia kuehniella* (Zeller) (Lepidoptera: Pyralidae) eggs as factitious hosts
92 under laboratory conditions in a climate chamber (20 ± 5˚C, RH 50 ± 5%,
93 L:D=12:12 h). *Wolbachia* presence was determined following the PCR
94 amplification protocol of Zhou et al 1998 in a presence/absence assessment
95 with known positive and negative control samples (Zhou et al., 1998). Natural
96 *Wolbachia* infections have previously been detected in Iranian populations of
97 *T. brassicae* (Farrokhi et al., 2010), but none of the Eurasian populations have
98 been known to support this symbiosis (Stouthamer, 1997; Stouthamer and
99 Huigens, 2003). Sequence coverage was calculated using the previously
100 established genome size estimate for *T. brassicae* of 246 Mbp (Johnston et al.,
101 2004).
102 **Isofemale line:**
103 Following confirmation of *Wolbachia* infection (Supplementary materials
104 S1.1.1), a single female from the general population was isolated (generation
105 0, G0), and given eggs *ad libitum*. In the resulting generation (G1), unmated

106 females were isolated and reared with eggs *ad libitum*. Offspring of the initial
107 isolations G0 and G1 were confirmed to be entirely female, suggesting
108 thelytokous parthenogenetic reproduction. Combined with isolating single
109 females, this maximizes genetic similarity of the following generation (G2) of
110 these G1 females. One of these G2 strains, S301, was boosted for multiple
111 generations over the period of one year. By the time of collection for
112 sequencing, both the S301 and general population no longer harboured
113 *Wolbachia* at detectable levels (Supplementary materials S1.1.2).

114 **gDNA extraction:**

115 Three separate extractions were prepared in 1.5 mL safelock tubes with each
116 several hundred *Trichogramma brassicae*. The tubes were frozen in liquid
117 nitrogen with approximately six 1-mm glass beads and shaken for 30 s in a
118 Silamat S6 shaker (Ivoclar Vivadent, Schaan, Liechtenstein). DNA was then
119 extracted using the Qiagen MagAttract Kit (Qiagen, Hilden, Germany).
120 Following an overnight lysis step with Buffer ATL and proteinase K at 56˚C,
121 extraction was performed according to the MagAttract Kit protocol. Elutions
122 were performed in two steps with Buffer AE (Tris-EDTA) each time (first 60 µL,
123 then 40 µL), yielding 100 µL. The two extractions yielding the largest amount of
124 DNA (5.49 µg and 8.24 µg) were combined for long-read sequencing, while the
125 remaining extraction (1.67 µg) was used for short-read sequencing. DNA
126 concentration was measured with an Invitrogen Qubit 2.0 fluorometer using the
127 dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, USA) while fragment
128 length was confirmed on gel.

129 **Library preparation and sequencing:**

130 Library preparation and sequencing was performed by Novogene
131 Bioinformatics Technology Co., Ltd., (Beijing, China). For Illumina sequencing,
132 gDNA was used to construct one paired-end (PE) library according to the
133 standard protocol for Illumina with an average insert size of 150 bp and was
134 sequenced using an Illumina HiSeq 2000 (Illumina, San Diego, USA). For Single
135 Molecule Real Time (SMRT) sequencing, gDNA was selected for optimal size
136 using a Blue Pippin size selection system (Sage Science, Beverley, USA) following
137 a standard library preparation. The library was then sequenced on a PacBio
138 Sequel (Pacific Biosciences, Menlo Park, USA) with 16 SMRT cells.

139 **Assembly and decontamination:**

140 Prior to assembly, Illumina reads were assessed for quality using *FASTQC*
141 (Andrews et al., 2015), then trimmed for quality in *CLC Genomics Workbench*
142 *11* using default settings (Qiagen). Trimmed Illumina reads were paired for
143 subsequent analysis.

144 In order to achieve the best possible assembly, three assembly pipelines were
145 evaluated: one for PacBio-only reads and two hybrid assemblers. The PacBio-
146 only were assembled with *Canu* (v1.6) with modifications based on PacBio
147 Sequel reads (`correctedErrorRate=0.085 corMhapSensitivity=normal`
148 `alongside corMhapSensivity=normal`) (Koren et al., 2017). This is assembly
149 version v1.0 in the subsequent discussion.

150 The first hybrid assembly pipeline using both long and short sequencing read
151 sets was *SPAdes* (v3.11.1) (Bankevich et al., 2012). The *SPAdes* genome toolkit
152 supports hybrid assemblies with the *hybridSPAdes* algorithm (Antipov et al.,
153 2016). Three iterations of the *SPAdes* pipeline were run with varying k-mer sizes

154 resulting in three different assembly versions: 21, 33, 55 (default, v2.1); k-mer sizes
155 21, 33, 55, 77 (v2.2); and a single k-mer size of 127 (v2.3).
156 The second hybrid assembly pipeline was *DBG2OLC* (Ye et al., 2016). The
157 *DBG2OLC* pipeline can be readily tweaked with other programs depending on
158 the job (Chakraborty et al., 2016). Following the *DBG2OLC* pipeline, de Bruijn
159 graph contigs were generated using *SparseAssembler* using default settings
160 and setting the expected genome size to 750 Mbp to ensure a genome size
161 output that is unrestricted (Ye et al., 2012). Contigs were transformed into read
162 overlaps using *DBG2OLC* with settings suggested for large genomes and
163 PacBio Sequel data (`k=17; AdaptiveTh=0.01; KmerCovTh=2;`
164 `MinOverlap=20; RemoveChimera=1`), according to the *DBG2OLC* manual
165 (https://github.com/yechengxi/DBG2OLC). This creates an assembly
166 backbone of the best overlaps between the short-read de Bruijn contigs and
167 the long reads. *minimap2* (v2.9) and *Racon* (v1.0.2) were used for consensus
168 calling remaining overlaps to the assembly backbone (Li, 2018; Vaser et al.,
169 2017). The resulting consensus assembly was polished twice using the Illumina
170 reads with *Pilon* (v1.22) (Walker et al., 2014). This final assembly is v3.0 in
171 subsequent discussion.
172 The best of the five assemblies generated was determined on the basis of N50,
173 genome size, and completeness (Table 1). Genome statistics such as N50,
174 number of contig, and genome size were determined using *Quast* (Gurevich
175 et al., 2013). Assembly completeness was assessed using *BUSCO* (v3.0.2) with
176 the insect_odb9 ortholog set and the fly training parameter (Simão et al., 2015).
177 Based on these characteristics, the decision was made to move forward with

178 assembly v3.0, which was then decontaminated for microbial sequences using
179 NCBI *BLASTn* (v2.2.31+) against the NCBI nucleotide collection (nr).

180 ***Wolbachia* contamination:**

181 Two contigs contained a large amount of *Wolbachia* content, with over 80%
182 of the scaffold containing material with 75% or higher homology to *Wolbachia*.
183 These contigs were assessed for homology against the NCBI nucleotide
184 collection (nr) and removed from the assembly (Supplementary material S1.2).
185 Post-decontamination, the assembly is referred to as v3.5.

186 **RNA extraction, library construction, and sequencing:**

187 *T. brassicae* wasps from the S301 line were collected for RNAseq for evidence-
188 based annotation. Hundreds of adult individuals (male and female) were killed
189 by freezing at -80˚C, then frozen in liquid nitrogen in a single 1.5 mL safelock
190 tube with approximately six 1-mm glass beads and shaken for 30 s in a Silamat
191 S6 shaker (Ivoclar Vivadent). The RNeasy Blood and Tissue Kit (Qiagen) was
192 used according to manufacturer's instructions, and final column elution was
193 achieved using 60 µL sterilized water. The sample was measured for quality and
194 RNA quantity using an Invitrogen Qubit 2.0 fluorometer and the RNA BR Assay
195 Kit (Thermo Fisher Scientific). The RNA sample was then processed by
196 Novogene Bioinformatics Technology Co., Ltd., (Beijing, China) using poly(A)
197 selection followed by cDNA synthesis with random hexamers and library
198 construction with an insert size of 300 bp. Paired-end sequencing was
199 performed on an Illumina HiSeq 4000 according to manufacturer's instruction.
200 Quality filtering was applied to remove adapters, reads with more than 10%

201 undetermined bases, and reads of low quality for more than 50% of the total
202 bases (Qscore less than or equal to 5).

203 ***Ab initio* gene finding, transcriptome assembly, and annotation:**

204 For the *ab initio* gene finding, a training set was established using the reference
205 genome of *Drosophila melanogaster* (Meigen) (Diptera: Drosophilidae)
206 (Genbank: GCA_000001215.4; Release 6 plus ISO1 MT) and the associated
207 annotation (Adams et al., 2000; Dos Santos et al., 2015). The training parameters
208 were used by *GlimmerHMM* (v3.0.1) for gene finding in the *T. brassicae* genome
209 assembly v3.5 (Majoros et al., 2004). For homology-based gene prediction,
210 *GeMoMa* v1.6 was used with the *D. melanogaster* reference genome
211 alongside our RNAseq data as evidence for splice site prediction (Keilwagen
212 et al., 2016). For evidence-based gene finding, the pooled RNAseq data was
213 mapped to the to the *T. brassicae* genome separately with *TopHat* (v2.0.14)
214 with default settings (Trapnell et al., 2009). After mapping, *Cufflinks* (v2.2.1) was
215 used to assemble transcripts (Trapnell et al., 2010). *CodingQuarry* (v1.2) was
216 used for gene finding in the genome using the assembled transcripts, with the
217 strandness setting set to 'unstranded' (Testa et al., 2015).

218 The tool *EVidenceModeler (EVM)* (v1.1.1) was used to combine the *ab initio*,
219 homology-based, and evidence-based information, with evidence-based
220 weighted 1, *ab initio* weighted 2, and homology-based weighted 3 (Haas et al.
221 2008). We annotated the predicted proteins with *BLASTp* (v2.2.31+) on a
222 custom database containing all SwissProt and Refseq genes of *D.*
223 *melanogaster* (Acland et al., 2014; Boutet et al., 2008; Camacho et al., 2009),

224 followed by an additional search in the NCBI non-redundant protein database
225 (nr) to obtain additional homology data.

226 **GO term analysis:**

227 A list of genes was constructed for Gene Ontology (GO) term classification by
228 deduplicating the annotated proteins and removing the non-annotated
229 proteins. These accession IDs were converted into UniProtKB accession IDs using
230 the UniProt ID mapping feature and deduplicated a final time (Boutet et al.,
231 2008). These UniProtKB accession IDs were in turn used with the *DAVID 6.8*
232 *Functional Annotation Tool* to assign GO terms to each accession ID with the
233 *D. melanogaster* background and generate initial functional analyses (Huang
234 et al., 2009a, 2009b) (see supplementary S1.3 for *DAVID* input list).

235 **Heterozygosity estimates:**

236 The heterozygosity of the S301 line was assessed using sequence reads and k-
237 mer counting, and compared to the congeneric *Trichogramma pretiosum*
238 (Riley) (Hymenoptera: Trichogrammatidae), for which sequence data exists for
239 both a thelytokous (asexual) *Wolbachia*-infected strain as well as an inbred
240 arrhenotokous (sexual) line (Lindsey et al., 2018). Using *jellyfish* (v2.3.0) to count
241 k-mers, the same trimmed and paired Illumina reads used for assembly were
242 assessed using the default k-mer size of 21 (m=21), with results exported to a
243 histogram (Marçais and Kingsford, 2011). This histogram file was then used with
244 *GenomeScope* (v1.0) to estimate heterozygosity of the reads based on a
245 statistical model, where a Poisson distribution is expected for a homozygous
246 sample while a bimodal distribution is expected for a homozygous distribution
247 (Vurture et al., 2017). This genome profiling gives a reliable estimate for

248 heterozygosity as well as estimates of repetitive content. The same *jellyfish* and
249 *GenomeScope* analyses were performed on *T. pretiosum* short-read sequence
250 data for the thelytokous strain (NCBI SRA database, SRR1191749) and the
251 arrhenotokous line (SRR6447489), with adaptions for reported insert sizes
252 (Lindsey et al., 2018).

253 **Ortholog cluster analysis:**

254 The complete gene set of *T. brassicae* was compared to that of *T. pretiosum*
255 (Lindsey et al., 2018), which was retrieved from the i5K Workspace (Poelchau
256 et al., 2016). An ortholog cluster analysis was performed on both gene sets via
257 *OrthoVenn2* with the default settings of E-values of 1e-5 and an inflation value
258 of 1.5 (Xu et al., 2019). For *T. brassicae* protein set, see supplementary materials
259 S1.5.

260 **Data availability:**

261 All sequence data are available at the EMBL-ENA database under BioProject
262 PRJEB35413, including assembly (CADCXV010000000.1). An additional,
263 complete annotation file (.gff) is also available (Ferguson, 2020). Additional
264 data, such as gel images, the *Wolbachia* contaminated contigs, input gene list
265 for *DAVID*, *GenomeScope* images, and complete protein set are available via
266 the supplementary materials.

267 **RESULTS AND DISCUSSION**

268 **Sequencing, assembly, and decontamination:**

269 Sequencing of the Illumina 150 bp paired-end library yielded 80,489,816 reads.
270 After quality filtering and trimming, 80,483,128 paired-end reads were retained.
271 Sequencing the PacBio Sequel library yielded 2,500,204 subreads with an

272 average length of 6377 bp. The genome size estimate for *T. brassicae* is 246
273 Mbp (Johnston et al., 2004) indicating that short-read coverage was 98x while
274 long-read coverage was 64x, resulting in a total coverage of 162x. Three
275 assembly pipelines were used, resulting in five potential assemblies where one,
276 v3.0, was eventually selected for further use. Results of these assemblies are
277 detailed in Table 1.
278 The first draft assembly generated with Canu with the altered settings for PacBio
279 Sequel data resulted in an assembly of approximately 70 Mbp in size, drastically
280 smaller than the 246 Mbp expected, and contained a total of 3,007 contigs
281 with an N50 of 27,303. The longest contig was 126,800 bp in size.
282 The second assembly strategy relied on hybrid assembly pipelines, and *SPAdes*
283 was used with the default k-mer settings, which resulted in an assembly of
284 approximately 227 Mbp in size with an N50 of 36,870 and a *BUSCO*
285 completeness of 96.8%. Three different assembly runs were done with differing
286 k-mer sizes: the default k-mer sizes of 21, 33, 55 (v2.1); default k-mer sizes plus 77
287 (v2.2); or the highest possible k-mer size of 127 (v2.3). Increasing the k-mer size
288 only improved N50 scores to a point, along with decreasing the number of
289 contigs, and stable *BUSCO* scores, however, the assembled genome size drops
290 dramatically with the third attempt shrinking down to 211 Mbp. Based on
291 *BUSCO* scores and N50 alone, the second *SPAdes* attempt, v2.2, would be the
292 best of the three, though all three are similar in most measures.
293 The third assembly strategy used the *DGB2OLC+Racon+Pilon* pipeline, which
294 resulted in assembly v3.0. Here, there is a large difference compared to the
295 previous *SPAdes* assemblies. Particularly, the number of contigs is reduced

296 dramatically from the 70,000 to 280,000 range of the *SPAdes* output down to a
297 mere 1,572. Meanwhile, the assembled genome size is now 235 Mbp and with
298 an N50 of 556,663 and a BUSCO score of 95.5%. The full completeness score for
299 this assembly, using the 1658 BUSCO groups within the insect_od09 BUSCO set,
300 returned 1531 (92.3%) complete and single-copy BUSCOs, 53 (3.2%) complete
301 and duplicated BUSCOs, 22 (1.3%) fragmented BUSCOs, and 52 (3.2%) missing
302 BUSCOs (Simão et al., 2015).
303 While the PacBio-only assembly in *Canu* could have been improved using
304 different settings or additional tools, we decided to focus on using the
305 additional sequence information of the Illumina reads in the subsequent hybrid
306 assembly strategies. The *SPAdes* assemblies (v2.1-3) were already decent but
307 could have been further improved using *Pilon*, a tool that improves assemblies
308 at the base pair level using high quality Illumina data. However, the v3.0
309 assembly was by far the best assembly based on assembled genome size, N50,
310 and BUSCO scores and therefore we chose this strategy for our *T. brassicae*
311 genome assembly.
312 Decontamination of this assembly (v3.0) resulted in the removal of two contigs
313 as the homology analysis using *BLASTn* with the NCBI nr database indicated
314 that both contigs were confirmed to be largely composed of *Wolbachia*
315 genomic content. Contig "Backbone_1176" is 9,448 bp in length and two areas
316 of the contig, representing over 80% of its length, showed high homology to
317 *Wolbachia*. Similarly, contig "Backbone_1392" is 17,350 bp and three separate
318 areas representing over 80% showed similar levels of homology to *Wolbachia*
319 After decontamination this final assembly (v3.5) was used for annotation.

320 ***Ab initio* gene finding, transcriptome assembly, and annotation:**

321 In our RNA sequencing experiment, we generated 26,479,830 150bp paired-
322 end cDNA reads. Filtering the reads for quality retained 99.3% of these reads to
323 be used for evidence-based gene finding via transcriptome assembly.

324 The annotations from the evidence-based gene finding were used alongside
325 homology-based findings and *ab initio* annotations in a weighted model,
326 resulting in a complete annotation for the assembly. In 865 mRNA tracks,
327 representing approximately 5.1% of the official gene set, a gene model could
328 not be annotated via the SwissProt database, and these tracks are named
329 "No_blast_hit." The majority of tracks are annotated with reference to SwissProt
330 or GenBank accession number of the top *BLASTp* hit.

331 Transcriptome assembly and mapping resulted in 45,876,158 mapped
332 transcripts (48,327,134 total). *CodingQuarry* predicted 45,454 evidence-based
333 genes from these mapped transcripts, while *ab initio* gene finding using
334 *GlimmerHMM* resulted in 16,877 genes and homology-based gene finding with
335 *GeMoMa* resulted in 6,675 genes. The final complete gene set was created
336 using *EVidenceModeler*, where a weighted model using all three inputs
337 resulted in a complete gene set of 16,905 genes.

338 **GO term analysis:**

339 The complete gene set of 16,905 genes was deduplicated and genes with no
340 correlating *BLASTp* hit were removed from this analysis. The remaining 9,373
341 genes were subjected to UniProtKB ID mapping, resulting in 8,247 genes with a
342 matching ID after another round of deduplication (828 duplicates found). The

343 remaining 755 accession IDs were not able to be matched, half of which are
344 obsolete proteins within the UniParc database (377).
345 The *DAVID Functional Annotation Tool* used 6,585 genes for the analysis and
346 showed that 80.8% (5,320) contribute to 530 biological processes, 77.5% (5,104)
347 contribute to 115 different cellular component categories, and 74.2% (4,889)
348 contribute to 93 molecular functions (genes can code to multiple GO terms).
349 The remaining 1,662 genes are uncategorised.
350 **Heterozygosity estimates:**
351 Using short-read data and k-mer counting, heterozygosity was estimated for
352 our isofemale S301 line and compared to both a parthenogenesis inducing
353 *Wolbachia*-infected strain and an arrhenotokous line of *T. pretiosum* (Lindsey et
354 al., 2018). The average estimated heterozygosity for our S301 *T. brassicae* line is
355 0.0332% with approximately 0.608% repetitive content (for full details, see Table
356 2). This is similar to the thelytokous *T. pretiosum* line, which has a slightly lower
357 estimated heterozygosity (0.0289%) and a lower amount of repetitive content
358 (0.482%). Both have a very distinct Poisson distribution, indicating a low
359 heterozygosity (Figure S1.4.1-2). The arrhenotokous *T. pretiosum* showed a
360 higher estimated heterozygosity (0.863%), a larger amount of repetitive content
361 (2.64%), and a slightly bimodal distribution (Figure S1.4.3). The fact that both
362 thelytokous *Trichogramma* species have a similar low level of heterozygosity
363 when compared to the arrhenotokous *T. pretiosum* suggests that in both cases
364 *Wolbachia* infection had a severe effect on genetic diversity. As the canonical
365 mechanism of parthenogenesis-induction in other *Wolbachia* infected
366 thelytokous *Trichogramma* species is gamete duplication (Pannebakker et al.,

367 2004; Stouthamer and Kazmer, 1994), in which unfertilized eggs are diploidized
368 and results in fully homozygous progeny in a single generation, the low genomic
369 heterozygosity rate suggests a similar mechanism for *Wolbachia*-induced
370 parthenogenesis in *T. brassicae*. However, the involvement of *Wolbachia* in
371 causing all-female offspring in this *T. brassicae* strain and the presence and
372 mechanisms of *Wolbachia* in other thelytokous *T. brassicae* strains (Farrokhi et
373 al., 2010; Poorjavad et al., 2018, 2012) does require further investigation.

374 **Ortholog cluster analysis:**

375 The complete gene set of *T. brassicae* was compared to that of *T. pretiosum*
376 using *OrthoVenn2* (full output in Table 3). Both species have a similar range of
377 proteins (16,905 in *T. brassicae* and 13,200 in *T. pretiosum*) that form a similar
378 number of clusters (6,537 in *T. brassicae* and 6,489 in *T. pretiosum*). The two
379 species share 6,158 clusters (of 16,899 proteins), while *T. brassicae* has 379
380 unique clusters (1,726 proteins) and *T. pretiosum* has 331 unique clusters (1,005
381 proteins), as shown in Figure 1. These unique clusters account for approximately
382 5% of the entire cluster set for both species, and may both indicate true areas
383 of differentiation, or result from differences in the annotation strategies. There is
384 a similar amount of singleton clusters (proteins that do not cluster with others) in
385 *T. brassicae* (5,291) and *T. pretiosum* (5,184). Both the unique clusters and the
386 unique single-copy genes could be novel proteins, regions of contamination,
387 evidence of unique horizontal gene transfer, or pseudogenes. More
388 investigation into these protein clusters in addition to a more comprehensive
389 manual annotation should shed some light on the differences between these
390 closely related yet geographically distinct parasitoid wasps.

391 **CONCLUSIONS AND PERSPECTIVES**

392 Here, we present the genome of biological control agent *Trichogramma*
393 *brassicae*, a chalcidoid wasp used throughout the world for augmentative
394 biological control as well as genetic and ecological research. This unique strain
395 hosted a parthenogenesis-inducing *Wolbachia* infection and is the first
396 European *Trichogramma* genome to be published, allowing for comparative
397 analyses with other *Trichogramma* genomes, as we have shown. Our genomic
398 data also illuminates the possible mechanism of parthenogenesis-induction by
399 *Wolbachia* in this strain. Furthermore, the variety of genomic and transcriptomic
400 data generated for this genome provide much-need resources to bring *T.*
401 *brassicae* into the -omics era of biological research.

402 A hybrid approach was used, resulting in a highly contiguous assembly of 1,572
403 contigs and 16,905 genes based on *ab initio*, homology-based, and evidence-
404 based annotation, for a total assembly size of 235 Mbp. Two scaffolds were
405 identified that were of *Wolbachia* origin and removed. Ortholog cluster analysis
406 showed 379 unique protein clusters containing 1,726 proteins. Future studies are
407 needed to show whether these clusters are truly unique. This genome and
408 annotation provides the basis for future, more in-depth comparative studies
409 into the genetics, evolution, ecology, and biological control use of
410 *Trichogramma* species.

411 **ACKNOWLEDGEMENTS AND FUNDING**

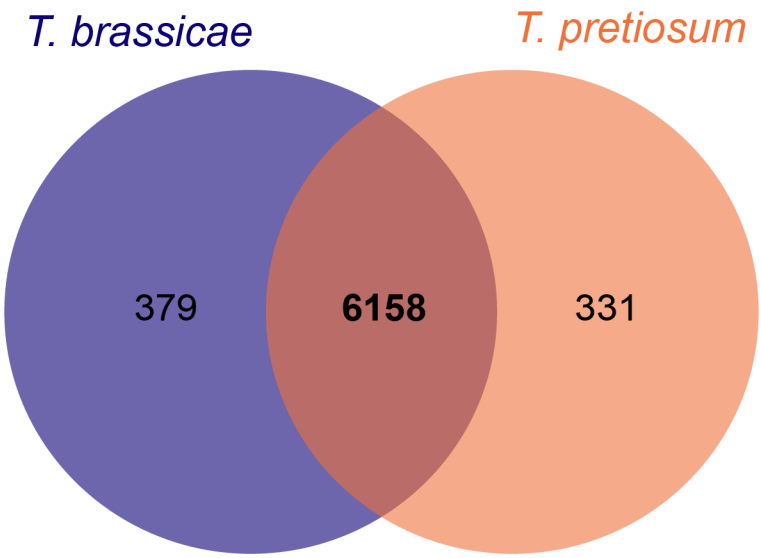
412 We would like to thank Bernd Wührer (AMW Nützlinge) for providing access to
413 specimens; Gabrielle Bukovinszkine Kiss, José van de Belt, Frank Becker
414 (Wageningen University), and Lorraine Latchoumane for assistance with

415 rearing, DNA, and RNA extraction; Richard Stouthamer and Nina Fatouros for
416 sharing their *Trichogramma* knowledge; and Sophie Chattington, Andra Thiel
417 (University of Bremen), and Bas Zwaan (Wageningen University) for their
418 assistance in this project. GenomeScan B. V. performed decontamination and
419 annotation analysis. This work has received funding from the European Union
420 Horizon 2020 research and innovation programme under the Marie
421 Skłodowska-Curie grant agreement no. 641456.

422 **SUPPLEMENTARY MATERIALS**

423 Additional supplementary material from this study is available in an attached
424 document, with some material found on the DANS EASY Repository,
425 https://doi.org/10.17026/dans-23w-a9tn (explanation within supporting
426 document).

427 FIGURES



428

429 Figure 1 Ortholog clusters analysis between *Trichogramma brassicae* and *T.*

430 *pretiosum* using OrthoVenn2 (Xu et al., 2019). The number of clusters shared

431 between the two organisms is in bold.

432

433 TABLES

Table 1. Statistics for five assemblies of *Trichogramma brassicae*. The first strategy was PacBio-only in *Canu*, while three hybrid assembly strategies were based on *SPAdes* and modulating k-mer sizes, and an additional hybrid assembly was based on an adapted *DBG2OLC+Racon+Pilon* protocol. BUSCO score is based on the insect_db09 dataset (Simão et al., 2015).

Assembler	**Version**	**Size** (bp)	**Contigs**	**Longest contig** (bp)	**N50** (bp)	**BUSCO** (Complete %)
Canu	v1.0	69,522,446	3,007	126,800	27,303	18.7
SPAdes (k=21, 33, 55)	v 2.1	227,096,967	282,988	474,998	36,870	96.8
SPAdes (k=21,33, 55, 77)	v 2.2	226,864,253	189,696	548,753	49,096	97.1
SPAdes (k=127)	v 2.3	211,402,326	73,567	537,817	63,558	96.4
DBG2OLC+ Racon+Pilon	v 3.0	235,413,774	1,572	2,953,580	556,663	95.5

434

435

Table 2. Heterozygosity and repetitive content analysis of *Trichogramma brassicae* (thelytokous), *Trichogramma pretiosum* (thelytokous), and *T. pretiosum* (arrhenotokous) lines based on sequence data.

	Heterozygosity (%)	**Repetitive content** (%)	**Source of sequence data**
T. brassicae, thelytokous S301 line	0.0332	0.608	This publication
T. pretiosum, thelytokous *Wolbachia* line	0.0289	0.482	Lindsey et al., 2018
T. pretiosum, arrhenotokous inbred line	0.863	2.64	Lindsey et al., 2018

436

437

Table 3. Output of OrthoVenn2 ortholog cluster analysis of *Trichogramma brassicae* and *Trichogramma pretiosum*.

Species	**Proteins**	**Clusters**	**Singletons**	**Source of gene set**
T. brassicae	16,905	6,537	5,291	This work (S1.5)
T. pretiosum	13,200	6,489	5,184	Lindsey et al., 2018; Poelchau et al., 2015

438

LITERATURE CITED

Acland, A., Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bryant, S.H., Canese, K., Church, D.M., Clark, K., Dicuccio, M., Dondoshansky, I., Federhen, S., Feolo, M., Geer, L.Y., Gorelenkov, V., Hoeppner, M., Johnson, M., Kelly, C., Khotomlianski, V., Kimchi, A., Kimelman, M., Kitts, P., Krasnov, S., Kuznetsov, A., Landsman, D., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Karsch-Mizrachi, I., Murphy, T., Ostell, J., O'Sullivan, C., Panchenko, A., Phan, L., Pruitt, D.P.K.D., Rubinstein, W., Sayers, E.W., Schneider, V., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Siyan, K., Slotta, D., Soboleva, A., Soussov, V., Starchenko, G., Tatusova, T.A., Trawick, B.W., Vakatov, D., Wang, Y., Ward, M., John Wilbur, W., Yaschenko, E., Zbicz, K., 2014. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 42, 8–13. https://doi.org/10.1093/nar/gkt1146

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H.C., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P. V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R.,

463 Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H.,
464 Cadieu, E., Center, A., Chandra, I., Michael Cherry, J., Cawley, S., Dahlke,
465 C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z.,
466 Deslattes Mays, A., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M.,
467 Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C.,
468 Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg,
469 N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Harley Gorrell, J., Gu,
470 Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez,
471 J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H.,
472 Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A.,
473 Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai,
474 Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei,
475 B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.
476 V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B.,
477 Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
478 Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard,
479 J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D.C.,
480 Scheeler, F., Shen, H., Christopher Shue, B., Siden-Kiamos, I., Simpson, M.,
481 Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R.,
482 Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X.,
483 Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams,
484 S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Alison Yao, Q., Ye, J.,
485 Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
486 Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A.,

487 Myers, E.W., Rubin, G.M., Craig Venter, J., 2000. The genome sequence of
488 Drosophila melanogaster. Science (80-.). 287, 2185–2195.
489 https://doi.org/10.1126/science.287.5461.2185
490 Andrews, S., Krueger, F., Seconds-Pichon, A., Biggins, F., Wingett, S., 2015.
491 FastQC. A quality control tool for high throughput sequence data.
492 Babraham Bioinformatics. Babraham Inst.
493 Antipov, D., Korobeynikov, A., McLean, J.S., Pevzner, P.A., 2016. HybridSPAdes:
494 An algorithm for hybrid assembly of short and long reads. Bioinformatics
495 32, 1009–1015. https://doi.org/10.1093/bioinformatics/btv688
496 Babendreier, D., Rostas, M., Höfte, M.C.J., Kuske, S., Bigler, F., 2003. Effects of
497 mass releases of Trichogramma brassicae on predatory insects in maize.
498 Entomol. Exp. Appl. 108, 115–124. https://doi.org/10.1046/j.1570-
499 7458.2003.00075.x
500 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S.,
501 Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V, Sirotkin,
502 A. V, Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a
503 new genome assembly algorithm and its applications to single-cell
504 sequencing. J. Comput. Biol. 19, 455–77.
505 https://doi.org/10.1089/cmb.2012.0021
506 Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bairoch, A., 2008.
507 UniProtKB/Swiss-Prot: The manually annotated section of the UniProt
508 KnowledgeBase. Methods Mol. Biol. 406, 89–112.
509 https://doi.org/10.1007/978-1-59745-535-0
510 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer,

511 K., Madden, T.L., 2009. BLAST+: Architecture and applications. BMC
512 Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421
513 Chakraborty, M., Baldwin-Brown, J.G., Long, A.D., Emerson, J.J., 2016.
514 Contiguous and accurate de novo assembly of metazoan genomes with
515 modest long read coverage. Nucleic Acids Res. 44, 1–12.
516 https://doi.org/10.1093/nar/gkw654
517 Cruaud, A., Groussier, G., Genson, G., Sauné, L., Polaszek, A., Rasplus, J.-Y.,
518 2018. Pushing the limits of whole genome amplification: successful
519 sequencing of RADseq library from a single microhymenopteran
520 (Chalcidoidea, *Trichogramma*). PeerJ 6, e5640.
521 https://doi.org/10.7717/peerj.5640
522 Cusumano, A., Weldegergis, B.T., Colazza, S., Dicke, M., Fatouros, N.E., 2015.
523 Attraction of egg-killing parasitoids toward induced plant volatiles in a
524 multi-herbivore context. Oecologia 179, 163–174.
525 https://doi.org/10.1007/s00442-015-3325-3
526 Delpuech, J.M., Delahaye, M., 2013. The sublethal effects of deltamethrin on
527 Trichogramma behaviors during the exploitation of host patches. Sci. Total
528 Environ. 447, 274–279. https://doi.org/10.1016/j.scitotenv.2012.12.096
529 Dos Santos, G., Schroeder, A.J., Goodman, J.L., Strelets, V.B., Crosby, M.A.,
530 Thurmond, J., Emmert, D.B., Gelbart, W.M., Brown, N.H., Kaufman, T.,
531 Werner-Washburne, M., Cripps, R., Broll, K., Gramates, L.S., Falls, K.,
532 Matthews, B.B., Russo, S., Zhou, P., Zytkovicz, M., Adryan, B., Attrill, H.,
533 Costa, M., Marygold, S., McQuilton, P., Millburn, G., Ponting, L., Stefancsik,
534 R., Tweedie, S., Grumbling, G., 2015. FlyBase: Introduction of the

535 Drosophila melanogaster Release 6 reference genome assembly and
536 large-scale migration of genome annotations. Nucleic Acids Res. 43,
537 D690–D697. https://doi.org/10.1093/nar/gku1099
538 Farrokhi, S., Ashouri, A., Shirazi, J., Allahvari, H., Huigens, M.E.E., 2010. A
539 comparative study on the functional response of Wolbachia-infected
540 and uninfected forms of the parasitoid wasp Trichogramma brassicae. J.
541 Insect Sci. 10, 167. https://doi.org/10.1673/031.010.14127
542 Fatouros, N.E., Huigens, M.E., 2012. Phoresy in the field: natural occurrence of
543 Trichogramma egg parasitoids on butterflies and moths. BioControl 57,
544 493–502. https://doi.org/10.1007/s10526-011-9427-x
545 Ferguson, K., 2020 Trichogramma brassicae hybrid genome annotation file.
546 https://doi.org/10.6084/m9.figshare.12073833.v1
547 Ghorbani, M., Saber, M., Bagheri, M., Vaez, N., 2016. Effects of diazinon and
548 fipronil on different developmental stages of Trichogramma brassicae
549 Bezdenko (Hym.; Trichogrammatidae). J. Agric. Sci. Technol. 18, 1267–
550 1278.
551 Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: Quality
552 assessment tool for genome assemblies. Bioinformatics 29, 1072–1075.
553 https://doi.org/10.1093/bioinformatics/btt086
554 Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009a. Bioinformatics enrichment
555 tools: Paths toward the comprehensive functional analysis of large gene
556 lists. Nucleic Acids Res. 37, 1–13. https://doi.org/10.1093/nar/gkn923
557 Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009b. Systematic and integrative
558 analysis of large gene lists using DAVID bioinformatics resources. Nat.

559 Protoc. 4, 44–57. https://doi.org/10.1038/nprot.2008.211
560 Huigens, M.E., Pashalidou, F.G., Qian, M.-H., Bukovinszky, T., Smid, H.M., van
561 Loon, J.J.A., Dicke, M., Fatouros, N.E., 2009. Hitch-hiking parasitic wasp
562 learns to exploit butterfly antiaphrodisiac. Proc. Natl. Acad. Sci. U. S. A.
563 106, 820–825. https://doi.org/10.1073/pnas.0812277106
564 Ivezić, A., Rugman-Jones, P., Stouthamer, R., Ignjatović-Ćupina, A., 2018.
565 Molecular identification of Trichogramma egg parasitoids of Ostrinia
566 nubilalis in northeastern Serbia. Arch. Biol. Sci. 70, 425–432.
567 https://doi.org/10.2298/ABS171103002I
568 Jamshidnia, A., Abdoli, S., Farrokhi, S., Sadeghi, R., 2018. Efficiency of spinosad,
569 Bacillus thuringiensis and Trichogramma brassicae against the tomato
570 leafminer in greenhouse. BioControl 63, 619–627.
571 https://doi.org/10.1007/s10526-018-9893-5
572 Johnston, J.S., Ross, L.D., Beani, L., Hughes, D.P., Kathirithamby, J., 2004. Tiny
573 genomes and endoreduplication in Strepsiptera. Insect Mol. Biol. 13, 581–
574 585. https://doi.org/10.1111/j.0962-1075.2004.00514.x
575 Keilwagen, J., Wenk, M., Erickson, J.L., Schattat, M.H., Grau, J., Hartung, F.,
576 2016. Using intron position conservation for homology-based gene
577 prediction. Nucleic Acids Res. 44. https://doi.org/10.1093/nar/gkw092
578 Klug, T., Meyhöfer, R., 2009. Performance of two Trichogramma brassicae
579 strains under greenhouse and field conditions for biocontrol of the silver Y
580 moth in spinach cultures. J. Pest Sci. (2004). 82, 73–79.
581 https://doi.org/10.1007/s10340-008-0224-y
582 Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M.,

583 2017. Canu: scalable and accurate long-read assembly via adaptive k-
584 mer weighting and repeat separation. Genome Res. 27, 722–736.
585 https://doi.org/10.1101/gr.215087.116
586 Kuske, S., Babendreier, D., Edwards, P.J., Turlings, T.C.J., Bigler, F., 2004.
587 Parasitism of non-target lepidoptera by mass released Trichogramma
588 brassicae and its implication for the larval parasitoid Lydella thompsoni.
589 BioControl 49, 1–19. https://doi.org/10.1023/B:BICO.0000009379.13685.47
590 Laurent, V., Wajnberg, E., Mangin, B., Schiex, T., Gaspin, C., Vanlerberghe-
591 Masutti, F., 1998. A composite genetic map of the parasitoid wasp
592 Trichogramma brassicae based on RAPD markers. Genetics 150, 275–82.
593 Lessard, E., Boivin, G., 2013. Effect of low temperature on emergence,
594 fecundity, longevity and host-feeding by Trichogramma brassicae.
595 BioControl 58, 319–329. https://doi.org/10.1007/s10526-012-9493-8
596 Li, H., 2018. Minimap2: Pairwise alignment for nucleotide sequences.
597 Bioinformatics 34, 3094–3100.
598 https://doi.org/10.1093/bioinformatics/bty191
599 Lindsey, A.R.I., Kelkar, Y.D., Wu, X., Sun, D., Martinson, E.O., Yan, Z., Rugman-
600 Jones, P.F., Hughes, D.S.T., Murali, S.C., Qu, J., Dugan, S., Lee, S.L., Chao,
601 H., Dinh, H., Han, Y., Doddapaneni, H.V., Worley, K.C., Muzny, D.M., Ye, G.,
602 Gibbs, R.A., Richards, S., Yi, S. V., Stouthamer, R., Werren, J.H., 2018.
603 Comparative genomics of the miniature wasp and pest control agent
604 Trichogramma pretiosum. BMC Biol. 16, 1–20.
605 https://doi.org/10.1186/s12915-018-0520-9
606 Liu, T.X., Zhang, Y., 2012. Side effects of two reduceδ-risk insecticides,

607 indoxacarb and spinosad, on two species of Trichogramma
608 (Hymenoptera: Trichogrammatidae) on cabbage. Ecotoxicology 21,
609 2254–2263. https://doi.org/10.1007/s10646-012-0981-5
610 Majoros, W.H., Pertea, M., Salzberg, S.L., 2004. TigrScan and GlimmerHMM: Two
611 open source ab initio eukaryotic gene-finders. Bioinformatics 20, 2878–
612 2879. https://doi.org/10.1093/bioinformatics/bth315
613 Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient
614 parallel counting of occurrences of k-mers. Bioinformatics 27, 764–770.
615 https://doi.org/10.1093/bioinformatics/btr011
616 Pannebakker, B.A., Pijnacker, L.P., Zwaan, B.J., Beukeboom, L.W., 2004.
617 Cytology of *Wolbachia*-induced parthenogenesis in *Leptopilina clavipes*
618 (Hymenoptera: Figitidae). Genome 303, 299–303.
619 https://doi.org/10.1139/G03-137
620 Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C.Y., Lin,
621 H., Lin, J.W., Hackett, K., 2015. The i5k Workspace@NAL-enabling genomic
622 data access, visualization and curation of arthropod genomes. Nucleic
623 Acids Res. 43, D714–D719. https://doi.org/10.1093/nar/gku983
624 Poelchau, M.F., Coates, B.S., Childers, C.P., Peréz De León, A.A., Evans, J.D.,
625 Hackett, K., Shoemaker, D.W., 2016. Agricultural applications of insect
626 ecological genomics. Curr. Opin. Insect Sci. 13, 61–69.
627 https://doi.org/10.1016/j.cois.2015.12.002
628 Polaszek, A., 2009. Species Diversity and Host Associations of Trichogramma in
629 Eurasia, in: Consoli, F.L., Parra, J.R.P., Zucchi, R.A. (Eds.), Egg Parasitoids in
630 Agroecosystems with Emphasis on Trichogramma. Springer Netherlands,

631 Dordrecht, pp. 237–266. https://doi.org/10.1007/978-1-4020-9110-0_9
632 Poorjavad, N., Goldansaz, S.H., Machtelinckx, T., Tirry, L., Stouthamer, R., van
633 Leeuwen, T., 2012. Iranian Trichogramma: ITS2 DNA characterization and
634 natural Wolbachia infection. BioControl 57, 361–374.
635 https://doi.org/10.1007/s10526-011-9397-z
636 Poorjavad, N., Goldansaz, S.H., van Leeuwen, T., 2018. Fertility life table
637 parameters, coi sequences and wolbachia infection in populations of
638 trichogramma brassicae collected from chilo suppressalis. Bull.
639 Insectology 71, 89–96.
640 Rahimi-Kaldeh, S., Ashouri, A., Bandani, A., 2018. Does Wolbachia Infection
641 Change the Overwintering Ability of Trichogramma brassicae
642 (Hymenoptera: Trichogrammatidae)? Neotrop. Entomol. 47, 583–590.
643 https://doi.org/10.1007/s13744-017-0549-4
644 Rugman-Jones, P.F., Stouthamer, R., 2017. High-resolution melt analysis without
645 DNA extraction affords rapid genotype resolution and species
646 identification. Mol. Ecol. Resour. 17, 598–607. https://doi.org/10.1111/1755-
647 0998.12599
648 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E.M.,
649 2015. BUSCO: Assessing genome assembly and annotation completeness
650 with single-copy orthologs. Bioinformatics 31, 3210–3212.
651 https://doi.org/10.1093/bioinformatics/btv351
652 Smith, S.M., 1996. Biological control with Trichogramma: advances, successes,
653 and potential of their use. Annu. Rev. Entomol. 41, 375–406.
654 https://doi.org/10.1146/annurev.ento.41.1.375

655 Stouthamer, R., 1997. Wolbachia-induced parthenogenesis, in: O'Neill, S.L.,
656 Hoffmann, A.A., Werren, J.H. (Eds.), Influential Passengers: Inherited
657 Microorganisms and Arthropod Reproduction. Oxford University Press,
658 Oxford, UK, pp. 102–124.
659 Stouthamer, R., Hu, J., Van Kan, F.J.P.M.P.M., Platner, G.R., Pinto, J.D., 1999.
660 The utility of internally transcribed spacer 2 DNA sequences of the nuclear
661 ribosomal gene for distinguishing sibling species of Trichogramma.
662 BioControl 43, 421–440. https://doi.org/10.1023/A:1009937108715
663 Stouthamer, R., Huigens, M., 2003. Parthenogenesis Associated With
664 Wolbachia, in: Bourtzis, K., Miller, T.A. (Eds.), Insect Symbiosis. CRC Press,
665 Boca Raton, pp. 247–266. https://doi.org/10.1201/9780203009918.ch15
666 Stouthamer, R., Kazmer, D.J., 1994. Cytogenetics of microbe-associated
667 parthenogenesis and its consequences for gene flow in Trichogramma
668 wasps. Heredity (Edinb). 73, 317–327. https://doi.org/10.1038/hdy.1994.139
669 Sumer, F., Tuncbilek, A.S., Oztemiz, S., Pintureau, B., Rugman-Jones, P.,
670 Stouthamer, R., 2009. A molecular key to the common species of
671 Trichogramma of the Mediterranean region. BioControl 54, 617–624.
672 https://doi.org/10.1007/s10526-009-9219-8
673 Suverkropp, B.P., Bigler, F., van Lenteren, J.C., 2010. Movement and host
674 finding of Trichogramma brassicae on maize plants. Bull. Insectology 63,
675 115–127.
676 Suverkropp, B.P., Bigler, F., van Lenteren, J.C., 2009. Dispersal behaviour of
677 Trichogramma brassicae in maize fields. Bull. Insectology 62, 113–120.
678 Testa, A.C., Hane, J.K., Ellwood, S.R., Oliver, R.P., 2015. CodingQuarry: Highly

679 accurate hidden Markov model gene prediction in fungal genomes using
680 RNA-seq transcripts. BMC Genomics 16, 1–12.
681 https://doi.org/10.1186/s12864-015-1344-4
682 Thubru, D.P., Firake, D.M., Behere, G.T., 2018. Assessing risks of pesticides
683 targeting lepidopteran pests in cruciferous ecosystems to eggs parasitoid,
684 Trichogramma brassicae (Bezdenko). Saudi J. Biol. Sci. 25, 680–688.
685 https://doi.org/10.1016/j.sjbs.2016.04.007
686 Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: Discovering splice
687 junctions with RNA-Seq. Bioinformatics 25, 1105–1111.
688 https://doi.org/10.1093/bioinformatics/btp120
689 Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J.,
690 Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and
691 quantification by RNA-Seq reveals unannotated transcripts and isoform
692 switching during cell differentiation. Nat. Biotechnol. 28, 511–5.
693 https://doi.org/10.1038/nbt.1621
694 Vaser, R., Sovic, I., Nagarajan, N., Sikic, M., 2017. Fast and accurate de novo
695 genome assembly from long uncorrected reads. Genome Res. 27, 737–
696 746. https://doi.org/10.1101/068122
697 Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Schatz, M.C., Gurtowski, J.,
698 Underwood, C.J., Vurture, G.W., Fang, H., 2017. GenomeScope: fast
699 reference-free genome profiling from short reads. Bioinformatics 33, 2202–
700 2204. https://doi.org/10.1093/bioinformatics/btx153
701 Wajnberg, E., 1993. Genetic variation in sex allocation in a parasitic wasp:
702 variation in sex pattern within sequences of oviposition. Entomol. Exp.

703 Appl. 69, 221–229. https://doi.org/10.1111/j.1570-7458.1993.tb01745.x
704 Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S.,
705 Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An
706 integrated tool for comprehensive microbial variant detection and
707 genome assembly improvement. PLoS One 9.
708 https://doi.org/10.1371/journal.pone.0112963
709 Wallberg, A., Bunikis, I., Pettersson, O.V., Mosbech, M.-B., Childers, A.K., Evans,
710 J.D., Mikheyev, A.S., Robertson, H.M., Robinson, G.E., Webster, M.T., 2019.
711 A hybrid de novo genome assembly of the honeybee, Apis mellifera, with
712 chromosome-length scaffolds. BMC Genomics 20, 275.
713 https://doi.org/10.1186/s12864-019-5642-0
714 Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y.Q.,
715 Coleman-Derr, D., Xia, Q., Wang, Y., 2019. OrthoVenn2: A web server for
716 genome wide comparison and annotation of orthologous clusters across
717 multiple species. Nucleic Acids Res. 47, W52–W58.
718 https://doi.org/10.1093/nar/gkz333
719 Ye, C., Hill, C.M., Wu, S., Ruan, J., Ma, Z., 2016. DBG2OLC: Efficient assembly of
720 large genomes using long erroneous reads of the third generation
721 sequencing technologies. Sci. Rep. 6, 1–9.
722 https://doi.org/10.1038/srep31900
723 Ye, C., Ma, Z.S., Cannon, C.H., Pop, M., Yu, D.W., 2012. Exploiting sparseness in
724 de novo genome assembly. BMC Bioinformatics 13 Suppl 6, S1.
725 https://doi.org/10.1186/1471-2105-13-S6-S1
726 Zhou, W., Rousset, F., O'Neill, S., 1998. Phylogeny and PCR-based classification

of Wolbachia strains using wsp gene sequences. Proc. R. Soc. B Biol. Sci. 265, 509–515. https://doi.org/10.1098/rspb.1998.0324

Zimmermann, O., 2004. Der Einsatz von Trichogramma-Schlupfwespen in Deutschland: Zum aktuellen Stand der Forschung und Nutzung von Eiparasitoiden gegen Schadlepidopteren im biologischen Pflanzen- und Vorratsschutz. Gesunde Pflanz. 56, 157–166. https://doi.org/10.1007/s10343-004-0031-1