Braconidae revisited: *Bracon brevicornis* genome showcases the potential of linked-read sequencing in identifying a putative *complementary sex determiner* gene

K. B. Ferguson*, B. A. Pannebakker*, A. Centurión†, J. van den Heuvel*, R. Nieuwenhuis‡, F. F. M. Becker*, E. Schijlen‡, A. Thiel†, B. J. Zwaan*, and E. C. Verhulst§.

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

† University of Bremen, FB02, Institute of Ecology, Population and Evolutionary Ecology Group, Bremen, Germany

‡ Wageningen University & Research, Bioscience, Wageningen, The Netherlands

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

DATA REFERENCE NUMBERS

ENA BioProject: PRJEB35412

Additional FASTA file, figshare DOI: 10.6084/m9.figshare.12674189.v2

Additional GFF file, figshare DOI: 10.6084/m9.figshare.12073911.v2

Supplementary data in DANS EASY repository, DOI: 10.17026/dans-xn6-pjm8

ABSTRACT

Bracon brevicornis is an ectoparasitoid of a wide range of larval-stage Lepidopterans, including several pests of important crops, such as the corn borer, *Ostrinia nubilalis*. It is also one of the earliest documented cases of complementary sex determination in Hymenoptera. Here, we present the linked-read genome of *B. brevicornis*, complete with an *ab initio*-derived annotation and protein comparisons with fellow braconids, *Fopius arisanus* and *Diachasma alloem*. We demonstrate the

potential of linked-read assemblies in exploring regions of heterozygosity and search for structural and homology-derived evidence of the *complementary sex determiner* gene (*csd*).

INTRODUCTION

Bracon brevicornis (Wesmael) is a gregarious ectoparasitoid of various Lepidoptera larvae, including many important pests, and is considered a cosmopolitan species (Temerak 1983b; Venkatesan *et al.* 2009). In the past *B. brevicornis* has been classified under the genus *Habrobracon* (Speicher and Speicher 1940), *Microbracon* (Narayanan *et al.* 1954), or classified as one species with *Habrobracon/Bracon hebetor* (Puttarudriah and Basavanna 1956), however recent research shows that *B. brevicornis* and *B. hebetor* are genetically two distinct species (Kittel and Maeto 2019). In the field, *B. brevicornis* has shown potential as a biological control agent against important pest species in stored corn stalks, such as *Ostrinia nubilalis* and *Sesamia cretica* (Kares *et al.* 2010), or against the coconut moth, *Opisinia arenosella* (Venkatesan *et al.* 2009). In the laboratory, *B. brevicornis* attacks a wide range of larval host such as *Ephesthia kuehniella*, *Galleria* spp., and *Spodoptera* spp. (Temerak 1983a).

Work on *B. brevicornis* has included both laboratory and semi-field set-ups to determine both its efficacy as a biological control agent as well as its suitability as a study system. There are several studies on the biology of *B. brevicornis*, e.g. on population growth potential (Srinivasan and Chandrikamohan 2017), their host range (Temerak 1983a), interspecific competition (Venkatesan *et al.* 2009), clutch size and fitness (Villacañas de Castro and Thiel 2017), mate choice (Thiel *et al.* 2013; Thiel and Weeda 2014), diet (Temerak 1983b), and efficacy (Kares *et al.* 2010).

Within a phylogenetic perspective, *B. brevicornis* falls within the subfamily Braconinae, the largest of the cyclostome-forming braconid wasps (Chen and van Achterberg 2019). The presence of a cyclostome (round mouthpart) is a defining feature within braconid wasps, as it represents an unresolved evolutionary and systematic question: is the cyclostome a derived trait within certain branches, or an ancestral trait that has been lost in others (Chen and van Achterberg 2019)? Within the Braconinae, there have been multiple switches from ectoparasitism to endoparasitism and *vice versa*, and this combination of cyclostome and endoparasitism has been described as a "controversial topic" by braconid researchers and taxonomists (Chen and van Achterberg 2019). These systematic issues are far from being resolved, and more genomic data would be useful for future phylogenetic analyses (Chen and van Achterberg 2019). Yet, a representative genome for the Braconinae is currently lacking. As previously stated, *B. brevicornis* is an ectoparasitoid, and its position within a family that contains both types of parasitism lifestyles holds promise for further phylogenetic comparisons.

In addition, as being part of the order Hymenoptera, *B. brevicornis* has a haplodiploid sex determination system where males develop from unfertilized eggs and females develop from fertilized eggs (Cook and Crozier 1995; Heimpel and de Boer 2008). From a genetic perspective, *B. brevicornis* belongs to an interesting genus where sex determination and diploid male production have been widely studied (*B. hebetor*, Whiting and Whiting 1925, *B. brevicornis*, Speicher and Speicher 1940, *B. serinopae* Clark, Bertrand, and Smith 1963, reviewed in van Wilgenburg, Driessen, and Beukeboom 2006; *B.* spec. near *hebetor*, Holloway *et al*. 1999; and *B. variator*, A. Thiel, pers. comm.). Indeed, the first description of the complementary

sex determination (CSD) mechanism was provided for *B. hebetor* (= *B. juglandis* by Whiting 1940, reviewed in Antolin *et al*. 2003), and recent work on *B. brevicornis* and polyploidy studies include diploid male fitness as well as ploidy-dependent mate choice behaviour (Thiel and Weeda 2014).

While straightforward to detect phenotypically through the formation of diploid males following inbreeding (van Wilgenburg *et al.* 2006), the molecular mechanism underlying CSD has thus far only been resolved to a small level of detail in the honeybee *Apis mellifera* (L.) (Hymenoptera: Apidae), with the identification of the *complementary sex determiner* (*csd*) gene (Beye *et al.* 2003). Heterozygosity at this gene leads to female development, while hemi- and homozygous individuals develop into haploid and diploid males respectively. Therefore, inbreeding often leads to diploid male production in species with a CSD mechanism as it increases homozygosity. *Csd* is a duplication of *feminizer* (*fem*), a *transformer* (*tra*) ortholog (Hasselmann *et al.* 2008) that is conserved across many insect orders as part of the sex determination cascade (Geuverink and Beukeboom 2014). When heterozygous, *csd* initiates the female-specific splicing of *fem*, which then autoregulates its own female-specific splicing, ultimately resulting in female development. Within the Hymenoptera, more duplications of *tra*/*fem* have been identified in species that are presumed to have CSD (Geuverink and Beukeboom 2014), but these *tra*/*fem* duplications have not been analysed for potential heterozygosity. Also, additional hymenopteran genomes are necessary to understand the evolutionary history of *tra*/*fem* duplications and identify the genes underlying CSD. However, an assembled genome is usually haploid as areas of heterozygosity are collapsed in the final stages of assembly. Yet recent advances in sequencing and analysis gave

us the ability to view heterozygous regions, known as "phases" in diploid assemblies, within a genome which allow us to investigate potential *csd* regions.

Here we report on the whole-genome sequencing of a pool of females from an isolated *B. brevicornis* strain using 10X Genomics technology that relies on linked-read sequencing (10x Genomics Inc., Pleasanton, CA, USA). Due to their long history of genetic isolation during laboratory rearing, the females in this strain are assumed to have a high level of homozygosity, whereas a *csd* locus would retain its heterozygosity. The 10X Genomics technology allows for generating phased data in which allelic variants can be identified after assembly. High-molecular weight DNA is partitioned into small droplets containing a unique barcode and adapter in such a way that only a few DNA molecules are present within each droplet. Within each droplet the DNA is broken into pieces and the barcode (Gel Bead-in-Emulsion, "GEM") is ligated to each of the DNA fragments. This resulting library can then be sequenced on an Illumina sequence platform. In the assembly step the reads originating from the same fragment are organized by barcode and put together into synthetic long-read fragments. Importantly, it is nearly impossible that two fragments with opposing allelic-variances are together in the same droplet (Weisenfeld *et al.* 2017). This technique therefore allowed us to identify potential *csd* candidates in the female-derived *B. brevicornis* genome after sequencing by studying the phased data containing the different haplotypes. Moreover, as *B. brevicornis* is a potential biological control agent of several pests, the availability of a full genome may provide effective ways to study and improve this species to grow it into an established biological control agent for Lepidopteran pests.

METHODS

Species description and general rearing:

Individuals of *B. brevicornis* were taken from the laboratory colony L06. The colony was initiated in 2006 from naturally parasitized *O. nubilalis* larvae collected in maize fields near Leipzig, Germany. Species identification was first carried out by Matthias Schöller and Cornelis van Achterberg based on morphological characteristics (Bernd Wührer, AMW Nützlinge, pers. comm.) Since collection, parasitoids have been reared on late instar larvae of the Mediterranean flour moth, *E. kuehniella* (Thiel and Weeda 2014). The species identity of strain L06 was recently revalidated based on molecular data and is entirely separate from its congeneric *B. hebetor* (Kittel and Maeto 2019).

DNA extraction:

Immediately following emergence, 100 to 120 female wasps were flash frozen in liquid nitrogen and ground with a mortar and pestle. Genomic DNA was extracted using a protocol modified from Chang, Puryear, and Cairney (Chang *et al.* 1993). Modifications include adding 300 µL BME to extraction buffer just before use. Instead of 10M LiCl, 0.7 volume isopropanol (100%) was added to the initial supernatant, after which it was divided into 1.5 mL Eppendorf tubes as 1 mL aliquots for subsequent extractions. The initial centrifugation step occurred at a slower rate and for a longer period of time to adjust for machine availability. Final pellets were dissolved in 50 µL autoclaved MQ and recombined at the end of the extraction process (1.0 mL). DNA concentration was measured with an Invitrogen Qubit 2.0 fluorometer using the dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, USA)

with final assessments for DNA quality, amount, and fragment size confirmed via BioAnalyzer 2100 (Agilent, Santa Clara, California, USA).

10X Genomics library preparation and sequencing:

As the genome of *B. brevicornis* is relatively small for the scale of the 10X platform, there is a higher risk of overlapping fragments within single GEMs. In order to reduce this risk, genomic DNA of a larger and previously analysed genome (Tomato, *Solanum lycopersicon* (L.) (Solanaceae), commercial variety Heinz 1607) (Hosmani *et al.* 2019) was used as 'carrier DNA'. DNA extraction of *S. lycopersicon* followed the protocol of Hosmani *et al*. (Hosmani *et al.* 2019). The DNA of both *B. brevicornis* and *S. lycopersicon* was pooled in a 1:4 molar ratio.

One nanogram of this pooled DNA was used for 10X Genomics linked-read library preparation following the Chromium Genome Reagent Kits Version 1 User Guide (CG-00022) (10x Genomics, Pleasanton, USA). Barcoded linked read DNA fragments were recovered for final Illumina library construction (Illumina, San Diego, USA). The library was used for 2 x 150 bp paired-end sequencing on one lane of an Illumina HiSeq 2500 at the business unit Bioscience of Wageningen University and Research (Wageningen, The Netherlands). Sequencing data was then used for basecalling and subsequent demultiplexing using Longranger (v2.2.2) (10X Genomics) (command `-mkfastq`), yielding 212,910,509 paired-end reads with a read length of 150 bp.

Assembly:

To filter sequence data from Heinz tomato (*S. lycopersicon*) carrier DNA sequences, 23bp (16bp GEM + 7 bp spacer) were removed from forward reads and all reads were subsequently mapped to an in-house high quality reference assembly of the

Heinz genome using BWA-MEM v0.7.17 (Li 2013). Using samtools v1.9 (Li *et al.* 2009), all unaligned read pairs (`-F=12`) were extracted and labelled non-Heinz. The assembly of the non-Heinz labelled read set was performed with 10X Supernova assembler v2.1.0 (10X Genomics), using default settings including commands for both pseudohap (`--style=pseudohap`) and pseudohap2 (`--style=pseudohap2`) outputs (Weisenfeld *et al.* 2017). These commands determine the output from Supernova, the first being the final scaffold output (pseudohap), while the second is the so-called 'parallel pseudohaplotype' (pseudohap2) scaffolds that represent areas of divergence or phases (Weisenfeld *et al.* 2017). Phasing is flattened in the pseudohap output by selecting the region with higher mapping coverage, whereas in the pseudohap2 output is differentiated by ".1" and ".2" at the end of each scaffold name to denote phasing, though not all scaffolds are phased at this point due to lack of divergence during assembly.

To verify whether there were no Heinz leftovers in the assembly, minimap2 v2.17-r941 (Li 2018) was used to align the assembly against the same Heinz assembly. Further examination on presence of possible non-*B. brevicornis* scaffolds, i.e. bacterial scaffolds from sample microbiome, was performed with BlobTools (v1.0) (Laetsch and Blaxter 2017), relying on megaBLAST against the NCBI NT-NR database (Acland *et al.* 2014)(2018-11-19) (`max_target_seqs=1, max_hsps=1, evalue=1e-25`) for taxonomical classification and BWA-MEM mapping of reads against scaffolds for coverage statistics. Reads mapping only against "Arthropoda" classified scaffolds were then extracted and used for a final k-mer analysis using jellyfish v2.1.1 (`-C m=21 -s=2000000000`) (Marçais and Kingsford 2011) and GenomeScope (Vurture *et al.* 2017) to infer heterozygosity.

Assembly completeness was determined using *BUSCO* (v3.0.2) with the `insect_odb9` ortholog set and the fly training parameter (Simão *et al.* 2015) while assembly statistics were determined using QUAST (Gurevich *et al.* 2013). The aforementioned pseudohap2 scaffolds were used in *csd* analysis, while the pseudohap scaffolds are now the assembly used for annotation.

Ab initio gene finding and protein comparison:

The coding sequences of two additional braconids (members of the subfamily Opiinae, and similar to the Braconinae belonging to the cyclostome subgroup (Li *et al.* 2013; Chen and van Achterberg 2019)) were used for gene prediction and protein comparisons: *Fopius arisanus* (Sonan) (Hymenoptera: Braconidae) and *Diachasma alloeum* (Muesebeck) (Hymenoptera: Braconidae). Both sets of coding sequences were retrieved from the NCBI Assembly Database, version ASM8063v1 for *F. arisanus* and version Dall2.0 for *D. alloem* (Acland *et al.* 2014; Geib *et al.* 2017; Tvedte *et al.* 2019).

For gene prediction, Augustus (v2.5.5) was first used to predict genes from the *B. brevicornis* assembly (Stanke and Morgenstern 2005). Using BLAST, coding sequences of *F. arisanus* were set as a query to the genome of *B. brevicornis* using default parameters (except `minIdentity=50`) (Camacho *et al.* 2009). The result was converted into a hints file that was used to predict the genes of *B. brevicornis* using *Nasonia vitripennis* (Walker) (Hymenoptera: Pteromalidae) as the species parameter in Augustus (`--species=nasonia --extrinsicCfgFile=extrinsic.E.cfg`).

After prediction, the protein sequences were retrieved and compared to both *F. arisanus* and *D. alloeum* (version Dall2.0) using Proteinortho (v6.0, `-p=blastp, -`

198 e=0.001) (Lechner *et al.* 2011). From the orthology grouping generated by
199 Proteinortho, gene names could be allocated to the predicted genes. Lengths of
200 both these *B. brevicornis* genes and the orthologs of *F. arisanus* and *D. alloem* were
201 retrieved using samtools for comparison (Li *et al.* 2009). Errors within the annotation
202 related to genome submission and validation were corrected with manual
203 annotation of exons (three cases) and removal of two predicted genes that were
204 more than 50% ambiguous nucleotides.

205 ***In silico* identification of *feminizer* as a putative *csd* locus:**

206 The pseudohap2 files were deduplicated using the dedupe tool within BBTools
207 (sourceforge.net/projects/bbmap/) (ac=f) to remove all parallel
208 pseudohaplotypes that were complete duplicates as these scaffolds were not
209 heterozygous. The remainder of the set contained both scaffolds that previously had
210 a duplicate, as well as solitary scaffolds that did not have a partner scaffold. These
211 unique scaffolds were removed using the "filter by name" tool in BBTools, leaving
212 258 scaffolds, or 129 pairs of pseudohap2 scaffolds. Pairs were pairwise aligned in
213 CLC Genomics Workbench 12 (Qiagen, Hilden, Germany) using default settings
214 (gap open cost=10, gap extension cost=1, end gap cost=free,
215 alignment=very accurate).

216 A local tBLASTn search against the entire *B. brevicornis* assembly was performed
217 using the *Apis mellifera* Feminizer protein (NP_001128300) as query in Geneious Prime
218 2019.1.3 (http://www.geneious.com, (Kearse *et al.* 2012)). The protein of gene
219 "g7607" (locus tag = BBRV_LOCUS33129) was used in an NCBI BLASTp against the nr
220 database with default settings (Camacho *et al.* 2009; Acland *et al.* 2014). Next a
221 region stretching from ~10Kbp upstream and downstream of the first and last

tBLASTn hit in scaffold 12, respectively, was annotated using HMM plus similar protein-based gene prediction (FGENESH+, Softberry, http://www.softberry.com/) with *Nasonia vitripennis tra* (NP_001128299) and *N. vitripennis* for the specific gene-finding parameters (Solovyev 2007). Only this combination of settings resulted in a full-length annotation from TSS to poly-A with seven exons. The resulting protein prediction was used in a BLASTp search with default settings against the nr database. To annotate the potential *fem* duplication, a stretch of ~10Kbp directly upstream of the annotated putative fem was again annotated using FGENESH+ (Softberry) with *Nasonia vitripennis tra* (NP_001128299) and *N. vitripennis* for the specific gene-finding parameters (Solovyv et al. 2007). The predicted annotation contained five exons but lacked the last coding segment with stop codon. A protein alignment was made in Geneious Prime 2019.1.3 with *A. mellifera csd* (ABU68670) and *fem* (NP_001128300)*; N. vitripennis tra* (XP_001604794) *and B. brevicornis* putative *fem* and *B. brevicornis* putative *fem* duplicate (*fem1*), using MAFFT v7.450 with the following settings: `Algorithm=auto, Scoring matrix=BLOSUM62, Gap open penalty=1.53, Offset value=0.123` (Katoh 2002; Katoh and Standley 2013).

Microsynteny analysis:

A microsynteny analysis was achieved by comparing the arrangement of a set of homologous genes directly upstream and downstream of *tra* or *fem* in *A. mellifera* and *N. vitripennis* using a combination of the online tool SimpleSynteny (Veltri *et al.* 2016) and tBLASTn searches using default settings in Geneious Prime. The scaffolds containing *fem* (*A. mellifera,* scaffold CM000059.5, 13.2Mbp in length), *tra* (*N. vitripennis,* scaffold NW_001820638.3, 3.7Mbp in length) or the putative *fem* (*B.*

brevicornis, scaffold 12, 4.5 Mbp in length) were extracted from their respective genomes (*Apis*: GCA_000002195.1_Amel_4.5_genomic, *Nasonia*: nvi_ref_Nvit_2.1, *Bracon*: *B. brevicornis* assembly from this study) and searched with protein sequence from the following genes: *tra* (GeneID: 00121203), LOC100121225, LOC100678616, LOC100680007 originating from *N. vitripennis*; and *fem* (GeneID:724970), *csd* (GeneID:406074), LOC408733, LOC551408, LOC724886 originating from *A. mellifera*. The advanced settings for SimpleSynteny were as follows: `BLAST E-value Threshold=0.01, BLAST Alignment type=Gapped, Minimum Query Coverage Cutoff=1%, Circular Genome Mode=Off`. If the gene was not found within the extracted scaffold, it was searched for in the full genome assembly. For the image settings, `Gene Display Mode=Project Full-Length Gene`. This generated image was used together with results from the tBLASTn searches as template to draw the final figure. The final figure that we present in the Results and Discussion section depicts ~0.9Mbp of genomic region for all three species.

Data availability:

Raw sequence data for *B. brevicornis* after removal of carrier DNA and contamination, as well as the assembly, can be found in the EMBL-EBI European Nucleotide Archive (ENA) under BioProject PRJEB35412, however, are currently being updated due to error in initial upload. In the meantime, both the assembly file (.fasta) (https://doi.org/10.6084/m9.figshare.12674189.v2) and the complete annotation file (.gff) (https://doi.org/10.6084/m9.figshare.12073911.v2) are available in a separate repository. Contaminated pseudohap scaffolds are available for download alongside the two pseudohap2 FASTA files, more details are provided in the supplementary materials at https://doi.org/10.17026/dans-xn6-pjm8.

RESULTS AND DISCUSSION

A total of 172 ng of *B. brevicornis* DNA was extracted, which was then reduced to 1 ng/µL for library preparation. Sequencing of the Heinz diluted library resulted in a total yield of 54 Gbp of data (corrected for 10X 23bp segment of forward reads). Mapping against the Heinz genome assembly showed a mapping percentage of 84.8%. There was a total of 30,278,915 unmapped pairs, comprising ~8.39 Gbp of data. This corresponds to the 4:1 ratio between Heinz and *B. brevicornis* DNA in the library. Further scaffold decontamination with BlobTools resulted in a separation of the assembly into *B. brevicornis* scaffolds and microbiome scaffolds. The final genome is 123,126,787 bp (123 Mbp) in size, comprised of 353 scaffolds (5.5% ambiguous nucleotides). This is similar to the projected physical genome size of 133 Mbp (J. G. de Boer, unpublished data, flow cytometry). BUSCO analysis indicates a completeness of 98.7% (single orthologs 97.0%, duplicate orthologs 1.7%).

K-mer analysis of the *B. brevicornis*-only read set showed an expected haploid genome length of ~115 mbp (105 Mbp unique, 10 Mbp repeat) and a heterozygosity of ~0.54%. Peak coverage was 27x.

Ab initio gene finding and protein comparison:

In total, 12,686 genes were predicted, with an average coding sequence length of 529.86 amino acids. The number of genes correspond well to those found in *F. arisanus* (11,775) and *D. alloem* (13,273), the two closest relatives of *B. brevicornis* for which public data is available. Proteinortho analysis resulted in 7660 three-way orthology groups (7,830 *B. brevicornis* genes), while 362 othology groups contained proteins of *B. brevicornis* and *F. arisanus* (382 *B. brevicornis* genes), and 451 groups contained *B. brevicornis* and *D. alloem* genes (479 *B. brevicornis* genes). A large

294 number of orthology groups (2,492) had no *B. brevicornis* genes, while 3,995
295 predicted genes remain ungrouped.

296 Compared to *F. arisanus*, the mean relative length of predicted *B. brevicornis* genes
297 was 1.016, while the mean relative length for the two- and three-way orthology
298 groups was 0.996. Similar results were obtained for comparisons to *D. alloem,* where
299 mean relative length for *B. brevicornis* genes was 1.011 and 0.988 for the two- and
300 three-way orthology groups. Furthermore, the pairwise lengths of all these proteins
301 resemble each other very well (Figure 1).

302 **Identification of a putative *feminizer* ortholog and duplication event:**

303 After deduplicating the similar parallel pseudohaplotype files, 6,706 scaffolds in
304 total, the remainder of the set contained 3,420 scaffolds, of which 3,286 scaffolds
305 were solitary and did not have a counterpart pseudohap2 for comparison. Some
306 had had a previous duplicate removed in the deduplication, while others never had
307 a partner scaffold in the first place. These unique scaffolds were removed, leaving
308 258 scaffolds, or 129 pairs of pseudohap2 scaffolds. These putatively heterozygous
309 scaffolds were good candidates to search for potential *csd* loci as these are
310 presumed to be heterozygous in females.

311 So far, a *csd* gene has been sequenced only in species of bees of the genus *Apis,*
312 and it is highly polymorphic, even within subspecies (Wang *et al.* 2012). It is located
313 adjacent to the more conserved *feminizer* (*fem*) (Hasselmann *et al.* 2008), and we
314 therefore started with localizing *feminizer* in the genome. As *feminizer* (or its ortholog
315 *transformer*, *tra*) was not identified in the *ab-initio* annotation, we used a local
316 tBLASTn search to find *fem* in the assembly. Four hits with E-value from 5.86e-04 to
317 8.59e-08 were found in scaffold 12. Searching the annotation using part of the

318 tBLASTn result shows that it is annotated as "g7607" (locus tag = BBRV_LOCUS33129)
319 which gave a first hit with protein O-glucosyltransferase 2 (*Diachasma alloeum*) after
320 a BLASTp search, and no *fem* or *tra* hits were found. A closer inspection showed that
321 "g7607" is annotated as fusion protein with the N-terminal part resembling *fem* and
322 the C-terminal part putatively encoding *O-glucosyltransferase 2*. Next, we used
323 FGENESH+ to re-annotate the genomic region, resulting in a full-length putative *B.*
324 *brevicornis feminizer* (*Bbfem*) ortholog containing seven exons (Figure 3). We found
325 that the two *fem*/*tra* signature domains in Hymenoptera, the Hymenoptera domain
326 (Verhulst *et al.* 2010) and CAM domain (putative autoregulatory domain) (Hediger
327 *et al.* 2010), are present in the putative *fem* ortholog, but are also duplicated
328 upstream of putative *Bbfem*. A second manual re-annotation step showed that a
329 partial *fem*-duplicate is encoded directly upstream of putative *Bbfem* containing
330 five exons (Figure 3), which we denote here as *Bbfem1* as suggested by Koch *et al.*
331 (Koch *et al.* 2014). The level of potential heterozygosity in the area encoding *Bbfem*
332 and *Bbfem1* is the highest when compared across all 129 pairs of pseudohap2
333 scaffolds (Figure 3).

334 A protein alignment showed that the full-length putative *Bbfem* as well as *Bbfem1*
335 are highly similar to each other and both contain all known *fem*/*tra* domains (Figure
336 S1). *Bbfem1* lacks a notably long first Arginine/Serine (RS)-rich region which is present
337 only in *Bbfem* (124-153aa), but it otherwise appears to encode for a full-length
338 protein. The *csd*-specific hypervariable domain (Figure S1, purple text; (Beye *et al.*
339 2003)) is not present in *Bbfem* nor in *Bbfem1*. Therefore, the gene name has been
340 updated as "g7607 putative Bbfem-Bbfem1 *csd*" in the official annotation.

Synteny analysis of putative *fem* encoding region:

We compared the orthologous gene arrangement of a number of genes up- and downstream of *N. vitripennis* tra and *A. mellifera fem* and *csd,* with the genomic organization of the *Bbfem* region (Figure 2). *N. vitripennis* LOC100680007 is present in the *tra/fem* containing scaffolds of all three genomes, while *A. mellifera* LOC408733 has both translocated closer to *Nasonia tra* and to a different scaffold in *B. brevicornis*. *N. vitripennis* LOC100121225 and LOC100678616 are encoded in opposing directions in both *A. mellifera* and *N. vitripennis* but are both downstream of *tra* in *N. vitripennis* and upstream of *fem* and *csd* in *A. mellifera*. There is no match for both genes in *B. brevicornis*. *A. mellifera* LOC724886 and LOC551408 are encoded in opposing directions with the same orientation in both *N. vitripennis* and *A. mellifera* but are reversed in *B. brevicornis* and downstream of *Bbfem* and *Bbfem1* while they are upstream of *csd* and *fem* in *A. mellifera*. In *N. vitripennis* both genes are not located in the *tra* containing scaffold but in another scaffold indicating that this region has undergone chromosomal rearrangements.

CONCLUSIONS AND PERSPECTIVES

Here, we present the genome of the braconid wasp *Bracon brevicornis*, a parasitoid wasp that not only has biological control applications, but also offers potential as a study system for future analyses into braconid phylogenetics and gene evolution. With no previous genomes available for the subfamily Braconinae, the most specious of the braconid wasps, the resources and investigations presented here fill this gap. Our linked-read library, assisted by carrier DNA of *S. lycopersicon,* has resulted in a highly contiguous, very complete assembly, comprised of just 353 scaffolds and 12,686 genes. This gene count is similar to related species, and in

365 further protein length comparisons, the proteins are highly similar. This indicates that
366 the predicted genes are highly complete, a necessary feature for any future
367 phylogenetic comparisons between species or families.

368 We utilized the 10X Genomics linked-read approach to obtain pseudohaploid
369 information that would allow us to search for potential *csd* loci *in silico*. As a
370 substantial number of scaffolds were putatively heterozygous, we used the notion
371 that in *A. mellifera, csd* is located adjacent to *fem* (Hasselmann *et al.* 2008) to limit
372 our search for *csd* candidates. We manually annotated a putative *B. brevicornis*
373 *fem* and a partial *Bbfem* duplicate that is highly similar, and both genes encode all
374 known *tra*/*fem* protein domains (Figure S1) (Verhulst *et al.* 2010). Both genes are in
375 a small region that is highly heterozygous, especially when compared to the
376 remainder of the scaffold, which would suggest true heterozygosity and not
377 assembly error, but also when compared to the level of heterozygosity in the other
378 128 aligned pseudohap2 scaffolds.

379 Our synteny analysis showed only little structural conservation between *B.*
380 *brevicornis*, *A. mellifera*, and *N. vitripennis* with the translocation of LOC408733 (*A.*
381 *mellifora*) and the absence of LOC100121225 and LOC100678616 (*N. vitripennis*) in
382 the *B. brevicornis* genome region. It is known that genomic regions encoding sex
383 determination genes are dynamic in nature, showing both duplications and
384 translocations (Dechaud *et al.* 2019). Also, *tra*/*fem* duplications have been shown
385 in CSD systems before, most notably in *A. mellifera* where a *fem* gene duplication
386 event resulted in it becoming a *csd* locus (Hasselmann *et al.* 2008; Gempe *et al.*
387 2009). However, also in non-CSD systems *tra* duplications have been observed
388 (Geuverink and Beukeboom 2014; Jia *et al.* 2016; Geuverink *et al.* 2018). Although

there is some debate on whether *fem* paralogs originated due to a single duplication event and functions as *csd* (Schmieder *et al.* 2012), or evolved multiple times independently and may have other functions (Koch *et al.* 2014), we suggest that the *Bbfem* paralog, *Bbfem1*, is a good *csd* gene candidate in *B. brevicornis*. However, in-depth analyses are required to verify this. Ultimately, our presented genome with its pseudohaploid information provides multiple opportunities for future studies, such as to improve the biological control opportunities with this species, but also to shed light on the evolutionary history of complementary sex determination systems.

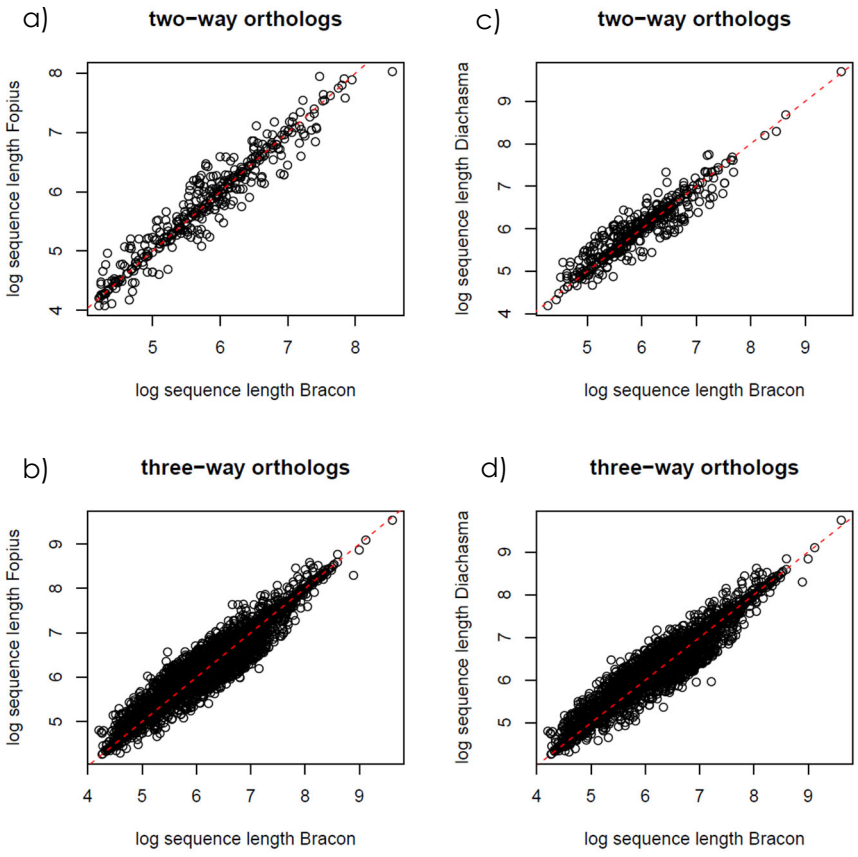
ACKNOWLEDGEMENTS AND FUNDING

We would like to acknowledge Jetske de Boer for information on *B. brevicornis* genome size, Martin Hasselmann for discussion on honeybee *csd*, and Elzemiek Geuverink for discussions on *tra/fem* duplicates in Hymenoptera. This project was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 641456.

SUPPLEMENTARY MATERIALS

Additional supplementary material from this study (contaminated scaffolds, pseudohap2 scaffolds) are available on the DANS EASY Repository, https://doi.org/10.17026/dans-xn6-pjm8

408 **FIGURES**



409

410 Figure 1. Protein length comparison between *Bracon brevicornis* and *Fopius*

411 *arisanus*, a) two- and b) three-way orthologs, and *B. brevicornis* and *Diachasma*

412 *alloem*, c) two- and d) three-way orthologs. Sequence lengths have been log-

413 transformed; red dashed line indicates synteny.

414

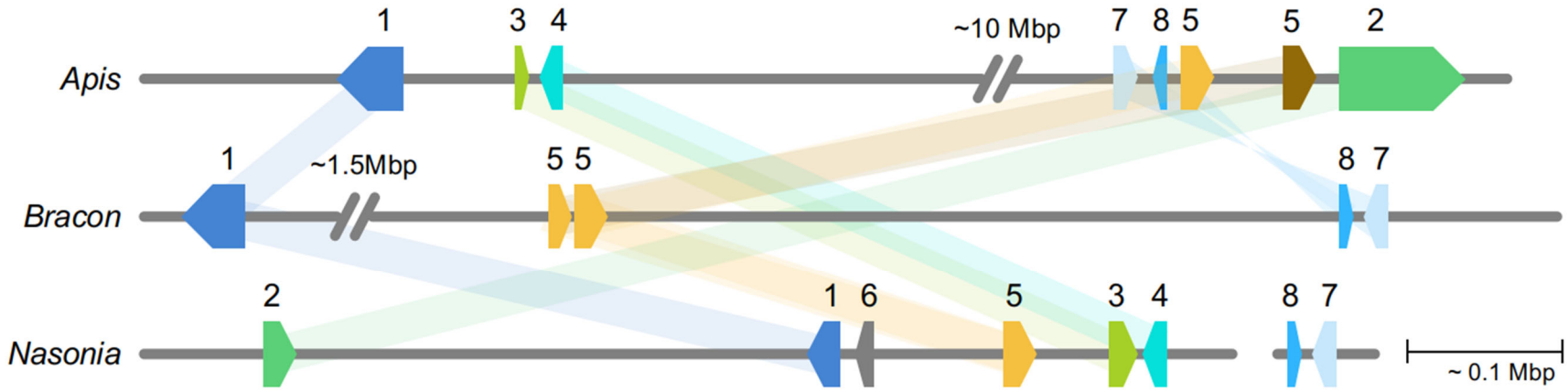


Figure 2. Microsynteny of genomic regions containing *tra/fem* paralogues. Shown is ~0.9 Mbp of genomic region of *Apis mellifera*, *Bracon brevicornis* and *Nasonia vitripennis*, containing the approximate coding region for 1. LOC100680007 (dark blue), 2. LOC408733 (green), 3. LOC100121225 (lime), 4. LOC100678616 (cyan), 5. *tra/fem/fem1* (yellow) and *csd* (brown), 6. LOC107980471 (gray), 7. LOC724886 (blue), 8. LOC551408 (light blue). Locus 2 is located on a different scaffold in *B. brevicornis*, locus 3 and 4 are not present in *B. brevicornis*. Locus 6 is unique to *N. vitripennis*, and locus 7 and 8 are located on a different scaffold in *N. vitripennis*, which is depicted on the right. Both 7 and 8 are in the same order and orientation as in *B. brevicornis*, but reversed in *A. mellifera*.

Figure 3. *Bracon brevicornis* annotation of *Bbfem* and *Bbfem1* on the alignment of pseudohaplotype track 1 and 2 in Geneious Prime 2019.1.3 (http://www.geneious.com, (Kearse *et al.* 2012)). Within the assembled genome, this section corresponds to a region on scaffold 12. The *Bbfem1* annotation lacks the last coding segment with stop codon. The identity track shows the amount of sequence identity across an arbitrary window (depending on zoom setting) and can be used as a proxy for heterozygosity. Green is identical, yellow is mismatch, and red is no match due to introduced gaps during alignment. The coding regions of *Bbfem1* and *Bbfem* are in a high putatively heterozygous region.

Figure S1. Protein alignment of *A. mellifera csd* (ABU68670) and *fem* (NP_001128300), *N. vitripennis tra* (XP_001604794), *B. brevicornis fem* and *fem1*. Purple shading indicates Hymenoptera domain (Verhulst *et al.* 2010), yellow shading indicates CAM domain (Hediger *et al.* 2010), blue shading indicates Proline (P)-rich region, red text colour indicates Arginine/Serine (RS)-rich regions, and purple text colour indicates hypervariable region in *csd* (Beye *et al.* 2003).

LITERATURE CITED

Acland, A., R. Agarwala, T. Barrett, J. Beck, D. A. Benson *et al.*, 2014 Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 42: 8–13. https://doi.org/10.1093/nar/gkt1146.

Antolin, M. F., P. J. Ode, G. E. Heimpel, R. B. O'Hara, and M. R. Strand, 2003 Population structure, mating system, and sex-determining allele diversity of the parasitoid wasp Habrobracon hebetor. Heredity (Edinb). 91: 373–381. https://doi.org/10.1038/sj.hdy.6800337.

Beye, M., M. Hasselmann, M. K. Fondrk, R. E. Page, and S. W. Omholt, 2003 The gene csd is the primary signal for sexual development in the honeybee and encodes an SR-type protein. Cell 114: 419–429. https://doi.org/10.1016/S0092-8674(03)00606-8.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: Architecture and applications. BMC Bioinformatics 10: 421. https://doi.org/10.1186/1471-2105-10-421.

Chang, S., J. Puryear, and J. Cairney, 1993 A simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Report. 11: 113–116. https://doi.org/10.1007/BF02670468.

Chen, X., and C. van Achterberg, 2019 Systematics, Phylogeny, and Evolution of Braconid Wasps: 30 Years of Progress. Annu. Rev. Entomol. 64: 335–358. https://doi.org/10.1146/annurev-ento-011118-111856.

Clark, A. M., H. A. Bertrand, and R. E. Smith, 1963 Life Span Differences between Haploid and Diploid Males of Habrobracon serinopae after Exposure as Adults to X Rays. Am. Nat. 97: 203–208. https://doi.org/10.1086/282271.

Cook, J. M., and R. H. Crozier, 1995 Sex determination and population biology in the hymenoptera. Trends Ecol. Evol. 10: 281–286. https://doi.org/10.1016/0169-5347(95)90011-X.

Dechaud, C., J. N. Volff, M. Schartl, and M. Naville, 2019 Sex and the TEs: Transposable elements in sexual development and function in animals. Mob. DNA 10: 42. https://doi.org/10.1186/s13100-019-0185-0.

Geib, S. M., G. H. Liang, T. D. Murphy, and S. B. Sim, 2017 Whole Genome Sequencing of the Braconid Parasitoid Wasp Fopius arisanus , an Important Biocontrol Agent of Pest Tepritid Fruit Flies. G3 Genes, Genomes, Genet. 7: 2407–2411. https://doi.org/10.1534/g3.117.040741.

Gempe, T., M. Hasselmann, M. Schiøtt, G. Hause, M. Otte *et al.*, 2009 Sex determination in honeybees: Two separate mechanisms induce and maintain the female pathway (B. S. Baker, Ed.). PLoS Biol. 7: e1000222.

https://doi.org/10.1371/journal.pbio.1000222.

Geuverink, E., and L. W. Beukeboom, 2014 Phylogenetic Distribution and Evolutionary Dynamics of the Sex Determination Genes doublesex and transformer in Insects. Sex. Dev. 8: 38–49. https://doi.org/10.1159/000357056.

Geuverink, E., K. Kraaijeveld, M. van Leussen, F. Chen, J. Pijpe *et al.*, 2018 Evidence for involvement of a transformer paralogue in sex determination of the wasp Leptopilina clavipes. Insect Mol. Biol. 27: 780–795. https://doi.org/10.1111/imb.12522.

Gurevich, A., V. Saveliev, N. Vyahhi, and G. Tesler, 2013 QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29: 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Hasselmann, M., T. Gempe, M. Schiøtt, C. G. Nunes-Silva, M. Otte *et al.*, 2008 Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. Nature 454: 519–522. https://doi.org/10.1038/nature07052.

Hediger, M., C. Henggeler, N. Meier, R. Perez, G. Saccone *et al.*, 2010 Molecular Characterization of the Key Switch F Provides a Basis for Understanding the Rapid Divergence of the Sex-Determining Pathway in the Housefly. Genetics 184: 155–170. https://doi.org/10.1534/genetics.109.109249.

Heimpel, G. E., and J. G. de Boer, 2008 Sex Determination in the Hymenoptera. Annu. Rev. Entomol. 53: 209–230. https://doi.org/10.1146/annurev.ento.53.103106.093441.

Holloway, A. K., G. E. Heimpel, M. R. Strand, and M. F. Antolin, 1999 Survival of diploid males in Bracon sp. near hebetor (Hymenoptera: Braconidae). Ann. Entomol. Soc. Am. 92: 110–116. https://doi.org/10.1093/aesa/92.1.110.

Hosmani, P. S., M. Flores-Gonzalez, H. van de Geest, F. Maumus, L. V Bakker *et al.*, 2019 An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. bioRxiv 767764. https://doi.org/https://doi.org/10.1101/767764.

Jia, L.-Y., J.-H. Xiao, T.-L. Xiong, L.-M. Niu, and D.-W. Huang, 2016 The transformer genes in the fig wasp C eratosolen solmsi provide new evidence for duplications independent of complementary sex determination. Insect Mol. Biol. 25: 191–201. https://doi.org/10.1111/imb.12210.

Kares, E. A., I. A. El-Sappagh, G. H. Ebaid, and I. M. Sabra, 2010 Efficacy of releasing Bracon brevicornis wesm. (Hymenoptera: Braconidae) for controlling hibernated Ostrinia nubilalis (Hübner) and Sesamia cretica led. larvae in stored corn stalks. Egypt. J. Biol. Pest Control 20: 155–159.

Katoh, K., 2002 MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30: 3059–3066. https://doi.org/10.1093/nar/gkf436.

Katoh, K., and D. M. Standley, 2013 MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30: 772–780. https://doi.org/10.1093/molbev/mst010.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung *et al.*, 2012 Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Kittel, R. N., and K. Maeto, 2019 Revalidation of Habrobracon brevicornis stat. rest. (Hymenoptera: Braconidae) Based on the CO1, 16S, and 28S Gene Fragments. J. Econ. Entomol. 112: 906–911. https://doi.org/10.1093/jee/toy368.

Koch, V., I. Nissen, B. D. Schmitt, and M. Beye, 2014 Independent evolutionary origin of femparalogous genes and complementary sex determination in hymenopteran insects (W. Hughes, Ed.). PLoS One 9: e91883. https://doi.org/10.1371/journal.pone.0091883.

Laetsch, D. R., and M. L. Blaxter, 2017 BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. F1000Research 6: 1287. https://doi.org/10.12688/f1000research.12232.1.

Lechner, M., S. Findeiß, L. Steiner, M. Marz, P. F. Stadler *et al.*, 2011 Proteinortho: Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics 12: 124. https://doi.org/10.1186/1471-2105-12-124.

Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv.

Li, H., 2018 Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34: 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, X. Y., C. van Achterberg, and J. C. Tan, 2013 Revision of the subfamily Opiinae (Hymenoptera, Braconidae) from Hunan (China), including thirty-six new species and two new genera. Zookeys 268: 1–186. https://doi.org/10.3897/zookeys.268.4071.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Marçais, G., and C. Kingsford, 2011 A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27: 764–770.

https://doi.org/10.1093/bioinformatics/btr011.

Narayanan, E. S., G. W. Angalet, B. R. Subba Rao, and G. I. D'Souza, 1954 Effect of refrigeration of the pupæ of microbracon brevicornis wesm. on the pigmentation of the adult. Nature 173: 503–504. https://doi.org/10.1038/173503b0.

Puttarudriah, M., and G. P. C. Basavanna, 1956 A study on the identity of bracon hebetor say and bracon brevicornis wesmael (hymenoptera: Braconidae). Bull. Entomol. Res. 47: 183–191. https://doi.org/10.1017/S0007485300046617.

Schmieder, S., D. Colinet, and M. Poirié, 2012 Tracing back the nascence of a new sex-determination pathway to the ancestor of bees and ants. Nat. Commun. 3: 895. https://doi.org/10.1038/ncomms1898.

Simão, 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. https://doi.org/10.1093/bioinformatics/btv351.

Solovyev, V., 2007 Statistical Approaches in Eukaryotic Gene Prediction, pp. 97–159 in *Handbook of Statistical Genetics*, edited by D. Balding, C. Cannings, and M. Bishop. Wiley-Interscience,.

Speicher, B. R., and K. G. Speicher, 1940 The Occurrence of Diploid Males in Habrobracon brevicornis. Am. Nat. 74: 379–382. https://doi.org/10.1086/280904.

Srinivasan, T., and Chandrikamohan, 2017 Population growth potential of Bracon brevicornis Wesmael (Braconidae: Hymenoptera): A life table analysis. Acta Phytopathol. Entomol. Hungarica 52: 123–130. https://doi.org/10.1556/038.52.2017.010.

Stanke, M., and B. Morgenstern, 2005 AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33: W465–W467. https://doi.org/10.1093/nar/gki458.

Temerak, S. A., 1983a Host preferences of the parasitoid Bracon brevicornis Wesmael (Hym., Braconidae) and host sensitivity to its venom. Zeitschrift für Angew. Entomol. 96: 37–41. https://doi.org/10.1111/j.1439-0418.1983.tb03638.x.

Temerak, S. A., 1983b Longevity of Bracon brevicornis [Hym: Braconidae] adults as influenced by nourishment on artificial and natural foods. Entomophaga 28: 145–150. https://doi.org/10.1007/BF02372138.

Thiel, A., and A. C. Weeda, 2014 Haploid, diploid, and triploid - Discrimination ability against polyploid mating partner in the Parasitic Wasp, Bracon

brevicornis (Hymenoptera: Braconidae). J. Insect Sci. 14: 1–7. https://doi.org/10.1093/jisesa/ieu153.

Thiel, A., A. C. Weeda, J. G. de Boer, and T. S. Hoffmeister, 2013 Genetic incompatibility drives mate choice in a parasitic wasp. Front. Zool. 10: 43. https://doi.org/10.1186/1742-9994-10-43.

Tvedte, E. S., K. K. O. Walden, K. E. McElroy, J. H. Werren, A. A. Forbes *et al.*, 2019 Genome of the parasitoid wasp diachasma alloeum, an emerging model for ecological speciation and transitions to asexual reproduction (J. Gonzalez, Ed.). Genome Biol. Evol. 11: 2767–2773. https://doi.org/10.1093/gbe/evz205.

Veltri, D., M. M. Wight, and J. A. Crouch, 2016 SimpleSynteny: a web-based tool for visualization of microsynteny across multiple species. Nucleic Acids Res. 44: W41–W45. https://doi.org/10.1093/nar/gkw330.

Venkatesan, T., S. K. Jalali, and K. Srinivasamurthy, 2009 Competitive interactions between Goniozus nephantidis and Bracon brevicornis, parasitoids of the coconut pest Opisina arenosella. Int. J. Pest Manag. 55: 257–263. https://doi.org/10.1080/09670870902914155.

Verhulst, E. C., L. van de Zande, and L. W. Beukeboom, 2010 Insect sex determination: it all evolves around transformer. Curr. Opin. Genet. Dev. 20: 376–383. https://doi.org/10.1016/j.gde.2010.05.001.

Villacañas de Castro, C., and A. Thiel, 2017 Resource-Dependent Clutch Size Decisions and Size-Fitness Relationships in a Gregarious Ectoparasitoid Wasp, Bracon brevicornis. J. Insect Behav. 30: 454–469. https://doi.org/10.1007/s10905-017-9632-2.

Vurture, G. W., F. J. Sedlazeck, M. Nattestad, M. C. Schatz, J. Gurtowski *et al.*, 2017 GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33: 2202–2204. https://doi.org/10.1093/bioinformatics/btx153.

Wang, Z., Z. Liu, X. Wu, W. Yan, and Z. Zeng, 2012 Polymorphism analysis of csd gene in six Apis mellifera subspecies. Mol. Biol. Rep. 39: 3067–3071. https://doi.org/10.1007/s11033-011-1069-7.

Weisenfeld, N. I., V. Kumar, P. Shah, D. M. Church, and D. B. Jaffe, 2017 Direct determination of diploid genome sequences. Genome Res. 27: 757–767. https://doi.org/10.1101/gr.214874.116.Freely.

Whiting, P. W., 1940 Multiple alleles in sex determination of Habrobracon. J. Morphol. 66: 323–355. https://doi.org/10.1002/jmor.1050660208.

Whiting, P. W., and A. E. Whiting, 1925 Diploid males from fertilized eggs in hymenoptera. Science (80-.). 62: 437.

https://doi.org/10.1126/science.62.1611.437.

van Wilgenburg, E., G. Driessen, and L. W. Beukeboom, 2006 Single locus complementary sex determination in Hymenoptera: an "unintelligent" design? Front. Zool. 3: 1. https://doi.org/https://doi.org/10.1186/1742-9994-3-1.