9-2021

A Chromosome-Level Genome Assembly of the Parasitoid Wasp, *Cotesia glomerata* (Hymenoptera: Braconidae)

Brendan J. Pinto
University of Texas - Austin

Jerome J. Weis
Milwaukee Public Museum

Tony Gamble
Marquette University, anthony.gamble@marquette.edu

Paul J. Ode
Colorado State University - Fort Collins

Ryan Paul
Colorado State University - Fort Collins

See next page for additional authors

Authors

Brendan J. Pinto, Jerome J. Weis, Tony Gamble, Paul J. Ode, Ryan Paul, and Jennifer M. Zaspel

A Chromosome-Level Genome Assembly of the Parasitoid Wasp, *Cotesia glomerata* (Hymenoptera: Braconidae)

Brendan J Pinto
Department of Integrative Biology, University of Texas at Austin, Austin, TX
Milwaukee Public Museum, Milwaukee, WI

Jerome J Weis
Milwaukee Public Museum, Milwaukee, WI

Tony Gamble
Milwaukee Public Museum, Milwaukee, WI
Department of Biological Sciences, Marquette University, Milwaukee, WI
Bell Museum of Natural History, University of Minnesota, St Paul, MN

Paul J Ode
Colorado State University, Ft. Collins, CO

Ryan Paul

Colorado State University, Ft. Collins, CO

Jennifer M Zaspel

Milwaukee Public Museum, Milwaukee, WI

Department of Biological Sciences, Marquette University, Milwaukee, WI

Abstract

Hymenopterans make up about 20% of all animal species, but most are poorly known and lack high-quality genomic resources. One group of important, yet understudied hymenopterans are parasitoid wasps in the family Braconidae. Among this understudied group is the genus *Cotesia*, a clade of ~1,000 species routinely used in studies of physiology, ecology, biological control, and genetics. However, our ability to understand these organisms has been hindered by a lack of genomic resources. We helped bridge this gap by generating a high-quality genome assembly for the parasitoid wasp, *Cotesia glomerata* (Braconidae; Microgastrinae). We generated this assembly using multiple sequencing technologies, including Oxford Nanopore, whole-genome shotgun sequencing, and 3D chromatin contact information (HiC). Our assembly is one of the most contiguous, complete, and publicly available hymenopteran genomes, represented by 3,355 scaffolds with a scaffold N50 of ~28 Mb and a BUSCO score of ~99%. Given the genome sizes found in closely related species, our genome assembly was ~50% larger than expected, which was apparently induced by runaway amplification of 3 types of repetitive elements: simple repeats, long terminal repeats, and long interspersed nuclear elements. This assembly is another step forward for genomics across this hyperdiverse, yet understudied order of insects. The assembled genomic data and metadata files are publicly available via Figshare (https://doi.org/10.6084/m9.figshare.13010549).

Keywords

Introduction

The insect order Hymenoptera (comprised >150,000 described species of wasps, bees, ants, and sawflies) is one of the most species-rich animal groups on the planet (Forbes et al. 2018). Much of this diversity lies within several groups of wasps with a "parasitoid" life history, where parasitic larvae develop into free-living adults (Peters et al. 2017). However, parasitoid wasps have generally been understudied taxonomically (Jones et al. 2009). In fact, undocumented parasitoid diversity could make the Hymenoptera the most species-rich order of animals (Forbes et al. 2018). Despite this profound species diversity and myriad of interesting biological phenomena within the Hymenoptera, high-quality genomic data for large swathes of the hymenopteran phylogeny are still lacking (Branstetter et al. 2018).

Braconidae is one of the largest hymenopteran families with >17,000 described species (Yu et al. 2005) and diverged from their closest relatives (Ichneumonidae) in the early-to-mid Jurassic (~175 million years ago) (Peters et al. 2017). The group includes many introduced species important for biological control of insect pests in field and greenhouse agriculture (Heimpel and Lundgren 2000; Van Driesche 2008; Acebes and Messing 2013). Perhaps most notably, the release of Eurasian parasitoid species, *Cotesia glomerata* (Braconidae; Microgastrinae), in North America, in the 1880s to control populations of imported cabbageworm (*Pieris rapae*) was one of the first documented cases of introductions for biological control in the world (Heimpel and Cock 2017). *Cotesia glomerata* has also been a valuable model in a wide range of topics including parasitoid–microbe interactions (Cusumano et al. 2018; Zhu et al. 2018), multitrophic interactions (Bukovinszky et al. 2009; Poelman et al. 2011, 2012), competition and life-history trade-offs (Geervliet et al. 2000; Gu and Dorn 2003; Vyas et al.

2019), nectar nutrition (Wäckers 2001; Wanner et al. 2006), neurobiology (Van Vugt et al. 2015), sex determination (Zhou et al. 2006), and host foraging (Horikoshi et al. 1997; Shiojiri et al. 2006; De Rijk et al. 2016). Continued exploration into these aspects, and others, of the biology and evolution of braconid wasps is currently hindered by the lack of available high-quality genomic data (Branstetter et al. 2018). To date, multiple genomic resources have been developed in braconid wasps. However, these resources are generally lacking in contiguity and phylogenetic breadth. To address this gap, we present a high-quality genome assembly for the parasitoid wasp, *C. glomerata*, generated using long-read, short-read, and long-range contact (HiC) sequencing data. The resulting *C. glomerata* genome assembly reported herein is among the most contiguous and complete assemblies currently available across all hymenopterans.

Methods

Data Generation

We obtained *C. glomerata* specimens from a colony maintained at Colorado State University. Field-collected founders of this colony were initially collected from cabbage fields at the Colorado State Agricultural Research, Development, and Education Center near Wellington, CO. *Cotesia glomerata* are haplodiploid with 2*n* females having 20 chromosomes and 1*n* males having 10 chromosomes (Zhou et al. 2006). We extracted high-molecular-weight DNA (HMW DNA) from 1 male [CgM1] and 1 female [CgF25] *C. glomerata* specimens using a custom low-input DNA extraction protocol designed for small vertebrates and adapted for single insect input (Supplementary Methods). For each individual, we generated a sequencing library for the Oxford Nanopore Technologies (ONT) platform (SQK-RAD004) and conducted long-read DNA sequencing using a MinION (2 FLO-MINSP6 flowcells). We quality-filtered male and female ONT reads using NanoFilt [v2.5.0] (De Coster et al. 2018) by discarding any sequence with an average quality of <7 and a read length <1000 bp. We then concatenated male and female ONT reads to produce our final ONT dataset, which contained 1 864 940 reads (9 440 627 459 bp) with a read N50 of 8222 bp (~32× coverage).

We extracted additional DNA from a second male *C. glomerata* specimen [CgM2] using the QIAamp DNA Micro Kit (Qiagen) and generated a proximally ligated DNA library for a pool of 5 male *C. glomerata* individuals using the Arima® HiC kit (Arima Genomics®, San Diego, CA). Then, we generated a total of 3 Illumina® sequencing libraries (genome re-sequencing for CgM1 and CgM2, and the proximally ligated HiC library) using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® with dual indices (New England BioLabs®). We sequenced re-sequencing libraries on an Illumina® HiSeqX (Psomagen®, Rockville, MD), and the HiC library was multiplexed on a NovaSeq lane (Novogene®, Davis, CA). We generated 2 male re-sequencing libraries for different individuals, as the remaining DNA after ONT sequencing (CgM1) was minute and produced a low-concentration library—which was supplemented by the second individual.

We processed raw illumina reads by quality and adapter trimming reads using Trim Galore! [v0.6.5] (https://github.com/FelixKrueger/TrimGalore; Martin 2011), filtering PCR duplicates using BBmap [v38.79] (Bushnell, 2014), and QC for each read set using FastQC [v0.11.7] (Andrews 2010). Data generated for each library/individual are summarized in Supplementary Table 1.

Genome Assembly

To maximize the quality of the final assembly with these data, we generated and compared several polished draft genome assemblies using 4 methods: Flye [v2.8.1] (Kolmogorov et al. 2019), wtdbg [v2.5] (Ruan and Li, 2020), Canu [v2.0] (Koren et al. 2017), and Miniasm [v0.3]/Racon [v1.4.13] (Li, 2016; Vaser et al. 2017). We then polished each assembly using ONT and Illumina reads with NextPolish [v1.3.1] (Hu et al. 2020) and assessed the contiguity and completeness of each step using the program Benchmarking Using Single-Copy Orthologs [v3.0] (BUSCO) with the arthropod ortholog database (odb9) (Simão et al. 2015), via the gVolante web server [v1.2.1]

(Nishimura et al. 2017) (Table 1). At this point, we chose the highest quality genome assembly (the Flye assembly; Table 1) to continue through HiC scaffolding and genome annotation. We broke and rescaffolded the polished Flye assembly using 2 iterations of 3D-DNA [v201008] (Dudchenko et al. 2017), which yielded 11 chromosome scale scaffolds with no apparent large-scale misassemblies. We visualized the final HiC contact map for misassemblies using Juicebox [v1.11] (Durand et al. 2016). However, we identified 2 arms of a single chromosome that were not scaffolded together and manually connected the scaffolds using Juicebox. Thus, postcuration, this assembly contained 10 chromosome-level scaffolds.

Table 1. Tracking quality and contiguity of the genome assemblies across versions using 5 common metrics: Contig/scaffold N50, size of the smallest scaffold comprising the largest 50% of the assembly; Scaffold L50, number of scaffolds comprising the largest 50% of the genome; Scaffolds, total number of scaffolds comprising the analyzed assembly; Size, the number of base pairs in the assembly; and BUSCO score using the arthropod database (odb9) (Nishimura et al. 2017)

Version	Step	N50	L50	Scaffolds	Size (Mb)	BUSCO (%)
Flye Assembly Statistics						
Flye Assembly v0.1	*Flye*	1 105 129	52	4767	287	79.0
Flye Assembly v0.2	*NextPolish*	1 122 554	52	4767	290	99.0
Flye Assembly v0.3	*3D-DNA*	27 795 534	5	3973	291	98.8
C. glomerata MPM v2.0	*Funannotate*	27 795 534	5	3355	290	98.9
Transcripts	*Funannotate*	Genome annotations				96.2
Peptides	*Funannotate*	Genome annotations				96.3
wtdbg2 Assembly Statistics						
wtdbg2 Assembly v0.1	*wtdbg2*	473 826	107	3316	302	37.24
wtdbg2 Assembly v0.2	*NextPolish*	482 022	109	3316	312	96.0
Miniasm/Racon Assembly Statistics						
Miniasm Assembly v0.1	*Miniasm/Racon*	329 657	242	1780	297	63.9
Miniasm Assembly v0.2	*NextPolish*	329 029	243	1780	299	89.9
Canu Assembly Statistics						
Canu Assembly v0.1	*Canu*	108 160	715	6368	338	58.9
Canu Assembly v0.2	*NextPolish*	110 205	718	6368	345	94.5

Genome Annotation

Prior to annotation, we trimmed the minimal scaffold length to 1000 bp (Table 1). To functionally annotate the genome assembly, we used the Funannotate pipeline [v1.5.0] (Palmer 2018) in an isolated docker computing environment (Merkel 2014). Briefly, the funannotate pipeline provides a simple pipeline to train and predict gene models using both curated databases (Simão et al. 2015) and RNAseq data (Haas et al. 2008; Keller et al. 2011; Hoff et al. 2016). To compile a posteriori annotation evidence, we fetched available paired-end RNAseq reads from previously conducted RNAseq experiments from Van Vugt et al. (2015) (Bioproject PRJNA289655: SRR2105801, SRR2105795, SRR2105791, SRR2105785). We quality and adapter trimmed reads before filtering PCR duplicates (as described above) of the concatenated RNAseq reads files. As one line of evidence, we mapped these reads directly to the genome assembly using HiSat2 [v2.1] (Kim et al. 2019). As another line of evidence, we assembled a de novo transcriptome using the *De novo* RNAseq Assembly Pipeline (DRAP) [v1.91] (Cabau et al. 2017) to assist in genome annotation. An in-depth description of the utility of DRAP in the

production of high-quality transcriptome assemblies can be found elsewhere (Cabau et al. 2017; Pinto et al. 2019). This transcriptome, along with the hymenopteran BUSCO database (odb9), was used to train gene prediction models during the annotation process. We conducted a systematic assessment of other, freely available hymenopteran genome assemblies using the same method to compare with our *C. glomerata* assembly using the same metrics as Table 1 (see Table 2).

Table 2. Representative genome assemblies for publicly available hymenopteran species compared with the new *C. glomerata* assembly (bold) across 5 common metrics: Scaffold N50, size of the smallest scaffold comprising the largest 50% of the assembly; Scaffold L50, number of scaffolds comprising the largest 50% of the genome; Scaffolds, total number of scaffolds comprising the analyzed assembly; Size, the number of base pairs in the assembly; and BUSCO score using the arthropod database

Species	Family	N50 (kb)	L50	Total scaffolds	Size (Mb)	BUSCO (%)	Citation
Ceratosolen solmsi	Agaonidae	9558	10	2457	277	99.2	Xiao et al. (2013)
Apis mellifera	Apidae	13 219	8	5320	250	99.5	Elsik et al. (2014)
Bombus terretris	Apidae	12 868	9	5609	249	99.0	Sadd et al. (2015)
Cotesia congregata	Braconidae	20 027	5	2039	199	97.5	Gauthier et al. (2021)
Cotesia glomerata	**Braconidae**	**27796**	**5**	**3355**	**290**	**98.9**	**This work**
Cotesia vestalis	Braconidae	51	818	39 682	172	97.3	Shi et al. (2019)
Diachasma alloeum	Braconidae	645	128	3968	389	98.6	Tvedte et al. (2019)
Fopius arisanus	Braconidae	978	49	1042	154	97.8	Burke, Simmonds, et al. (2018)
Macrocentrus cingulum	Braconidae	192	179	5696	132	98.8	Yin et al. (2018)
Microplitis demolitor	Braconidae	1139	50	1794	241	99.0	Burke, Walden, et al. (2018)
Copidosoma floridanum	Encyrtidae	1037	153	5445	555	97.1	i5k:GCA_000648655
Atta cephalotes	Formicidae	5154	21	2835	318	97.3	Suen et al. (2011)
Monomorium pharaonis	Formicidae	15 646	7	9622	259	97.7	Warner et al. (2019)
Ooceraea biroi	Formicidae	16 888	6	139	224	99.3	Oxley et al. (2014)
Temnothorax longispinosus	Formicidae	514	137	3987	261	98.1	Kaur et al. (2019)
Vollenhovia emeryi	Formicidae	1346	61	13 258	288	99.8	Miyakawa and Mikheyev (2015)
Diadromus collaris	Ichneumonidae	1030	107	63 392	399	99.6	Shi et al. (2019)
Orussus abietinus	Orussidae	612	63	1789	186	98.5	i5k:GCA_000612105
Nasonia giraulti	Pteromalidae	759	64	4912	284	97.1	Werren et al. (2010)

Nasonia longicornis	Pteromalidae	758	65	5214	286	97.2	Werren et al. (2010)
Nasonia vitripennis	Pteromalidae	897	21	6098	296	97.1	Werren et al. (2010)
Trichogramma pretiosum	Trichogrammatidae	3706	19	357	195	99.3	Lindsey et al. (2018)

Results and Discussion

Data Description

Across individuals and sequencing platforms, we generated a data set of ~130× coverage for genome assembly, ~32× in Oxford Nanopore (ONT) and ~100× PE Illumina reads. We also generated a dataset of 140× Illumina reads derived from 3D chromatin capture experiment (HiC). Our initial assembly depended on the ONT long reads, Illumina reads for downstream polishing, and HiC for scaffolding. Data are summarized further in Supplementary Table 1.

Genome Assembly

By almost every metric, Flye produced the best draft genome assembly. It was the most contiguous (N50) and had the highest per-base quality of all 4 assemblies (represented by a higher BUSCO score, given that each assembly was derived from the same dataset). In contrast, Canu provided the lowest-quality assembly by all metrics including an order of magnitude increase in runtime, while Miniasm/Racon and wtdbg generated intermediate assemblies. Of note, assemblers varied in speed by orders of magnitude with wtdbg being the fastest from reads to consensus (~2 h), then Miniasm and Racon (~9 h), Flye (~8 h), and finally Canu (>96 h). However, after polishing, only the wtdbg assembly approached the quality of the Flye assembly (summarized further in Table 1). Thus, we chose the Flye assembly for additional scaffolding using the HiC data.

Our *C. glomerata* genome assembly achieved "chromosome-level" status with 90% of the genome attributed to the largest 10 scaffolds; representative of each of *C. glomerata*'s 10 chromosomes (1*n* = 10, sensu Zhou et al. 2006). Additionally, 95.6% of the final assembly is represented in 46 scaffolds that are longer than 50 kb. Quality-control read-mapping of Illumina back to the final assembly produced extremely high mapping rates for both CgM1 (98.9% reads mapped; 96.1% properly paired) and CgM2 (98.4% reads mapped; 95.4% properly paired). The base composition across the genome assembly was relatively homogenous with an overall GC content of 31% (±6%). The BUSCO statistics for the final assembly, relative to a set of 1066 conserved arthropod orthologs, included 97.3% complete single-copy genes with 1.5% complete and duplicated gene versions, 0.5% fragmented gene copies, and 0.7% missing genes. Genome annotation, using RNAseq data derived from experiments of expression in the heads of multiple female *C. glomerata*, successfully passed 29 455 annotations, including 24 170 genes. Interestingly, the final genome assembly approached 300 Mb in size, which far exceeded our initial predictions for its genome size given the available data from their closest relatives (Figure 1).
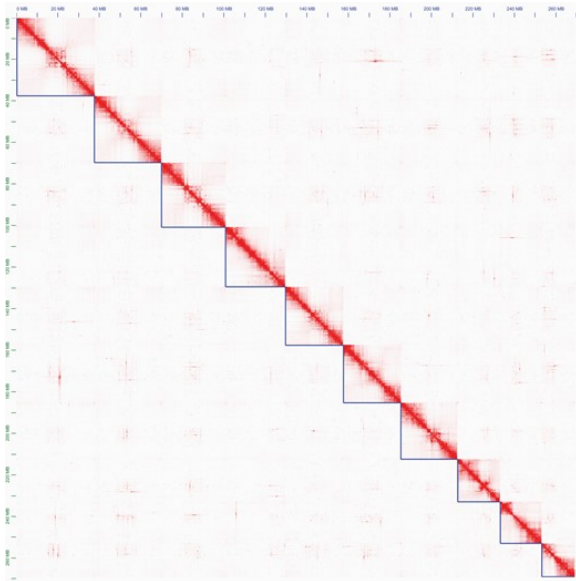
Figure 1. The HiC contact map of the final *Cotesia glomerata* genome assembly cropped to the 10 largest scaffolds generated in Juicebox (Durand et al. 2016).

Genome sizes across Hymenoptera can vary greatly between taxa, such as the 3-fold difference between *Macrocentrus cingulum* and *Diadromus collaris* within the superfamily Ichneumonoidea (Table 2). Previous genome assemblies within *Cotesia*, *C. congregata* (Gauthier et al. 2021) and *C. vestalis* (Shi et al. 2019), had haploid genome sizes that were far closer to ~200 Mb (Table 2). We set out to compare across the genomes of these taxa to shed light on this apparent genome size discrepancy. However, drastic differences in sequencing technologies and genome quality across these species prevented direct comparisons between all 3 species. To close this gap, we dropped *C. vestalis* from downstream analyses and focused solely on comparing the *C. glomerata* and *C. congregata* [v2.0] (Gauthier et al. 2021) genomes.

We investigated this genome size discrepancy by interrogating the repeat content using Illumina reads to account for differences in sequencing technologies between assemblies—along with a estimating total repeat content from the assemblies. First, we confirmed this was the haploid genome size by estimating the *C. glomerata* (CgM2) and *C. congregata* (ENA: ERR3829581) genome size using kmers with Jellyfish [v2.2.10] (Marçais and Kingsford 2011). The estimated genome size of *C. glomerata* roughly approximated the assembled genome size at 272 Mb, whereas the estimated genome size of *C. congregata* was 142 Mb (Figure 2). Next, using RepeatModeler [v1.0.11] (Smit and Hubley 2008), we calculated percent repeat content for each genome 54.75% (159.28 Mb) of the *C. glomerata* assembly and 35.04% (69.72 Mb) of the *C. congregata* assembly. This >100% increase in repeat content is sufficient to account for the large increase in genome size in *C. glomerata*, relative to *C. congregata*. Lastly, we approximated the categories of genomic repeats responsible for the discrepancies in genome size between these 2 *Cotesia* species using dnaPipeTE [v1.3.1] (Goubert et al. 2015). We identified 3 primary categories of repetitive elements that increased proportionally between *C. glomerata* relative to *C. congregata*. A large proportion of this amplified repetitive content appears to be made up of simple repeats, long terminal repeats, and long interspersed nuclear elements (Figure 1).
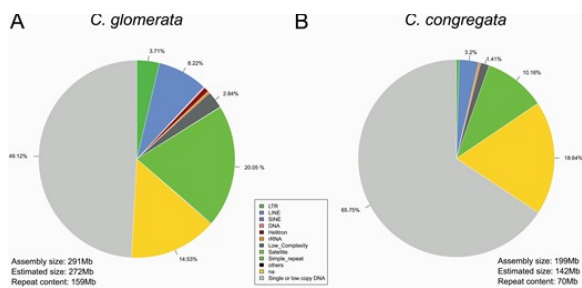
Figure 2. Comparative genome TE content between *Cotesia glomerata* and *C. congregata*. "Assembly Size" is the total assembled genome size for each species; "Estimated Size" is the genome size estimation calculated using kmer counting with Jellyfish; and "Repeat Content" is the total repeat content estimated using RepeatModeler. Categories of genomic repeat content and pie charts generated by dnaPipeTE, unspecified % pie slices represent <1% estimated content. It should also be noted that the color palette for each wheel begins at the top of both the key and the wheels (with LTR) and continues clockwise down the palette in the order listed in the key.

Here, we present a chromosome-level genome assembly for *C. glomerata*. This assembly is currently the most complete and, in part due to its exceptionally large genome size (larger chromosomes), is also one of the most contiguous hymenopteran genomes to date, i.e., having the highest N50 (Table 2). Interestingly, the genome size was much larger than initially hypothesized based on information from its closest relatives, and this discrepancy appears to be due to runaway amplification of repetitive elements in the *C. glomerata* lineage (Figure 1). As more high-quality braconid genomes continue to be generated, fruitful areas of research may include diving deeper into the unanswered questions, such as the impact of repeat content on the parasitoid lifestyle and the evolution of sex determination in Hymenoptera—among others. Indeed, this assembly is another step forward for genomics across this hyperdiverse, yet understudied, order of insects.

Acknowledgments

Author Contributions

B.J.P, T.G., P.O., J.Z., and J.W. conceptualized the study. R.P. and P.O. conducted fieldwork. B.J.P. and J.W. extracted DNA and B.J.P. generated sequencing libraries, conducted ONT sequencing, analyzed data, and drafted the manuscript. All authors edited, read, and approved the final manuscript.

Data Availability

Sequence data is submitted to the NCBI Sequence Read Archive (SRA) under BioProject PRJNA715233. Assembled data is publicly available at https://doi.org/10.6084/m9.figshare.13010549.

References

Acebes AL, Messing RH. 2013. Comparative susceptibility to hyperparasitism of Binodoxys communis and Aphidius colemani, primary aphid parasitoids introduced to Hawaii. Biol Control. 65:286–292.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Branstetter MG, Childers AK, Cox-Foster D, Hopper KR, Kapheim KM, Toth AL, Worley KC. (2018). Genomes of the Hymenoptera. Curr Opin Insect Sci. 25:65–75.

Bukovinszky T, Poelman EH, Gols R, Prekatsakis G, Vet LE, Harvey JA, Dicke M. 2009. Consequences of constitutive and induced variation in plant nutritional quality for immune defence of a herbivore against parasitism. Oecologia. 160:299–308.

Burke GR, Simmonds TJ, Sharanowski BJ, Geib SM. 2018. Rapid viral symbiogenesis via changes in parasitoid wasp genome architecture. Mol Biol Evol. 35:2463–2474.

Burke GR, Walden KKO, Whitfield JB, Robertson HM, Strand MR. 2018. Whole genome sequence of the parasitoid wasp Microplitis demolitor that harbors an endogenous virus mutualist. G3 (Bethesda). 8:2875–2880.

Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner (No. LBNL-7065E). Berkeley (CA): Lawrence Berkeley National Lab (LBNL).

Cabau C, Escudié F, Djari A, Guiguen Y, Bobe J, Klopp C. 2017. Compacting and correcting Trinity and Oases RNA-Seq de novo assemblies. PeerJ. 5:e2988.

Cusumano A, Zhu F, Volkoff AN, Verbaarschot P, Bloem J, Vogel H, Dicke M, Poelman EH. 2018. Parasitic wasp-associated symbiont affects plant-mediated species interactions between herbivores. Ecol Lett. 21:957–967.

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 34:2666–2669.

De Rijk M, Krijn M, Jenniskens W, Engel B, Dicke M, Poelman EH. 2016. Flexible parasitoid behaviour overcomes constraint resulting from position of host and nonhost herbivores. Anim Behav. 113:125–135.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science. 356:92–95.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016. Juicebox provides a visualization system for hi-C Contact maps with unlimited zoom. Cell Syst. 3:99–101.

Elsik CG, Worley KC, Bennett AK, Beye M, Camara F, Childers CP. 2014. Finding the missing honey bee genes: lessons learned from a genome upgrade. BMC Genomics. 15(1):86.

Forbes AA, Bagley RK, Beer MA, Hippee AC, Widmayer HA. 2018. Quantifying the unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal order. BMC Ecol. 18:21.

Gauthier J, Boulain H, van Vugt JJFA, Baudry L, Persyn E, Aury JM, Noel B, Bretaudeau A, Legeai F, Warris S, et al. 2021. Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. Commun Biol. 4:104.

Geervliet JB, Verdel MS, Snellen H, Schaub J, Dicke M, Vet LE. 2000. Coexistence and niche segregation by field populations of the parasitoids Cotesia glomerata and C. rubecula in the Netherlands: predicting field performance from laboratory data. Oecologia. 124:55–63.

Goubert C, Modolo L, Vieira C, ValienteMoro C, Mavingui P, Boulesteix M. 2015. De novo assembly and annotation of the Asian tiger mosquito (Aedes albopictus) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (Aedes aegypti). Genome Biol Evol. 7:1192–1205.

Gu H, Dorn S. 2003. Mating system and sex allocation in the gregarious parasitoid Cotesia glomerata. Anim. Behav. 66:259–264.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9:R7.

Heimpel GE, Cock MJW. 2017. Shifting paradigms in the history of classical biological control. BioControl. 63:27–37.

Heimpel GE, Lundgren JG. 2000. Sex ratios of commercially reared biological control agents. Biol Control. 19:77–93.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 32:767–769.

Horikoshi M, Takabayashi J, Yano S, Yamaoka R, Ohsaki N, Sato Y. 1997. Cotesia glomerata female wasps use fatty acids from plant–herbivore complex in host searching. J Chem Ecol. 23:1505–1515.

Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics. 36:2253–2255.

Kaur R, Stoldt M, Jongepier E, Feldmeyer B, Menzel F, Bornberg-Bauer E, Foitzik S. 2019. Ant behaviour and brain gene expression of defending hosts depend on the ecological success of the intruding social parasite. Philos Trans R Soc Lond B Biol Sci. 374:20180192.

Keller O, Kollmar M, Stanke M, Waack S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics. 27(6):757–763.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 37(8):907–915.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 37:540–546.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27(5):722–736.

Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics. 32(14):2103–2110.

Lindsey AR, Kelkar YD, Wu X, Sun D, Martinson EO, Yan Z, ... Werren JH. 2018. Comparative genomics of the miniature wasp and pest control agent Trichogramma pretiosum. BMC Biol. 16:1–20.

Jones OR, Purvis A, Baumgart E, Quicke DLJ. 2009. Using taxonomic revision data to estimate the geographic and taxonomic distribution of undescribed species richness in the Braconidae (Hymenoptera: Ichneumonoidea). Insect Conserv Diver. 2:204–212.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 27(6):764–770.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. 17(1):10–12.

Merkel D. 2014. Docker: lightweight Linux containers for consistent development and deployment. Linux J. 239:2.

Miyakawa MO, Mikheyev AS. 2015. QTL mapping of sex determination loci supports an ancient pathway in ants and honey bees. PLoS Genet. 11:e1005656.

Nishimura O, Hara Y, Kuraku S. 2017. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. Bioinformatics. 33(22):3635–3637.

Oxley PR, Ji L, Fetter-Pruneda I, McKenzie SK, Li C, Hu H, Kronauer DJC. 2014. The genome of the clonal raider ant Cerapachys biroi. Curr Biol. 24(4):451–458.

Palmer J. 2018. Funannotate: eukaryotic genome annotation pipeline. doi:10.5281/zenodo.2604804

Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Niehuis O. 2017. Evolutionary history of the hymenoptera. Curr Biol. 27(7):1013–1018.

Pinto BJ, Card DC, Castoe TA, Diaz RE Jr., Nielsen SV, Trainor PA, Gamble T. (2019). The transcriptome of the Veiled Chameleon (Chamaeleo calyptratus): a resource for studying the evolution and development of vertebrates. Dev Dyn. 248:702–708.

Poelman EH, Bruinsma M, Zhu F, Weldegergis BT, Boursault AE, Jongema Y, Dicke M. 2012. Hyperparasitoids use herbivore-induced plant volatiles to locate their parasitoid host. PLoS Biol. 10(11):e1001435.

Poelman EH, Zheng S-J, Zhang Z, Heemskerk NM, Cortesero A-M, Dicke M. 2011. Parasitoid-specific induction of plant responses to parasitized herbivores affects colonization by subsequent herbivores. Proc Natl Acad Sci USA. 108(49):19647–19652.

Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 17(2):155–158.

Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, Elsik CG, Lozier JD. Robertson HM, Smagghe G, Stolle E, et al. 2015. The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol. 16(76):1–31.

Shi M, Wang Z, Ye X, Xie H, Li F, Hu X, Wang Z, Yin C, Zhou Y, Gu Q, et al. 2019. The genomes of two parasitic wasps that parasitize the diamondback moth. BMC Genomics. 20(893):1–13.

Shiojiri K, Ozawa R, Matsui K, Kishimoto K, Kugimiya S, Takabayashi J. 2006. Role of the lipoxygenase/lyase pathway of host-food plants in the host searching behavior of two parasitoid species, Cotesia glomerata and Cotesia plutellae. J Chem Ecol. 32:969–979.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31(19):3210–3212.

Smit AFA, Hubley R. 2008–2015. RepeatModeler Open-1.0. https://www.repeatmasker.org

Suen G, Teiling C, Li L, Holt C, Abouheif E, Bornberg-Bauer E, Caldera EJ, Cash E, Cavanaugh A, Denas O. 2011. The genome sequence of the leaf-cutter ant Atta cephalotes reveals insights into its obligate symbiotic lifestyle. PLoS Genet. 7(2):e1002007.

Tvedte ES, Walden KKO, McElroy KE, Werren JH, Forbes AA, Hood GR, Logsdon JM, Feder JF, Robertson HM. 2019. Genome of the parasitoid wasp diachasma alloeum, an emerging model for ecological speciation and transitions to asexual reproduction. Genome Biol Evol. 11(10):2767–2773.

Van Driesche RG. 2008. Biological control of Pieris rapae in New England: host suppression and displacement of Cotesia glomerata by Cotesia rubecula (Hymenoptera: Braconidae). Fla Entomol. 91:22–25.

Van Vugt JJ, Hoedjes KM, Van de Geest HC, Schijlen EW, Vet LE, Smid HM. 2015. Differentially expressed genes linked to natural variation in long-term memory formation in Cotesia parasitic wasps. Front Behav Neurosci. 9:255.

Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27(5):737–746.

Vyas DK, Harvey JA, Paul RL, Heimpel GE, Ode PJ. 2019. Ecological dissociation and re-association with a superior competitor alters host selection behavior in a parasitoid wasp. Oecologia. 191:261–270.

Wäckers FL. 2001. A comparison of nectar- and honeydew sugars with respect to their utilization by the hymenopteran parasitoid Cotesia glomerata. J Insect Physiol. 47:1077–1084.

Wanner H, Gu H, Dorn S. 2006. Nutritional value of floral nectar sources for flight in the parasitoid wasp, Cotesia glomerata. Physiol Entomol. 31:127–133.

Warner MR, Qiu L, Holmes MJ, Mikheyev AS, Linksvayer TA. 2019. Convergent eusocial evolution is based on a shared reproductive groundplan plus lineage-specific plastic genes. Nat Commun. 10:2651.

Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Werren JH, Richards S, Desjardins CA, Niehuis O, et al. ; Nasonia Genome Working Group. 2010. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science. 327:343–348.

Xiao JH, Yue Z, Jia LY, Yang XH, Niu LH, Wang Z, Zhang P, Sun B-F, He S-M, Li Z. 2013. Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. Genome Biol. 14(12):R141.

Yin C, Li M, Hu J, Lang K, Chen Q, Liu J, ... Chen X. 2018. The genomic features of parasitism, polyembryony and immune evasion in the endoparasitic wasp Macrocentrus cingulum. BMC Genomics. 19(420):1–18.

Yu D, van Achterberg K, Horstmann K. 2005. World Ichneumonidea 2004. Taxonomy, biology, morphology and distribution. Vancouver (Canada): Taxapad.

Zhou Y, Gu H, Dorn S. 2006. Single-locus sex determination in the parasitoid wasp Cotesia glomerata (Hymenoptera: Braconidae). Heredity (Edinb). 96:487–492.

Zhu F, Cusumano A, Bloem J, Weldegergis BT, Villela A, Fatouros NE, van Loon JJA, Dicke M, Harvey JA, Vogel H, Poelman EH. 2018. Symbiotic polydnavirus and venom reveal parasitoid to its hyperparasitoids. Proc Natl Acad Sci USA. 115(20):5205–5210.