

RESEARCH ARTICLE

# A partial genome assembly of the miniature parasitoid wasp, *Megaphragma amalphanum*

Fedor S. Sharko<sup>1,2</sup>, Artem V. Nedoluzhko<sup>1,2,3</sup>, Brandon M. Lê<sup>4</sup>, Svetlana V. Tsygankova<sup>2</sup>, Eugenia S. Boulygina<sup>2</sup>, Sergey M. Rastorguev<sup>2</sup>, Alexey S. Sokolov<sup>1</sup>, Fernando Rodriguez<sup>4</sup>, Alexander M. Mazur<sup>1</sup>, Alexey A. Polilov<sup>5</sup>, Richard Benton<sup>6</sup>, Michael B. Evgen'ev<sup>7</sup>, Irina R. Arkhipova<sup>4</sup>, Egor B. Prokhortchouk<sup>1,5\*</sup>, Konstantin G. Skryabin<sup>1,2,5†</sup>

**1** Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia, **2** National Research Center "Kurchatov Institute", Moscow, Russia, **3** Nord University, Faculty of Biosciences and Aquaculture, Bodø, Norway, **4** Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, Massachusetts, United States of America, **5** Lomonosov Moscow State University, Faculty of Biology, Moscow, Russia, **6** Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland, **7** Institute of Molecular Biology RAS, Moscow, Russia

☉ These authors contributed equally to this work.

† Deceased.

\* nedoluzhko@gmail.com (AN); prokhortchouk@gmail.com (EP)



**OPEN ACCESS**

**Citation:** Sharko FS, Nedoluzhko AV, Lê BM, Tsygankova SV, Boulygina ES, Rastorguev SM, et al. (2019) A partial genome assembly of the miniature parasitoid wasp, *Megaphragma amalphanum*. PLoS ONE 14(12): e0226485. <https://doi.org/10.1371/journal.pone.0226485>

**Editor:** John Parkinson, Hospital for Sick Children, CANADA

**Received:** January 18, 2019

**Accepted:** November 26, 2019

**Published:** December 23, 2019

**Copyright:** © 2019 Sharko et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All genome and transcriptome assemblies and all SRA sequence data are publicly available at the NCBI BioProject: PRJNA344956.

**Funding:** This work has been carried out using computing resources of the federal collective usage center Complex for Simulation and Data Processing for Mega-science Facilities at NRC "Kurchatov Institute" (ministry subvention under agreement RFMEFI62117X0016) (<http://ckp.nrcki.ru/>). Support of this project was provided by the

## Abstract

Body size reduction, also known as miniaturization, is an important evolutionary process that affects a number of physiological and phenotypic traits and helps animals conquer new ecological niches. However, this process is poorly understood at the molecular level. Here, we report genomic and transcriptomic features of arguably the smallest known insect—the parasitoid wasp, *Megaphragma amalphanum* (Hymenoptera: Trichogrammatidae). In contrast to expectations, we find that the genome and transcriptome sizes of this parasitoid wasp are comparable to other members of the Chalcidoidea superfamily. Moreover, compared to other chalcid wasps the gene content of *M. amalphanum* is remarkably conserved. Intriguingly, we observed significant changes in *M. amalphanum* transposable element dynamics over time, in which an initial burst was followed by suppression of activity, possibly due to a recent reinforcement of the genome defense machinery. Overall, while the *M. amalphanum* genomic data reveal certain features that may be linked to the unusual biological properties of this organism, miniaturization is not associated with a large decrease in genome complexity.

## Introduction

Miniaturization in animals is an evolutionary process that is frequently accompanied by structural simplification and size reduction of organs, tissues and cells [1, 2]. The parasitoid wasp *Megaphragma amalphanum* (Hymenoptera: Trichogrammatidae, subfamily Oligositinae) is one of the smallest known insects, whose size (250 µm adult length) is comparable with unicellular eukaryotes and even some bacteria (Fig 1). Parasitoids from the genus *Megaphragma* parasitize greenhouse thrips *Heliethrips haemorrhoidalis* (Thysanoptera: Thripidae) developing on the shrubs *Viburnum tinus* (Adoxaceae) and *Myrtus communis* (Myrtaceae) [3], and possibly

Russian Scientific Foundation grant #14-24-00175 to K.S. and E.P. (<http://rscf.ru/en/>). Support of this project was also provided by the Russian Scientific Foundation grant #14-50-00060 to M.E. (<http://rscf.ru/en/>). B.L. was supported by the Rosenthal Brown-MBL internship and the REU supplement to NSF MCB-1121334 to I.A. (<https://www.nsf.gov/>). R.B. was supported by a European Research Council Consolidator Grant (615094), (<https://erc.europa.eu/>).

**Competing interests:** The authors have declared that no competing interests exist.

*Hercinothrips femoralis* (Thysanoptera: Thripidae) [4]. The wasp spends most of its life cycle in host eggs, while the imago stage is very short and lasts only a few days [3, 4]. *M. amalphanum* belongs to chalcid wasps, which represent one of the largest insect superfamilies (~23,000 described species)[5]. The higher-level taxonomic relationships of Trichogrammatidae, Chalcidoidea and Hymenoptera have been investigated in several recent studies [6–10] that helped to establish the placement of this unique taxon that related to Mymaridae and Pteromalidae.

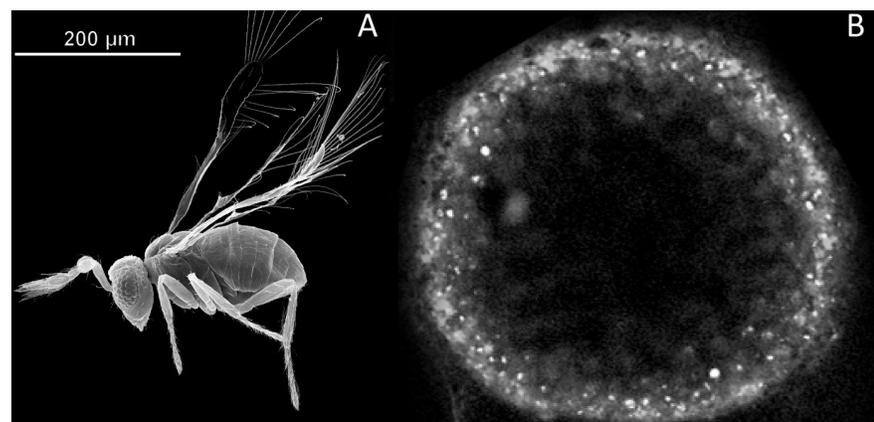
Amongst notable anatomical features of *M. amalphanum*, this species has only ~4,600 neurons in its brain, which is substantially fewer than in the brains of other wasps, e.g., the parasitoid chalcid wasp *Trichogramma pretiosum* (Trichogrammatidae: Trichogrammatinae) (~18,000 neurons), *Hemiptarsenus sp.* (Chalcidoidea: Eulophidae) (~35,000 neurons), and the honey bee *Apis mellifera* (Apidae) (~850,000–1,200,000 neurons). Moreover, by the final stage of *M. amalphanum* development, up to 95 percent of the neurons of the central nervous system have lost their nuclei [12, 13]. Nevertheless, adult wasps, which have an average lifespan of 5 days, still preserve the basic functional traits of hymenopteran insects including flight, mating and oviposition in hosts [14].

In this study, we present a *M. amalphanum* partial genome assembly and the adult transcriptome, and compare these with several parasitoid wasp species of different body sizes from the Chalcidoidea and Ichneumonoidea hymenopteran superfamilies. We performed general gene ontology and pathway analyses as well as specific gene categories of interest, such as chemosensory receptors and venom components. Additionally, we investigated transposable element (TE) content and dynamics across *M. amalphanum* and other parasitoid wasp species and analyzed the major components of the genome defense machinery. As body size reduction and loss of physiological or phenotypic traits is often correlated with genome size diminution [15, 16] and/or gene networks reduction [17], including chromatin diminution from the somatic tissues during embryogenesis[18, 19], we initially anticipated that the *M. amalphanum* genome would be greatly simplified.

## Material and methods

### Detailed information is presented in *Supplementary Information*

**Nucleic acid extraction and library construction.** *M. amalphanum* individuals were reared in the laboratory conditions from eggs of *Heliothrips haemorrhoidalis* (Thysanoptera:



**Fig 1. Size comparison of the parasitoid wasp *M. amalphanum* and bacterium *Thiomargarita namibiensis*.** (A) An adult stage of the parasitoid wasp *M. amalphanum* (image adapted from [5]), (B) *T. namibiensis*—the largest known bacterium (modified from Schulz et al. 1999) [11].

<https://doi.org/10.1371/journal.pone.0226485.g001>

Thripidae) collected in Santa Margherita, Northern Italy (44.32, 9.20). Unfortunately, we could collect only a dozen *M. amalphitanum* individuals because their habitats are difficult to detect (culture in the laboratory is currently impossible), the imago life span is short (5 days), and the animal is extremely small. With several insects we could cleanly recover, we were therefore able to obtain only around 1–5 ng of genomic DNA for the each paired-end DNA and cDNA libraries. DNA was extracted from ten individuals (males and females) using NucleoSpin Tissue XS kit (Macherey-Nagel, Germany) for each DNA-library. Three DNA libraries (DNA-library1 –whole insects; DNA-library2 –thorax and abdomen; DNA-library3 – head) were constructed using Ovation Ultralow Systems V2 kit (NuGEN, USA). Limited amount of biological material and low quantity of starting material (1–3 ng) did not permit construction of mate-paired libraries. Genome libraries were sequenced using Illumina HiSeq 1500 (Illumina, USA) with 150 bp paired-end reads. RNA was extracted from ten *M. amalphitanum* individuals (males and females) using the Trizol reagent (Thermo Fisher Scientific, USA) by a standard protocol, and cDNA libraries were constructed using Ovation RNA-Seq System V2 kit (NuGEN, USA) with poly(A) enrichment.

**Genome de novo assembly.** The output from Illumina sequencing of the genomic DNA library (source format \*.fastq) was used for *de novo* genome assembly. To assemble the genome of *M. amalphitanum*, we used 102,188,833 paired-end reads. Genome assemblies were constructed using different assembly algorithms, and their performance was compared to each other (S2 Fig). Then, *M. amalphitanum* reads were mapped to the final partial assembly with 92.3% conformity. Additionally, genomic DNA-libraries from thorax and abdomen (DNA-library2) of *M. amalphitanum* (SRR5982987) and from head (DNA-library3) of *M. amalphitanum* (SRR5982986) were prepared. In total, 79,317,970 (paired-end sequencing: 2×100 bp) and 85,409,775 (single-end sequencing: 50 bp) DNA reads were sequenced and used for *M. amalphitanum* coverage increase and as additional evidence during the search for potentially missing genes (S1 Table). Then, these reads were used for *de novo* building of the *M. amalphitanum* genome sequence by the SPAdes assembler (v.3.6.1) [20].

**Transcriptome de novo assembly.** Illumina RNA sequencing generated a total of 59,790,973 paired-end reads. Transcriptome *de novo* assembly was conducted using the default k-mer size in the Trinity software package (v. 2.4.0) [21], which combines three assembly algorithms: Inchworm, Chrysalis and Butterfly. Annotation of the *M. amalphitanum* transcriptome assembly was performed using the Trinotate pipeline [22].

**Transposable element (TE) de novo identification and analysis.** For *de novo* TE library construction, we used the REPET package [23] which combines three mutually complementing repeat identification tools (RECON, GROUPEP and PILER), yielding a combined repeat library with the average consensus sequence length of 1.66 kb (ranging from 157–14,640 bp). The outputs were subject to additional classification with the RepeatClassifier tool from the RepeatMasker package ([www.repeatmasker.org](http://www.repeatmasker.org)), which was also used to build the corresponding TE landscape divergence plots.

## Results and discussion

### Genome and transcriptome sequencing and assembly of *M. amalphitanum*

To gain insight into the genomic signatures of miniaturization that would distinguish *M. amalphitanum* from other Hymenoptera, we performed whole-genome shotgun sequencing of DNA (DNA-library1) isolated from ten adult individuals (males and females), using the Illumina platform (S1 Table). The resulting partial genome assembly (PRJNA344956) has a cumulative length of 346 megabases (Mb), with a scaffold N50 of 10,296 bp. The total genome coverage is 88.6-fold. Thus, the genome of *M. amalphitanum* is comparable in size with other

Chalcidoidea wasps, such as *Copidosoma floridanum*, *T. pretiosum* or *Nasonia vitripennis* [24, 25]. The best-performing combination of assembly software yielded contig N50 of 4,285 bp and allowed us to assemble 94,687 scaffolds from the low amounts of starting DNA material (Table 1; S2 Table; S1 Fig).

The *M. amalphantum* genome assemblies were evaluated with the BUSCO v3 (benchmarking universal single-copy orthologs) Hymenoptera gene set [26], which uses 4,415 near-universal single-copy orthologs to assess the relative completeness of genome assemblies. Through this analysis, 7.55% of the conserved genes were initially identified in the *M. amalphantum* assembly as putatively missing (S3 Table). More detailed information on our extensive search for the missing genes in *M. amalphantum* genome is presented below.

We also performed whole-body transcriptome analysis using RNA extracted from ten *M. amalphantum* individuals (males and females). Transcriptome *de novo* assembly (PRJNA344956) was performed using the Trinity software [21]. A total of 46,841 contigs were assembled with a mean length of 586 bp and an N50 of 633 bp from the low amounts of starting RNA material (S4 Table). The Illumina paired-end RNA-Seq data from *M. amalphantum* were mapped to the previously assembled genome using Bowtie2 [27]. Inspection of the alignments revealed that 79.95% of reads could be mapped to the genome. The BUSCO v3 statistics for the transcriptome assembly is also presented in Table 1; S3 Table.

The BUSCO analysis shows the low completeness of the present partial genome and transcriptome assemblies, with 28–29% of BUSCO genes listed as fragmented. This may be caused by inability to use mate-paired DNA-libraries or single-molecule sequencing (because of low amount of starting DNA material); possible high heterozygosity and/or significant structural variation between different parasitoid wasp individuals that were used for genome and transcriptome assemblies; BUSCO database incompleteness; and other factors. An additional factor in poor transcriptome completeness could be a high number of short and chimeric isotigs: while nearly 80% of transcriptome reads map to the genome, only 24% of assembled contigs are represented in the complete BUSCO set.

**Table 1. Final statistics of the genome and transcriptome assemblies of parasitoid wasp *Megaphragma amalphantum*.**

Genome assembly	
Number of contigs	94,687
Median (n:N50)	7,843
Contig N50 size	10,296
Maximum contig length, bp	895,906
Cumulative assembly size, bp	3.46×10 <sup>8</sup>
BUSCO assembly completeness, %	80.4
Fragmented, %	9.8
Transcriptome assembly	
Number of contigs	46,841
Median (n:N50)	13,109
Contig N50 size	633
Maximum contig length, bp	9,503
Cumulative assembly size, bp	3.74×10 <sup>7</sup>
BUSCO assembly completeness, %	24.65
Fragmented, %	28.12
Number of transcripts (BLASTX)	12,238

<https://doi.org/10.1371/journal.pone.0226485.t001>

### Gene ontology analysis

We used Gene Ontology (GO) analysis terms to describe characteristics of *M. amalphanum* gene products in three independent categories: biological processes (S2 Fig), molecular function (S3 Fig), and cellular components (S4 Fig). BLASTX outputs were used to retrieve the associated gene names and GO terms in all three categories (Table 2).

All *M. amalphanum* transcripts were matched to the Clusters of Orthologous Groups (COG) database to predict and classify their functions. In total, 8,810 genes were assigned to 25 COG functional categories. One of the largest groups is represented by the cluster for post-translational modification, protein turnover, and chaperones (988 counts; 10.7%), followed by intracellular trafficking, secretion, and vesicular transport (659 counts; 7.2%), DNA replication, recombination and repair (606 counts; 6.6%), signal transduction mechanisms (599 counts, 6.5%) and transcription (587; 6.4%) (S5 Fig).

To better understand incorporation of genes into diverse pathways, all annotated transcripts were mapped against the KEGG database for pathway-based analysis. As a result, 6,130 transcripts out of a total of 46,841 were assigned to a KEGG pathway, and were present in 328 different KEGG pathways. The KEGG pathway distribution is summarized in S6 Fig. The top pathways are biosynthesis of secondary metabolites (150 counts; 2.4%), RNA transport (100 counts; 1.6%), biosynthesis of antibiotics (95 counts; 1.5%), and spliceosome (94 counts; 1.5%).

The annotation of *M. amalphanum* and the available transcriptome assemblies of other parasitoid wasps from the families Trichogrammatidae (*T. pretiosum*, a lepidopteran egg parasitoid) and Braconidae including *Cotesia vestalis* (a diamondback moth parasitoid), *Dia-chasma alloeum* (an apple maggot parasitoid) and *Fopius arisanus* (tephritid fruit fly parasitoid) were used for comparative analysis of the most represented gene functions in parasitoids. We also used transcriptome assemblies from the Agaonidae fig wasp, *Ceratosolen solmsi*. We found significant similarities between *M. amalphanum*, *T. pretiosum* and *C. vestalis* major GO enrichment categories (S7–S9 Figs). At the same time, a significant number of transcripts related to DNA integration relative to other parasitoid wasps was found in *D. alloeum* and *M. amalphanum* (S7 Fig) (see below). Complete information about reference datasets used for *M. amalphanum* genome and transcriptome data analysis is shown in S5 Table. The Trinotate statistics for annotation of *M. amalphanum*, *C. solmsi*, *D. alloeum*, *F. arisanus*, *C. vestalis* and *T. pretiosum* transcriptome assemblies is presented in S6 Table.

### Potentially missing genes and missing or rapidly evolving gene clusters in the *M. amalphanum* genome

Given the incomplete nature of the *M. amalphanum* genome assembly (BUSCO coverage of ~80%, Table 1), we could perform only a preliminary assessment of potentially missing genes

Table 2. Basic Gene Ontology (GO) analysis terms for *M. amalphanum* gene products.

GO assignments of the transcripts	Transcript counts and percentage of total		
Biological processes	8,812 counts, 49.72%		
	Transcription	Regulation of transcription	DNA integration
	15%	10%	8%
Cellular components	4,802 counts, 27.10%		
	Nucleus and cytoplasm components	Integral membrane components	Plasma membrane components
	18%	9%	7%
Molecular functions	4,108 counts, 23.18%		
	ATP binding	Metal ion binding	Zinc ion binding
	17%	12%	10%

<https://doi.org/10.1371/journal.pone.0226485.t002>

and/or rapidly evolving gene clusters compared to other species. We clustered gene orthologs and identified gene clusters for each hymenopteran taxa (Chalcidoidea: *M. amalphanum*, *T. pretiosum*, *C. solmsi*, *C. floridanum*, and *N. vitripennis*; Ichneumonoidea: *D. alloeum* and *F. arisanus*; Apoidea: *A. mellifera*) using OrthoMCL [28]. The core gene set of all the hymenopteran species was composed of 6,278 gene clusters, 122 gene clusters were unique to the chalcid clade. 262 gene clusters were not detected in any of the chalcids analyzed (Supplementary Dataset 2; NCBI BioProject: PRJNA344956), but found in all the other hymenopterans, consistent with a similar recent analysis [29]. Our findings suggest that the loss of these genes apparently occurred in the last common ancestor of chalcids, or point to the possibility of parallel genome evolution across these species. Interestingly, the missing/rapidly evolving genes include homologs of genes that have important roles in embryonic patterning and development in other insects (e.g., *krueppel-1*, *knirps* or *short gastrulation* [29]).

To determine whether miniaturization in *M. amalphanum* is associated with significant gene loss that could be detected even in a partial genome assembly, we used genomic data of six larger hymenopteran species (*T. pretiosum*, *C. vestalis*, *C. floridanum*, *F. arisanus*, *N. vitripennis*, and *N. giraulti*), as well as the well-annotated genome of the honeybee (*A. mellifera*) as reference (body sizes are presented in S5 Table). We mapped the *M. amalphanum* (DNA-library1), *T. pretiosum*, *C. vestalis*, *C. floridanum*, *F. arisanus*, *N. vitripennis*, *N. giraulti* DNA reads on the *A. mellifera* genome sequence (PRJNA13343, PRJNA10625) (S10 Fig), and detected 115 genes that were not represented by *M. amalphanum* sequencing reads but were present in other parasitoid wasps. We then increased the coverage of the *M. amalphanum* genome to 146.8-fold by adding the reads from additional libraries (DNA-library2 and DNA-library3) (S1 Table) and observed the apparent absence of 114 of the 115 genes. An additional TBLASTX search identified 36 of these genes as present, yielding a total of 78 putatively missing genes (S7 Table). However, querying the *M. amalphanum* genome with the corresponding amino acid sequences from the closest wasp ortholog (*N. vitripennis* or *T. pretiosum*) in TBLASTN searches reduced the number of putatively missing genes to just five: centrosomin, phosphoglycerate mutase 5, phosphoglycerate mutase 5–2, 26S proteasome complex subunit DSS1, and mucin-1/nucleoporin NSP1-like. We detected short *M. amalphanum* genome sequences encoding protein fragments (~8–23 amino acid residues) with some similarity to four of them, suggesting that they may be in the process of degeneration in this species. Despite a thorough search, we were unable to find any homologous sequence related to *centrosomin* (*cnn*) gene either in the partially assembled genome or in our cDNA libraries. Although *cnn* is regarded as rapidly evolving [30], sequence homology can be readily discerned and orthologs are present in every other insect, including the parasitoid *T. pretiosum*, suggesting that this gene is specifically absent in *M. amalphanum*. In *Drosophila melanogaster*, Cnn has important roles at the centrosome in mitotic spindle formation, cytoskeleton organization and neuronal morphogenesis [31, 32], although these functions may not be indispensable because this species (and possibly other insects) possesses centrosome-independent mechanisms for spindle nucleation [33]. A fungal homolog of Cnn is involved in nuclear migration [34–36]. Since the presented genome assembly has only partial BUSCO coverage, the absence of *cnn* remains tentative. Globally, however, the analysis of the available genome assembly argues for relatively little gene loss in *M. amalphanum*. Confident identification of true gene losses in this species will require additional DNA sequencing and improved genome assembly.

### Chemosensory genes in the *M. amalphanum* genome

Chemosensory receptors are encoded by some of the largest gene families in insect genomes, reflecting their important and wide-ranging roles in detection of environmental odors and

tastants. We asked how these gene families have evolved in *M. amalphitanum*, whose central and peripheral nervous systems are highly reduced [2, 14]. The highly divergent sequences of chemosensory receptors and relatively short genomic contig lengths available for *M. amalphitanum* precluded accurate annotation of full-length sequences in this species for the majority of loci. Nevertheless, comparison with chemosensory receptor repertoires of other insects allowed us to define probable orthologous relationships with receptors of known function in other species and obtain initial estimates of the size of each family.

The most deeply conserved family of chemosensory receptors in insects are the Ionotropic Receptors (IRs), which are distantly related to ionotropic glutamate receptors [37, 38]. IRs function in heteromeric protein complexes comprising more broadly-expressed co-receptors with selectively expressed “tuning” IRs that determines sensory specificity. We identified orthologs of each of the co-receptors (*Ir8a*, *Ir25a* (two paralogs), *Ir93a* and *Ir76b*), as well as four genes encoding tuning IRs related to acid-sensing receptors in other species. We also identified orthologs of IR68a, which functions in hygrosensation [39] and IR21a, which functions in cool temperature-sensing [40, 41]. Overall, the repertoire of IRs in *M. amalphitanum* is therefore very similar in size and content to that of *N. vitripennis* [38].

Insects possess a second superfamily of chemosensory ion channels—distinguished by a heptahelical protein structure—comprising Odorant Receptor (OR) and Gustatory Receptor (GR) subfamilies, which generally function in detection of volatile and non-volatile stimuli, respectively [42–45]. Similar to IRs, ORs function in heteromeric complexes of a conserved co-receptor (ORCO) and a tuning OR. We identified an *M. amalphitanum* ortholog of *Orco* and 83 additional *Or*-related sequences. We caution that many of these *Or* sequences are small fragments (often located near the end of the assembled contigs), so it is currently difficult to determine whether these are intact genes or pseudogenes. Within the GR repertoire, we identified genes encoding proteins related to GR43a, a sensor of both external and internal fructose [46], two others similar to other insect sugar-sensing GRs [47], and 25 additional *Gr* gene fragments. The sizes of these repertoires are smaller than in *N. vitripennis* (300 *Ors* (including 76 pseudogenes) and 58 *Grs* (including 11 pseudogenes) [48]), but similar to non-miniaturized parasitoid wasps *Meteorus pulchricornis* and *Macrocentrus cingulum* [49, 50]. However, precise comparison with the latter two species is difficult, as receptors in these wasps were identified from antennal transcriptomes, thereby representing only one of these insects’ chemosensory organs.

In sum, these analyses reveal that despite drastic nervous system reduction, *M. amalphitanum* has retained the conserved chemosensory receptors of larger wasps (and other insects), and appears to have numerous additional order- or species-specific receptors to allow detection of environmental chemical cues.

### Venom components in the *M. amalphitanum* transcriptomic data

Parasitoid wasps often use venom to modify the metabolism of their hosts; toxins and their known or presumed biological functions are described in various species [51]. We investigated the presence of homologs of *N. vitripennis* toxin constituents in *M. amalphitanum* and other parasitoid wasps (*Megastigmus spermotrophus*, *N. vitripennis*, *C. solmsi*, *T. pretiosum*), using previously published venom data [52, 53] and the transcriptomes of chalcid wasps (S5 Table). We identified 28 transcripts encoding putative venom proteins (Fig 2; S8 Table); homologs of these are found in all investigated Chalcidoidea species (Table 3). Assuming that most of these candidates are truly conserved venom proteins among Chalcidoids, *M. amalphitanum* venom diversity does not seem to have been significantly affected by size reduction.

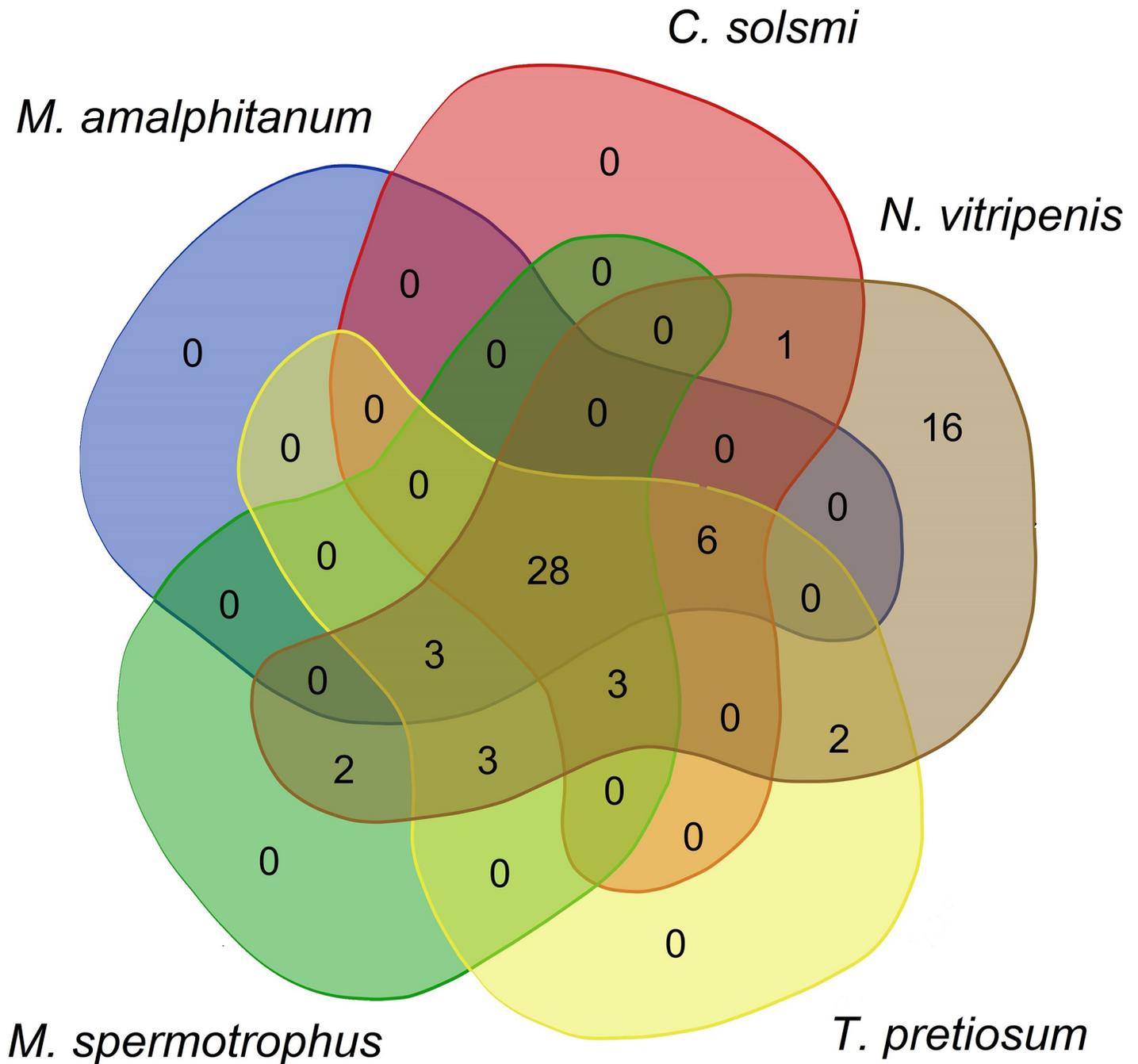


Fig 2. A Venn diagram showing *Nasonia vitripennis* venom components in other Chalcidoidea species: *M. spermotrophus*, *C. solmsi*, *T. pretiosum* and *M. amalphanum*.

<https://doi.org/10.1371/journal.pone.0226485.g002>

### *M. amalphanum* transposable elements and genome defense

Transposable elements (TEs) constitute a measurable fraction of virtually all eukaryotic genomes, and can play important roles in their function and evolution. In insects, TE activity has been implicated in evolution of eusociality, based on comparison of ten bee genomes with

**Table 3. Number of homologs of *N. vitripennis* venom (*N. vitripennis* toxin constituents) in *M. amalphanum* and other Chalcidoidea species based on Universal Chalcidoidea Database [54].**

Parasitoid wasp species	Families of Chalcidoidea	Number of <i>N. vitripennis</i> venom constituents	Body size, mm	Approximate number of hosts
<i>M. amalphanum</i>	Trichogrammatidae	37	0.25	2 insect species from one order
<i>C. solmsi</i>	Agaonidae	38	2.7	2 plant species from one family
<i>M. spermotrophus</i>	Torymidae	41	2.8	13 plant species from one family
<i>T. pretiosum</i>	Trichogrammatidae	45	0.5	>140 insect species from 4 orders
<i>N. vitripennis</i>	Pteromalidae	64	2.2	6* insect species from one order [55]

\* Universal Chalcidoidea Database lists >110 insect species from 8 orders [54]

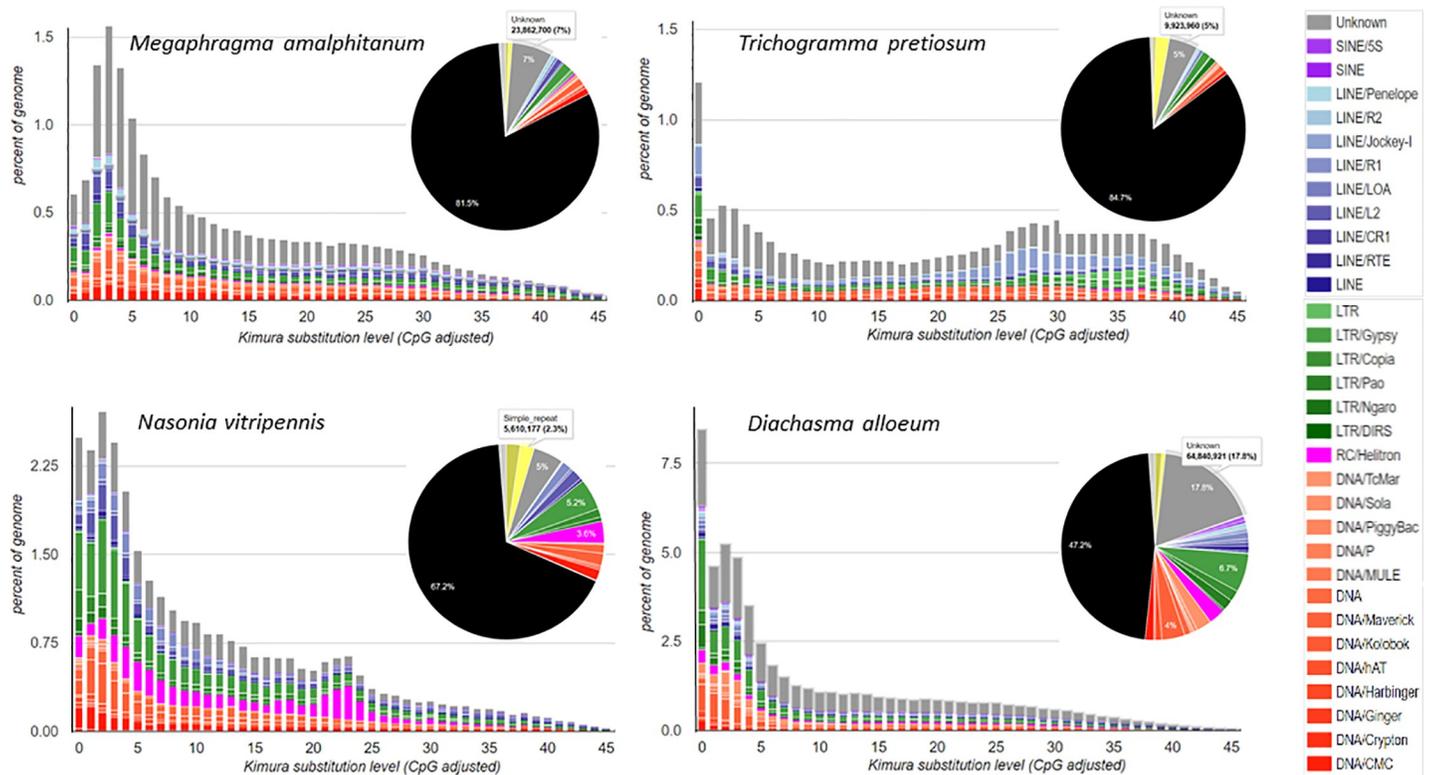
<https://doi.org/10.1371/journal.pone.0226485.t003>

increasing degrees of social complexity [56]. We performed *de novo* TE identification and comparative analysis of TE dynamics in *M. amalphanum* and in a representative set of larger wasp genomes for which TE content has previously been reported: the parasitoid *N. vitripennis* and two primitively eusocial aculeate wasps *Polistes canadensis* and *Polistes dominula* [12, 25, 57]. Additionally, we analyzed TEs in the genomes of parasitoid wasps *T. pretiosum* from the family Trichogrammatidae and *D. alloeum* from the family Braconidae.

For uniformity of measurements, we applied the same workflow to all genomes, without relying on pre-existing repeat libraries. We employed the REPET package for *de novo* TE identification (also used in [56]), and RepeatMasker for repeat classification and construction of TE landscape divergence plots. Comparison of the overall repeat content across six wasp species did not reveal substantial differences between four species (18.5% in *M. amalphanum* vs. 18.1%, 17.7% and 14.2% in *P. canadensis*, *P. dominula* and *T. pretiosum*, respectively). The *N. vitripennis* genome was 32.5% repetitive, in close agreement with the published estimate [25], and *D. alloeum* was highly repetitive at 52.8% (pie charts in Fig 3; S11 Fig). TE dynamics over time, which is shown on the corresponding TE landscape divergence plots, was found to differ substantially for *M. amalphanum*, which displayed a pronounced decline in recent TE activity after an initial increase, a pattern that is rarely observed in other hymenopterans [58, 59] (Fig 3).

While TE dynamics may be affected by different factors, the observed drop in active TE content in *M. amalphanum* may be relevant to the unique biology of this highly miniaturized insect. Its closest relative, *T. pretiosum*, is about 2-fold larger in body length. *Wolbachia* infection, which typically results in *T. pretiosum* parthenogenesis, can afterwards indirectly affect TE mobility in the host as a consequence of asexual reproduction, resulting in proliferation of specific TE families [58, 60, 61]. Other wasps do not display notable drops or spikes in current TE activity; TE inactivation was reported in two asexual mites [58], however it appears to be ancient and may have occurred prior to the abandonment of sex. Overall, the continued decline in *M. amalphanum* TE activity over the span of several million years—not observed in *T. pretiosum* which shares the most recent common ancestor with *M. amalphanum*—represents a rather unusual genomic feature compared to other hymenopteran we examined, including ants (not shown). We note, however, a recent comprehensive study [59] described two hymenopterans with a similar decline in recent TE activity (see below). No traces of *Wolbachia* infection or other representatives of the Rickettsiaceae family were found in *M. amalphanum* individuals [62], while the sequenced *T. pretiosum* carries the *Wolbachia* symbiont [63]; the sequenced *Nasonia* strain was maintained on antibiotics to cure it of infection.

To gain insights into possible reasons for reduction in TE activity after the initial burst, we investigated the major components of the genome defense machinery in *M. amalphanum*, including Dicer (Dcr)-like and Argonaute (Ago)/Piwi-like protein-coding genes. In insects,

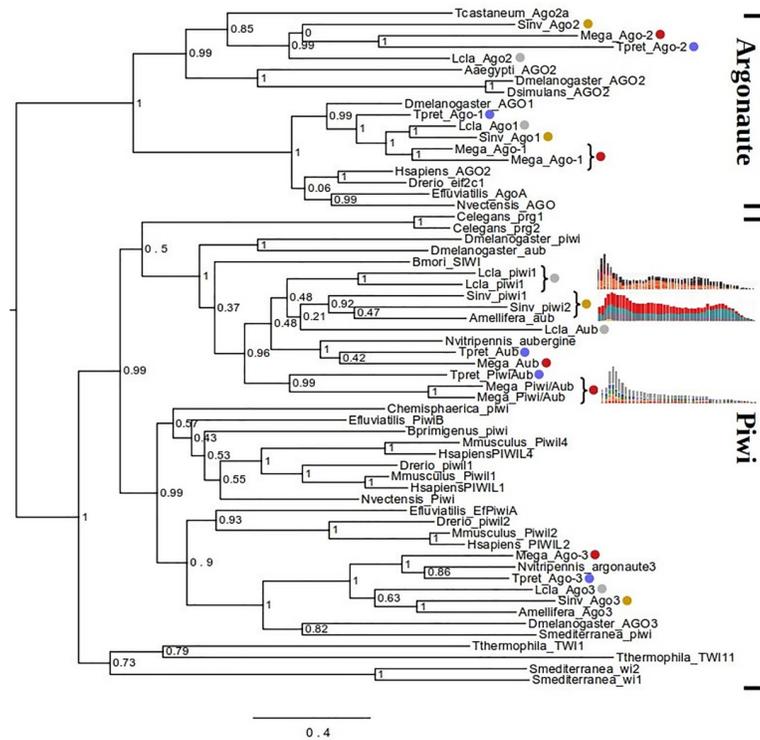


**Fig 3. Comparison of TE landscape divergence plots and TE genome fraction pie charts in four parasitoid wasp species: *M. amalphanum*, *T. pretiosum*, *N. vitripennis* and *D. alloenum*.**

<https://doi.org/10.1371/journal.pone.0226485.g003>

*Ago-1* and *Dcr-1* homologs represent the key components of the miRNA pathway; *Ago-2* and *Dcr-2* mediate antiviral RNA interference; and *Piwi* and *Ago-3/Aub* suppress TE activity in the germline [64]. Both *M. amalphanum* and *T. pretiosum* possess equal numbers of *Dcr-1* and *Dcr-2* homologs, as well as *Ago-2* and *Ago-3* homologs (S12 Fig). However, in *M. amalphanum*, the *Ago-1* and the *Piwi/Aub* homologs underwent a relatively recent duplication in comparison to *T. pretiosum* (Fig 4). This may indicate additional layers of enforcement in the miRNA and piRNA pathways of *M. amalphanum*, both of which should result in suppression of TE activity. Indeed, after inspecting the genomes of two other sequenced hymenopteran species showing recent declines in TE activity (*Leptopilina clavipes* and *Solenopsis invicta*; [58, 59]), we found that they also display relatively recent duplications of *Piwi*-like proteins (Fig 4).

The drop in TE activity is also evident from the transcriptome analysis. The GO radar plot (S7 Fig) shows a substantial number of short contigs related to DNA integration, most of which upon inspection were found to represent separate fragments of gypsy-like and copia-like LTR retrotransposons, and a few belong to Polinton, P and Ginger DNA TEs. Transcriptionally active copies fall into two groups: first, those which apparently proliferated during the burst of TE activity and have since accumulated debilitating mutations making them incapable of transposition, but still retain a certain level of transcriptional activity; second, those that originate from recent infections by retrovirus-like TEs and contain uninterrupted ORFs, but are not actively proliferating and are present at very few genomic loci. Comparison of BLASTN hits for *M. amalphanum* integrase-related TE transcripts showed that high-copy hits represent MITEs (S13 Fig). We hypothesize that actively proliferating TE copies represent recent arrivals, possibly brought about by viruses or host-parasite interactions [65].



**Fig 4. Maximum likelihood analysis of phylogenetic relationships between Piwi/Argonaute coding sequences.** Colored dots denote sequences from *T. pretiosum* (blue), *L. clavipes* (gray), *S. invicta* (yellow) and *M. amalphantanum* (red). Recent duplications in the latter three hymenopterans are indicated by curly brackets, and the corresponding TE divergence plots from [58, 59] and Fig 3 are placed next to each curly bracket. Phylogeny analysis and notations are as in S12 Fig.

<https://doi.org/10.1371/journal.pone.0226485.g004>

### Concluding remarks

Our study provides a first view of the genomic content of one of the smallest insects currently known, the parasitoid wasp *M. amalphantanum*. In contrast to the expectation that the small body size, in combination with the parasitic lifestyle, should lead to significant reduction in the amount of genomic DNA and in gene content, we do not observe a drastic reduction in the overall genome size or in the number of expressed genes in comparison with larger parasitic wasps. However, the multiple experimental constraints described above limit the quality of genome and transcriptome assemblies. In the future, improved genomic studies in this species (and other Hymenoptera) will be essential to confidently assess specific genetic adaptations that may be linked with body miniaturization.

Interestingly, transposable element dynamics over time were found to differ substantially between the analyzed wasp species, with *M. amalphantanum* displaying a relatively recent decline in TE activity preceded by a burst, a pattern not observed in most other parasitoid wasps. The decline in TE activity may have been associated with evolution of additional *Ago* and *Piwi* copies, not present in *T. pretiosum*, which could have reinforced the genome defense machinery to prevent uncontrolled TE expansion. This hypothesis is strengthened by identifying duplications of Piwi-like proteins accompanied by a decline in TE activity over time in two additional species of Hymenoptera; by contrast, most other hymenopterans show no such decline.

The relationship between body size and genome size has been discussed for a long time. Significant correlations of these values have been described for flatworms and copepods [16]; by contrast, such correlations were not found in ants [66]. Our results show that body size reduction in hymenopterans is not accompanied by greatly decreased transcriptomic and genomic complexity. This observation begs the question of how miniaturization is encoded genetically. We hypothesize that changes in regulatory sequences, rather than gene content, were important in the process of body size reduction, similar to mechanisms of morphological evolution that have driven adaptive diversification in all animals, great or small [67].

## Supporting information

**S1 Fig. *M. amalphantum* genome assembly statistics using ABySS, SPAdes, and Velvet software.** K-mer sizes were matched for ABySS, SPAdes and Velvet. Note: CLC Genomics Workbench does not use k-mer size; CLC assembly was performed with default settings, and the statistics are given in [S2 Table](#).

(PNG)

**S2 Fig. Gene ontology analysis of *M. amalphantum* transcriptome for contigs with assigned GO: Biological processes.**

(TIF)

**S3 Fig. Gene ontology analysis of *M. amalphantum* transcriptome for contigs with assigned GO: Molecular function.**

(TIF)

**S4 Fig. Gene ontology analysis of *M. amalphantum* transcriptome for contigs with assigned GO: Cellular components.**

(TIF)

**S5 Fig. The Clusters of Orthologous Groups (COG) for *M. amalphantum* transcriptome (top pathways).**

(TIF)

**S6 Fig. KEGG pathway analysis for the *M. amalphantum* transcriptome.**

(TIFF)

**S7 Fig. Radar plot for the *M. amalphantum*, *C. solmsi*, *D. alloenum*, *F. arisanus*, *C. vestalis*, *T. pretiosum* transcriptome GO-category related to biological processes showing numbers of transcripts in this GO-category for six Chalcidoid species.**

(TIF)

**S8 Fig. Radar plot for the *M. amalphantum*, *C. solmsi*, *D. alloenum*, *F. arisanus*, *C. vestalis*, *T. pretiosum* transcriptome GO-category related to cellular components showing numbers of transcripts in this GO-category for six Chalcidoid species.**

(TIF)

**S9 Fig. Radar plot for the *M. amalphantum*, *C. solmsi*, *D. alloenum*, *F. arisanus*, *C. vestalis*, *T. pretiosum* transcriptome GO-category related to molecular processes showing numbers of transcripts in this GO-category for six Chalcidoid species.**

(TIF)

**S10 Fig. Potentially missing genes in the *M. amalphantum* partial genome assembly.** Y-axis: number of genes; X-axis: number of hymenopteran genomes analysed.

(TIF)

**S11 Fig. Effects of re-classification of “unknown” repeats in the *de novo* library for *M. amalphanum* and *P. dominula* (Supplementary Notes B6). v2, re-classified.**

(TIF)

**S12 Fig. Maximum likelihood analysis of phylogenetic relationships among eukaryotic Dicer homologs from animals, plants, and fungi. *M. amalphanum* and *T. pretiosum* Dcr-1 and Dcr-2 homologs are denoted by red dots.** Multiple alignments of CDS sequences were performed using Muscle v3.8 with default settings. Phylogenetic trees were generated under the maximum likelihood criterion using PhyML 3.0 (GTR model, NNI topological moves and likelihood branch supports). All manipulations of phylogenetic trees were performed using FigTree. Scale bar, nucleotide substitutions per site.

(PNG)

**S13 Fig. Box plot of percent identity between BLASTN hits for *M. amalphanum* integrase-related TE transcripts, binned by copy count. High-copy hits represent MITES.**

(PNG)

**S14 Fig. An overview of the missing gene analysis pipeline and its results.**

(TIF)

**S1 Table. Paired-end DNA-libraries used for *M. amalphanum* genome sequencing.**

(DOCX)

**S2 Table. *M. amalphanum* genome assembly statistics using ABySS, SPAdes, CLC and Velvet software (contigs).**

(DOCX)

**S3 Table. Evaluation of the *M. amalphanum* genome and transcriptome assemblies using the BUSCO v3 (benchmarking universal single-copy orthologs) Hymenoptera gene set.**

(DOCX)

**S4 Table. *M. amalphanum* and *C. solmsi* transcriptome assembly statistics using Trinity software (contigs).**

(DOCX)

**S5 Table. Reference data sets used for *M. amalphanum* genome and transcriptome data analysis.**

(DOCX)

**S6 Table. Trinotate statistics for *M. amalphanum*, *C. solmsi*, *D. allozum*, *F. arisanus*, *C. vestalis*, *T. pretiosum* transcriptome assemblies.**

(DOCX)

**S7 Table. A set of 78 genes (paralogs and homologs) not covered by *M. amalphanum* reads.**

(DOCX)

**S8 Table. Common putative venom constituents in Chalcidoidea parasitoid wasps *M. amalphanum*, *C. solmsi*, *M. spermotrophus*, *T. pretiosum*, *N. vitripennis*.**

(DOCX)

## Acknowledgments

Konstantin Skryabin passed away before the submission of the final version of this manuscript. Artem Nedoluzhko accepts responsibility for the integrity and validity of the data collected

and analyzed. This work was carried out using high-performance computing resources of federal center for collective usage at NRC “Kurchatov Institute”, <http://computing.kiae.ru/>. Support of this project was provided by the Russian Scientific Foundation (RSF) grant #14-24-00175 and RSF grant #14-50-00060. B.L. was supported by the Rosenthal Brown-MBL internship and the REU supplement to NSF MCB-1121334 to I.A. R.B. was supported by a European Research Council Consolidator Grant (615094).

## Author Contributions

**Conceptualization:** Artem V. Nedoluzhko, Richard Benton, Irina R. Arkhipova, Egor B. Prokhortchouk, Konstantin G. Skryabin.

**Data curation:** Fedor S. Sharko, Artem V. Nedoluzhko, Irina R. Arkhipova, Egor B. Prokhortchouk.

**Formal analysis:** Fedor S. Sharko, Artem V. Nedoluzhko, Brandon M. Lê, Sergey M. Rastorguev, Alexey S. Sokolov, Fernando Rodriguez, Richard Benton, Michael B. Evgen'ev, Irina R. Arkhipova, Egor B. Prokhortchouk.

**Funding acquisition:** Egor B. Prokhortchouk, Konstantin G. Skryabin.

**Investigation:** Artem V. Nedoluzhko, Alexey A. Polilov, Richard Benton, Michael B. Evgen'ev, Irina R. Arkhipova, Egor B. Prokhortchouk.

**Methodology:** Artem V. Nedoluzhko, Brandon M. Lê, Eugenia S. Boulygina, Richard Benton, Irina R. Arkhipova, Egor B. Prokhortchouk.

**Project administration:** Alexander M. Mazur, Egor B. Prokhortchouk.

**Resources:** Alexey A. Polilov, Egor B. Prokhortchouk.

**Software:** Brandon M. Lê, Fernando Rodriguez, Richard Benton, Irina R. Arkhipova.

**Supervision:** Irina R. Arkhipova, Egor B. Prokhortchouk.

**Validation:** Artem V. Nedoluzhko, Irina R. Arkhipova, Egor B. Prokhortchouk.

**Visualization:** Artem V. Nedoluzhko, Brandon M. Lê, Richard Benton, Irina R. Arkhipova.

**Writing – original draft:** Fedor S. Sharko, Artem V. Nedoluzhko, Richard Benton, Michael B. Evgen'ev, Irina R. Arkhipova, Egor B. Prokhortchouk.

**Writing – review & editing:** Artem V. Nedoluzhko, Svetlana V. Tsygankova, Eugenia S. Boulygina, Sergey M. Rastorguev, Alexey A. Polilov, Richard Benton, Michael B. Evgen'ev, Irina R. Arkhipova, Egor B. Prokhortchouk, Konstantin G. Skryabin.

## References

1. Hanken J, Wake DB. Miniaturization of Body-Size—Organismal Consequences and Evolutionary Significance. *Annu Rev Ecol Syst.* 1993; 24:501–19. <https://doi.org/10.1146/annurev.es.24.110193.002441> ISI:A1993MJ37100018.
2. Polilov AA. Small is beautiful: features of the smallest insects and limits to miniaturization. *Annu Rev Entomol.* 2015; 60:103–21. Epub 2014/10/24. <https://doi.org/10.1146/annurev-ento-010814-020924> PMID: 25341106.
3. Bernardo U, Viggiani G. Biological data on *Megaphragma amalphanum* Viggiani and *Megaphragma mymaripenne* Timberlake (Hymenoptera: Trichogrammatidae), egg-parasitoid of *H. haemorrhoidalis* Bouché (Thysanoptera: Thripidae) in southern Italy. *Bollettino del Laboratorio di Entomologia Agraria Filippo Silvestri.* 2002; 58:77–85.

4. Pintureau B, Lassabliere F, Khatchadourian C, Daumal J. Eggs parasitoids and symbionts of two European Thrips. *Annales de la Société Entomologique de France*. 1999; 35:416–20. ISI:000085892500074.
5. Polilov AA. Anatomy of adult *Megaphragma* (Hymenoptera: Trichogrammatidae), one of the smallest insects, and new insight into insect miniaturization. *PLoS One*. 2017; 12(5):e0175566. Epub 2017/05/04. <https://doi.org/10.1371/journal.pone.0175566> PMID: 28467417
6. Branstetter MG, Danforth BN, Pitts JP, Faircloth BC, Ward PS, Buffington ML, et al. Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of Ants and Bees. *Current Biology*. 2017; 27(7):1019–25. <https://doi.org/10.1016/j.cub.2017.03.027> ISI:000398061700023. PMID: 28376325
7. Nedoluzhko AV, Sharko FS, Boulygina ES, Tsygankova SV, Sokolov AS, Mazur AM, et al. Mitochondrial genome of *Megaphragma amalphanum* (Hymenoptera: Trichogrammatidae). *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016; 27(6):4526–7. Epub 2015/12/01. <https://doi.org/10.3109/19401736.2015.1101546> PMID: 26617282.
8. Owen AK, George J, Pinto JD, Heraty JM. A molecular phylogeny of the Trichogrammatidae (Hymenoptera: Chalcidoidea), with an evaluation of the utility of their male genitalia for higher level classification. *Syst Entomol*. 2007; 32(2):227–51. <https://doi.org/10.1111/j.1365-3113.2006.00361.x> ISI:000245745800003.
9. Heraty JM, Burks RA, Cruaud A, Gibson GAP, Liljeblad J, Munro J, et al. A phylogenetic analysis of the megadiverse Chalcidoidea (Hymenoptera). *Cladistics*. 2013; 29(5):466–542. <https://doi.org/10.1111/cla.12006> ISI:000324622500003.
10. Peters RS, Niehuis O, Gunkel S, Blaser M, Mayer C, Podsiadlowski L, et al. Transcriptome sequence-based phylogeny of chalcidoid wasps (Hymenoptera: Chalcidoidea) reveals a history of rapid radiations, convergence, and evolutionary success. *Mol Phylogenet Evol*. 2018; 120:286–96. Epub 2017/12/17. <https://doi.org/10.1016/j.ympev.2017.12.005> PMID: 29247847
11. Schulz HN, Brinkhoff T, Ferdelman TG, Marine MH, Teske A, Jorgensen BB. Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science*. 1999; 284(5413):493–5. Epub 1999/04/16. <https://doi.org/10.1126/science.284.5413.493> PMID: 10205058.
12. Standage DS, Berens AJ, Glastad KM, Severin AJ, Brendel VP, Toth AL. Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. *Mol Ecol*. 2016; 25(8):1769–84. Epub 2016/02/10. <https://doi.org/10.1111/mec.13578> PMID: 26859767.
13. Makarova AA, Polilov AA. Peculiarities of the Brain Structure and Ultrastructure in Small Insects Related to Miniaturization. 1. The Smallest Coleoptera (Ptiliidae). *Zool Zh*. 2013; 92(5):523–33. <https://doi.org/10.7868/S0044513413050073> ISI:000321472300003.
14. Polilov AA. The smallest insects evolve anucleate neurons. *Arthropod Struct Dev*. 2012; 41(1):29–34. Epub 2011/11/15. <https://doi.org/10.1016/j.asd.2011.09.001> PMID: 22078364.
15. Gregory TR. Synergy between sequence and size in large-scale genomics. *Nat Rev Genet*. 2005; 6(9):699–708. Epub 2005/09/10. nrg1674 [pii] <https://doi.org/10.1038/nrg1674> PMID: 16151375.
16. Gregory TR, Hebert PDN, Kolasa J. Evolutionary implications of the relationship between genome size and body size in flatworms and copepods. *Heredity*. 2000; 84(2):201–8. <https://doi.org/10.1046/j.1365-2540.2000.00661.x> ISI:000086196500009. PMID: 10762390
17. Erives AJ. Genes conserved in bilaterians but jointly lost with *Myc* during nematode evolution are enriched in cell proliferation and cell migration functions. *Dev Genes Evol*. 2015; 225(5):259–73. Epub 2015/07/16. <https://doi.org/10.1007/s00427-015-0508-1> PMID: 26173873; PubMed Central PMCID: PMC4568025.
18. Akif'ev AP, Grishanin AK, Degtiarev SV. [Chromatin diminution is a key process explaining the eukaryotic genome size paradox and some mechanisms of genetic isolation]. *Genetika*. 2002; 38(5):595–606. Epub 2002/06/19. PMID: 12068542.
19. Wang J, Mitreva M, Berriman M, Thorne A, Magrini V, Koutsovoulos G, et al. Silencing of germline-expressed genes by DNA elimination in somatic cells. *Dev Cell*. 2012; 23(5):1072–80. Epub 2012/11/06. <https://doi.org/10.1016/j.devcel.2012.09.020> PMID: 23123092
20. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012; 19(5):455–77. Epub 2012/04/18. <https://doi.org/10.1089/cmb.2012.0021> PubMed Central PMCID: PMC3342519. PMID: 22506599
21. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*. 2011; 29(7):644–U130. <https://doi.org/10.1038/nbt.1883> ISI:000292595200023. PMID: 21572440
22. Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, et al. A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Rep*. 2017;

- 18(3):762–76. Epub 2017/01/19. <https://doi.org/10.1016/j.celrep.2016.12.063> PMID: 28099853; PubMed Central PMCID: PMC5419050.
23. Flutre T, Duprat E, Feuillet C, Quesneville H. Considering transposable element diversification in de novo annotation approaches. *PLoS One*. 2011; 6(1):e16526. Epub 2011/02/10. <https://doi.org/10.1371/journal.pone.0016526> PMID: 21304975; PubMed Central PMCID: PMC3031573.
24. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J Hered*. 2013; 104(5):595–600. Epub 2013/08/14. <https://doi.org/10.1093/jhered/est050> PMID: 23940263; PubMed Central PMCID: PMC4046820.
25. Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, et al. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science*. 2010; 327(5963):343–8. Epub 2010/01/16. <https://doi.org/10.1126/science.1178028> PMID: 20075255
26. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015; 31(19):3210–2. Epub 2015/06/11. <https://doi.org/10.1093/bioinformatics/btv351> PMID: 26059717.
27. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9(4):357–9. Epub 2012/03/06. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286; PubMed Central PMCID: PMC3322381.
28. Li L, Stoeckert CJ Jr., Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003; 13(9):2178–89. Epub 2003/09/04. <https://doi.org/10.1101/gr.1224503> PMID: 12952885; PubMed Central PMCID: PMC403725.
29. Lindsey ARI, Kelkar YD, Wu X, Sun D, Martinson EO, Yan Z, et al. Comparative genomics of the miniature wasp and pest control agent *Trichogramma pretiosum*. *BMC Biol*. 2018; 16(1):54. Epub 2018/05/20. <https://doi.org/10.1186/s12915-018-0520-9> PMID: 29776407; PubMed Central PMCID: PMC5960102.
30. Eisman RC, Kaufman TC. Probing the boundaries of orthology: the unanticipated rapid evolution of *Drosophila* centrosomin. *Genetics*. 2013; 194(4):903–26. Epub 2013/06/12. <https://doi.org/10.1534/genetics.113.152546> PMID: 23749319; PubMed Central PMCID: PMC3730919.
31. Feng Z, Caballe A, Wainman A, Johnson S, Haensele AFM, Cottee MA, et al. Structural Basis for Mitotic Centrosome Assembly in Flies. *Cell*. 2017; 169(6):1078–89 e13. Epub 2017/06/03. <https://doi.org/10.1016/j.cell.2017.05.030> PMID: 28575671
32. Yalgin C, Ebrahimi S, Delandre C, Yoong LF, Akimoto S, Tran H, et al. Centrosomin represses dendrite branching by orienting microtubule nucleation. *Nat Neurosci*. 2015; 18(10):1437–45. Epub 2015/09/01. <https://doi.org/10.1038/nn.4099> PMID: 26322925.
33. Basto R, Lau J, Vinogradova T, Gardiol A, Woods CG, Khodjakov A, et al. Flies without centrioles. *Cell*. 2006; 125(7):1375–86. <https://doi.org/10.1016/j.cell.2006.05.025> ISI:000239104500019. PMID: 16814722
34. Zheng Z, Gao T, Hou Y, Zhou M. Involvement of the anucleate primary sterigmata protein FgApsB in vegetative differentiation, asexual development, nuclear migration, and virulence in *Fusarium graminearum*. *FEMS Microbiol Lett*. 2013; 349(2):88–98. Epub 2013/10/15. <https://doi.org/10.1111/1574-6968.12297> PMID: 24117691.
35. Veith D, Scherr N, Efimov VP, Fischer R. Role of the spindle-pole-body protein ApsB and the cortex protein ApsA in microtubule organization and nuclear migration in *Aspergillus nidulans*. *J Cell Sci*. 2005; 118(Pt 16):3705–16. Epub 2005/08/18. <https://doi.org/10.1242/jcs.02501> PMID: 16105883
36. Suelmann R, Sievers N, Galetzka D, Robertson L, Timberlake WE, Fischer R. Increased nuclear traffic chaos in hyphae of *Aspergillus nidulans*: molecular characterization of *apsB* and in vivo observation of nuclear behaviour. *Mol Microbiol*. 1998; 30(4):831–42. Epub 1999/03/27. <https://doi.org/10.1046/j.1365-2958.1998.01115.x> PMID: 10094631.
37. Benton R, Vannice KS, Gomez-Diaz C, Vossball LB. Variant Ionotropic Glutamate Receptors as Chemosensory Receptors in *Drosophila*. *Cell*. 2009; 136(1):149–62. <https://doi.org/10.1016/j.cell.2008.12.001> ISI:000262318400022. PMID: 19135896
38. Rytz R, Croset V, Benton R. Ionotropic Receptors (IRs): Chemosensory ionotropic glutamate receptors in *Drosophila* and beyond. *Insect Biochem Molec*. 2013; 43(9):888–97. <https://doi.org/10.1016/j.ibmb.2013.02.007> ISI:000323801100011. PMID: 23459169
39. Knecht ZA, Silbering AF, Cruz J, Yang L, Croset V, Benton R, et al. Ionotropic Receptor-dependent moist and dry cells control hygrosensation in *Drosophila*. *Elife*. 2017; 6. Epub 2017/06/18. <https://doi.org/10.7554/eLife.26654e26654> [pii] PMID: 28621663; PubMed Central PMCID: PMC5495567.
40. Knecht ZA, Silbering AF, Ni L, Klein M, Budelli G, Bell R, et al. Distinct combinations of variant ionotropic glutamate receptors mediate thermosensation and hygrosensation in *Drosophila*. *Elife*. 2016; 5. Epub 2016/09/24. <https://doi.org/10.7554/eLife.17879e17879> [pii] PMID: 27656904; PubMed Central PMCID: PMC5052030.

41. Ni L, Klein M, Svec KV, Budelli G, Chang EC, Ferrer AJ, et al. The Ionotropic Receptors IR21a and IR25a mediate cool sensing in *Drosophila*. *Elife*. 2016; 5. Epub 2016/04/30. <https://doi.org/10.7554/eLife.13254> PMID: 27126188
42. Benton R. Multigene Family Evolution: Perspectives from Insect Chemoreceptors. *Trends Ecol Evol*. 2015; 30(10):590–600. Epub 2015/09/29. <https://doi.org/10.1016/j.tree.2015.07.009> PMID: 26411616
43. Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, et al. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet*. 2010; 6(8):e1001064. Epub 2010/09/03. <https://doi.org/10.1371/journal.pgen.1001064> PMID: 20808886
44. Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB. Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron*. 2004; 43(5):703–14. Epub 2004/09/02. <https://doi.org/10.1016/j.neuron.2004.08.019> PMID: 15339651
45. Joseph RM, Carlson JR. *Drosophila* Chemoreceptors: A Molecular Interface Between the Chemical World and the Brain. *Trends in Genetics*. 2015; 31(12):683–95. <https://doi.org/10.1016/j.tig.2015.09.005> ISI:000366346000002. PMID: 26477743
46. Miyamoto T, Slone J, Song XY, Amrein H. A Fructose Receptor Functions as a Nutrient Sensor in the *Drosophila* Brain. *Cell*. 2012; 151(5):1113–25. <https://doi.org/10.1016/j.cell.2012.10.024> ISI:000311423500022. PMID: 23178127
47. Fujii S, Yavuz A, Slone J, Jagge C, Song X, Amrein H. *Drosophila* sugar receptors in sweet taste perception, olfaction, and internal nutrient sensing. *Curr Biol*. 2015; 25(5):621–7. Epub 2015/02/24. <https://doi.org/10.1016/j.cub.2014.12.058> PMID: 25702577
48. Robertson HM, Gadau J, Wanner KW. The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis*. *Insect Mol Biol*. 2010; 19 Suppl 1:121–36. Epub 2010/02/20. <https://doi.org/10.1111/j.1365-2583.2009.00979.x> IMB979 [pii] PMID: 20167023.
49. Ahmed T, Zhang T, Wang Z, He K, Bai S. Gene set of chemosensory receptors in the polyembryonic endoparasitoid *Macrocentrus cingulum*. *Sci Rep*. 2016; 6:24078. Epub 2016/04/20. <https://doi.org/10.1038/srep24078> PMID: 27090020
50. Sheng S, Liao CW, Zheng Y, Zhou Y, Xu Y, Song WM, et al. Candidate chemosensory genes identified in the endoparasitoid *Meteorus pulchricornis* (Hymenoptera: Braconidae) by antennal transcriptome analysis. *Comp Biochem Physiol Part D Genomics Proteomics*. 2017; 22:20–31. Epub 2017/02/12. <https://doi.org/10.1016/j.cbd.2017.01.002> PMID: 28187311
51. Moreau SJM, Asgari S. Venom Proteins from Parasitoid Wasps and Their Biological Functions. *Toxins*. 2015; 7(7):2385–412. <https://doi.org/10.3390/toxins7072385> ISI:000359191200004. PMID: 26131769
52. de Graaf DC, Aerts M, Brunain M, Desjardins CA, Jacobs FJ, Werren JH, et al. Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatic and proteomic studies. *Insect Mol Biol*. 2010; 19 Suppl 1:11–26. Epub 2010/02/20. <https://doi.org/10.1111/j.1365-2583.2009.00914.x> PMID: 20167014
53. Paulson AR, Le CH, Dickson JC, Ehling J, von Aderkas P, Perlman SJ. Transcriptome analysis provides insight into venom evolution in a seed-parasitic wasp, *Megastigmus spermotrophus*. *Insect Molecular Biology*. 2016; 25(5):604–16. <https://doi.org/10.1111/imb.12247> ISI:000383345000008. PMID: 27286234
54. Universal Chalcidoidea Database. World Wide Web electronic publication. [Internet]. 2019. Available from: <http://www.nhm.ac.uk/chalcidooids>.
55. Peters RS. Host Range and Offspring Quantities in Natural Populations of *Nasonia vitripennis* (Walker, 1836) (Hymenoptera: Chalcidoidea: Pteromalidae). *J Hymenopt Res*. 2010; 19(1):128–38. WOS:000285775800012.
56. Kapheim KM, Pan H, Li C, Salzberg SL, Puiu D, Magoc T, et al. Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. *Science*. 2015; 348(6239):1139–43. Epub 2015/05/16. <https://doi.org/10.1126/science.aaa4788> PMID: 25977371
57. Patalano S, Vlasova A, Wyatt C, Ewels P, Camara F, Ferreira PG, et al. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc Natl Acad Sci U S A*. 2015; 112(45):13970–5. Epub 2015/10/21. <https://doi.org/10.1073/pnas.1515937112> PMID: 26483466
58. Bast J, Schaefer I, Schwander T, Maraun M, Scheu S, Kraaijeveld K. No Accumulation of Transposable Elements in Asexual Arthropods. *Mol Biol Evol*. 2016; 33(3):697–706. Epub 2015/11/13. <https://doi.org/10.1093/molbev/msv261> PMID: 26560353; PubMed Central PMCID: PMC4760076.
59. Petersen M, Armisen D, Gibbs RA, Hering L, Khila A, Mayer G, et al. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC Evol Biol*. 2019; 19(1):11. Epub 2019/01/11. <https://doi.org/10.1186/s12862-018-1324-9> PMID: 30626321; PubMed Central PMCID: PMC6327564.

60. Kraaijeveld K, Bast J. Transposable element proliferation as a possible side effect of endosymbiont manipulations. *Mob Genet Elements*. 2012; 2(5):253–6. Epub 2013/04/04. <https://doi.org/10.4161/mge.22878> PMID: 23550173; PubMed Central PMCID: PMC3575435.
61. Kraaijeveld K, Zwanenburg B, Hubert B, Vieira C, De Pater S, Van Alphen JJ, et al. Transposon proliferation in an asexual parasitoid. *Mol Ecol*. 2012; 21(16):3898–906. Epub 2012/05/03. <https://doi.org/10.1111/j.1365-294X.2012.5582.x> PMID: 22548357.
62. Nedoluzhko AV, Sharko FS, Tsygankova SV, Boulygina ES, Sokolov AS, Rastorguev SM, et al. Metagenomic analysis of microbial community of a parasitoid wasp *Megaphragma amalphanum*. *Genom Data*. 2017; 11:87–8. Epub 2017/01/10. <https://doi.org/10.1016/j.gdata.2016.12.007> PMID: 28066711; PubMed Central PMCID: PMC5200880.
63. Lindsey AR, Werren JH, Richards S, Stouthamer R. Comparative Genomics of a Parthenogenesis-Inducing *Wolbachia* Symbiont. *G3 (Bethesda)*. 2016; 6(7):2113–23. Epub 2016/05/20. <https://doi.org/10.1534/g3.116.028449> PMID: 27194801
64. Obbard DJ, Gordon KH, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. *Philos Trans R Soc Lond B Biol Sci*. 2009; 364(1513):99–115. Epub 2008/10/18. <https://doi.org/10.1098/rstb.2008.0168> PMID: 18926973; PubMed Central PMCID: PMC2592633.
65. Schaack S, Gilbert C, Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol*. 2010; 25(9):537–46. Epub 2010/07/02. <https://doi.org/10.1016/j.tree.2010.06.001> PMID: 20591532
66. Tsutsui ND, Suarez AV, Spagna JC, Johnston JS. The evolution of genome size in ants. *BMC Evol Biol*. 2008; 8:64. Epub 2008/02/28. <https://doi.org/10.1186/1471-2148-8-64> PMID: 18302783; PubMed Central PMCID: PMC2268675.
67. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*. 2008; 134(1):25–36. Epub 2008/07/11. <https://doi.org/10.1016/j.cell.2008.06.030> PMID: 18614008