

RESEARCH

Open Access



Chromosome-level genome assemblies of two parasitoid biocontrol wasps reveal the parthenogenesis mechanism and an associated novel virus

Sarah N. Inwood^{1†}, John Skelly^{1,2†}, Joseph G. Guhlin³, Thomas W.R. Harrop⁴, Stephen L. Goldson⁵ and Peter K. Dearden^{1,3*}

Abstract

Background Biocontrol is a key technology for the control of pest species. *Microctonus* parasitoid wasps (Hymenoptera: Braconidae) have been released in Aotearoa New Zealand as biocontrol agents, targeting three different pest weevil species. Despite their value as biocontrol agents, no genome assemblies are currently available for these *Microctonus* wasps, limiting investigations into key biological differences between the different species and strains.

Methods and findings Here we present high-quality genomes for *Microctonus hyperodae* and *Microctonus aethiopoides*, assembled with short read sequencing and Hi-C scaffolding. These assemblies have total lengths of 106.7 Mb for *M. hyperodae* and 129.2 Mb for *M. aethiopoides*, with scaffold N50 values of 9 Mb and 23 Mb respectively. With these assemblies we investigated differences in reproductive mechanisms, and association with viruses between *Microctonus* wasps. Meiosis-specific genes are conserved in asexual *Microctonus*, with *in-situ* hybridisation validating expression of one of these genes in the ovaries of asexual *Microctonus aethiopoides*. This implies asexual reproduction in these *Microctonus* wasps involves meiosis, with the potential for sexual reproduction maintained. Investigation of viral gene content revealed candidate genes that may be involved in virus-like particle production in *M. aethiopoides*, as well as a novel virus infecting *M. hyperodae*, for which a complete genome was assembled.

Conclusion and significance These are the first published genomes for *Microctonus* wasps which have been deployed as biocontrol agents, in Aotearoa New Zealand. These assemblies will be valuable resources for continued investigation and monitoring of these biocontrol systems. Understanding the biology underpinning *Microctonus* biocontrol is crucial if we are to maintain its efficacy, or in the case of *M. hyperodae* to understand what may have influenced the significant decline of biocontrol efficacy. The potential for sexual reproduction in asexual *Microctonus* is significant given that empirical modelling suggests this asexual reproduction is likely to have contributed to biocontrol decline. Furthermore the identification of a novel virus in *M. hyperodae* highlights a previously unknown aspect of this biocontrol system, which may contribute to premature mortality of the host pest. These findings have potential to be exploited in future in attempt to increase the effectiveness of *M. hyperodae* biocontrol.

[†]S.N Inwood and J Skelly contributed equally to this work.

*Correspondence:

Peter K. Dearden

peter.dearden@otago.ac.nz

Full list of author information is available at the end of the article



Keywords Genome assembly, Metagenome assembly, *Microctonus*, Endoparasitoid wasp, Biocontrol, Parthenogenesis, Virus

Background

Species in the genus of *Microctonus* (Wesmael, 1835) (Hymenoptera: Braconidae) have been found to be effective biocontrol agents against forage production weevil pests in Aotearoa New Zealand (NZ), against which they are the primary form of control (Fig. 1A). *M. hyperodae* has been found to provide effective biological control against the Argentine stem weevil (ASW) *Listronotus bonariensis* (Kuschel) (Coleoptera: Curculionidae), a severe pest of Gramineae which causes an estimated NZD\$200 million in damage p.a. [1]. There are multiple strains of *M. aethiopoulos* Loan, where strains refers to *M. aethiopoulos* which were collected from different locations, have different reproductive biology and have different primary hosts, which are also used as biocontrol agents in NZ. Different *M. aethiopoulos* strains are effective against the clover root weevil *Sitona obsoletus* (Gmelin) (Coleoptera: Curculionidae), which causes an estimated NZD\$235 in damage p.a. [1], and the lucerne weevil *Sitona discoideus* (Gyllenhal) (Coleoptera: Curculionidae) (Fig. 1A).

M. hyperodae was released for biocontrol at numerous sites across NZ's ecoclimatic range in the 1990's [2, 3], after collection from eight locations throughout their home range in temperate South America [4]. *M. hyperodae* was initially found to exert parasitism at rates of 70–90% [5–7] resulting in useful suppression of the weevil [8]. However, after several years parasitism rates dropped significantly, by up to 50% in some locations [9, 10] leading to renewed concerns about damage to pasture [1, 11]. This decline of biocontrol efficacy has been associated with the asexual reproduction of *M. hyperodae* via thelytokous parthenogenesis, resulting in the parasitoid being unable to evolve fast enough to accommodate the effects of selection pressure on the genetically diverse and sexually reproducing *L. bonariensis* populations [10, 12, 13] potentially through increased parasitoid evasion behaviours by the weevil [14, 15].

The lucerne weevil *Sitona discoideus* is controlled by a sexually reproducing strain of *M. aethiopoulos* from the Mediterranean area [16] and is often referred to as the Moroccan strain [17], released for biocontrol in 1982. This *M. aethiopoulos* strain was subsequently found to have no appreciable effect on the invasive populations of the clover root weevil *Sitona obsoletus* [18], resulting in a widespread European search for a suitable parasitoid. A sexually reproducing French strain of *M. aethiopoulos* [19] was found to be effective against

S. obsoletus, however in quarantine it was found to hybridise with the Moroccan strain producing offspring with greatly reduced efficacy against both *Sitona* spp, precluding its use for biocontrol in NZ [20]. Continued searching led to the discovery of an asexual European strain of *M. aethiopoulos* in Ireland, which reproduces asexually via thelytokous parthenogenesis which would not hybridise with the Moroccan strain [21]. This permitted its widespread rearing and release and the eventual suppression of the weevil [17, 22].

Presence of two reproductive strategies within the *M. aethiopoulos* spp. raises interesting questions as to the underlying mechanisms of asexuality in the *Microctonus* genus. While the *M. hyperodae* population in NZ was found to be asexual, four impotent male *M. hyperodae* were discovered amongst the founding 251 the adult *M. hyperodae* reared from weevils from South America [4], indicating some ability to reproduce sexually. Preliminary examination of the biological underpinnings of reproductive mode in *M. hyperodae* using allozymes revealed an absence of recombination, inferred from a lack of certain homozygous genotypes despite prevalent heterozygotes in populations [23]. From this lack of recombination, it was suggested that *M. hyperodae* parthenogenesis might be apomictic, using mitosis rather than meiosis. However, more recently the conservation and expression of core meiosis genes in *M. hyperodae* ovaries has been demonstrated, indicating that *M. hyperodae* parthenogenesis may involve meiosis thereby retaining the potential for sexual reproduction [24]. No such investigation of the parthenogenesis mechanism of the *M. aethiopoulos* Irish strain has been performed to date.

Another difference between *Microctonus* wasps is in the presence and transmission of virus particles during parasitism. Endogenous viral elements (EVEs) in the form of polydnviruses (PDVs) or virus-like particles (VLPs) as well as exogenous viruses, often play a role in the parasitism processes of koinobiont endoparasitoids, particularly in host immune-suppression [25–29]. Barratt et al. [30, 31] detected the presence of viral particles in the ovarian epithelial cells of the sexual Moroccan strain of *M. aethiopoulos*, with no viral particles found in other *Microctonus* strains or species. There has been no further investigation into the presence of EVEs or infectious viruses associated with the *Microctonus* wasps, and no investigation using genomic data.

A more comprehensive investigation into the potential presence of viruses or EVEs in *M. hyperodae* is

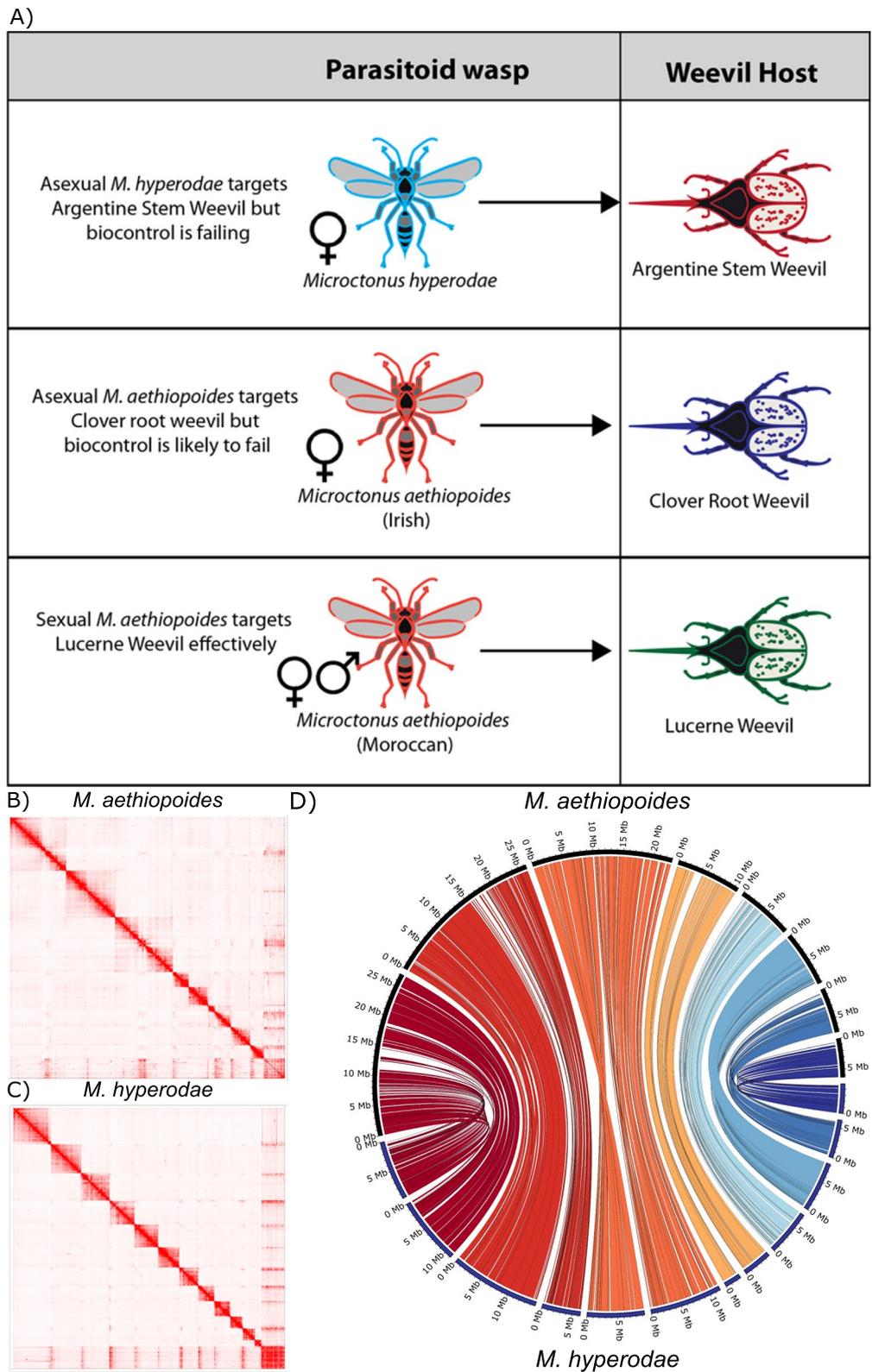


Fig. 1 Key biological differences between *Microctonus* wasps and Hi-C scaffolding results. **A** Host target and reproductive mechanisms of *Microctonus* wasps used as biocontrol agents in New Zealand, Hi-C interaction maps for **B** *M. aethioides* Irish and **C** *M. hyperodae* displaying chromosome scaffolds, and **D** a Circos plot displaying chromosomal synteny between both Hi-C assemblies

required due to a premature mortality phenomenon observed in *L. bonariensis*. This phenomenon is where ASW exposed to adult *M. hyperodae* have a significantly higher rate of mortality than ASW which are not exposed to *M. hyperodae*, with this mortality occurring too early to be a result of parasitism [32–34]. The cumulative ASW mortality increases with increasing numbers of adult *M. hyperodae*, but is less apparent with increased time of exposure [32–34]. The current hypothesis is that there may be a toxin-antitoxin system acting during parasitism, whereby *M. hyperodae* transmits something toxic during an unsuccessful ovipositional attempt, which is offset by an ovarian extract during successful parasitism, with increased *M. hyperodae* exposure times increasing the likelihood of a successful parasitism attempt to avoid this premature mortality [34]. This toxin-antitoxin phenomenon has also been observed with the parasitoid *Asobara japonica* Forster (Hymenoptera: Braconidae) and its *Drosophila* host, with the source of toxicity revealed to be virus particles [35, 36]. The variance in virus particle detection between *Microctonus* species, and the premature mortality phenomenon associated with *M. hyperodae* therefore necessitate an investigation into the viral gene content of the *Microctonus* spp. genome assemblies, and further investigation into the virome of *M. hyperodae*.

Given the range of host targets and reproductive modes, and the importance of biocontrol to a pastoral economy, the genomes of *M. hyperodae* and *M. aethioides* have the potential to determine factors that influence biocontrol efficacy, such as genomic correlates of the reproductive mode and virus associations, offering a basis for a better understanding the biology of these biocontrol systems. Here we present the chromosome-level genomes of two *Microctonus* species, *M. hyperodae* and *M. aethioides*, with additional assemblies for a further two *M. aethioides* strains with divergent biology, and that of a previously undetected virus found to be infecting *M. hyperodae*. These assemblies provide a unique insight into parasitoid reproduction and commensal viruses that play key roles in parasitic life history, and are valuable genomic resources for understanding the biology of *Microctonus* biocontrol and ongoing investigations into its success or decline in NZ, which is crucial if we are to maintain its efficacy.

Results and Discussion

Read taxonomic classification

Bacterial endosymbionts, particularly *Rickettsia*, *Wolbachia*, and *Cardinium* can induce parthenogenesis in insects [37], while maintaining the potential for sexual reproduction [38, 39]. As *M. hyperodae* and the Irish

strain of *M. aethioides* reproduce asexually, the presence of such endosymbionts was investigated using Kraken2 read classification. Kraken2 analysis resulted in the classification of 17.7%, 17.3%, 17.4% and 7.25% of reads from *M. aethioides* Irish, French, and Moroccan strains, and *M. hyperodae*. For all samples a large majority of classified reads were classified as Eukaryotic (14.53%, 14.22%, 14.22%, and 6.06% respectively) while most reads remaining unclassified (as the database does not contain insects).

Prokaryotic classification was much lower for all samples, with reads classified as bacterial from *M. aethioides* Irish, French, and Moroccan strains, and *M. hyperodae* as follows; 2.72%, 2.61%, 2.65% and 1.04%. Classification of reads as known parthenogenesis-inducing endosymbiont genera was low, with 0.05% or less as *Rickettsia*, 0.01% or less of reads classified as *Wolbachia*, and 0.00% as *Cardinium*. There was no correlation between the percentage of reads assigned to the three genera and the reproductive mechanism of the parasitoids. These results are consistent with previous RNA-seq read classification and PCR results from *M. hyperodae* [24], and with antibiotic and heat treatment failing to revert the asexual reproduction mechanism of *Microctonus* wasps [40]. This does not preclude that asexual reproduction could be associated with another bacterial genera, which would likely require more targeted 16S sequencing to investigate. The percentage of reads classified as viral were lower again at 0.23%, 0.24%, 0.25% and 0.05% from *M. aethioides* Irish, French, and Moroccan strains, and *M. hyperodae*. The reduced microbial diversity in *M. hyperodae* for both bacteria and viruses may be of interest given the decline of successful biocontrol with *M. hyperodae* specifically, though may instead be due to species in the *M. hyperodae* microbiome being too phylogenetically diverged from those present in the Kraken2 database for reads to be classified.

Wasp Genome assembly

Using short-read Illumina sequencing, draft genomes were produced for *M. hyperodae*, and the Irish, French and Moroccan strains of *M. aethioides*, containing 105–128 Mb of total sequence, with a BUSCO completeness of 86.8–93.2% (Table 1). Hi-C scaffolding of the *M. hyperodae* and Irish *M. aethioides* assemblies improved these assemblies substantially to chromosome-scale scaffolds, with N50 lengths increasing from 15 Kb to 9 Mb for *M. hyperodae* and from 64 Kb to 23 Mb for *M. aethioides* Irish (Table 1). The contiguity of these scaffolded *Microctonus* assemblies (Table 1) are comparable to the model insects *D. melanogaster*, *A. mellifera* and *N. vitripennis*, which have N50 lengths of 25.2 Mb, 13.6 Mb and 24.7 Mb, and L50 values of 3, 5 and 7 [41–43]. Compared

Table 1 Assembly and annotation statistics for Meraculous and Hi-C scaffolded assemblies for *M. hyperodae* and *M. aethiopoidea* strains. Genome and Annotation BUSCO completeness (%) scores refer to the number of BUSCO genes found complete (whether single-copy or duplicated) in the genome assemblies and respective annotations

Assembly statistics	Hi-C assemblies		Meraculous assemblies			
	<i>M. hyperodae</i>	<i>M. aethiopoidea</i> Irish	<i>M. hyperodae</i>	<i>M. aethiopoidea</i> Irish	<i>M. aethiopoidea</i> French	<i>M. aethiopoidea</i> Moroccan
Total length (Mb)	106.7	129.2	105.8	128.8	118.9	120.3
Number of scaffolds	3663	2844	14,207	6322	10,395	13,735
Scaffold N50 (bp)	9,364,176	23,025,277	15,392	64,081	30,183	17,930
L50	5	3	1758	484	881	1634
GC (%)	29.5	29.4	29.5	29.4	29.3	29.4
Genome BUSCO complete (%)	89.8	93.9	86.8	93.2	90.7	89.1
Predicted genes	11,744	12,474	13,667	13,521	14,229	14,475
Annotation BUSCO complete (%)	87.9	94.0	85.2	90.7	88.4	85.8
Repeat content (%)	12.85	21.67	12.76	22.08	17.63	17.96

to other scaffolded Hymenopteran genomes available on NCBI (<https://www.ncbi.nlm.nih.gov/data-hub/genome/?taxon=7399>, accessed 14/09/22, excluding contig assemblies) the *M. aethiopoidea* Irish Hi-C assembly has an N50 length higher than 90% of assemblies, and *M. hyperodae* higher than 75% of assemblies, indicating that these genomes are more contiguous than most scaffolded Hymenopteran genomes, with a high level of BUSCO completeness (Table 1).

The lengths for these assemblies are smaller than many other high-quality insect genomes, e.g. *D. melanogaster* at 144 Mb, *A. mellifera* at 225 Mb, and *N. vitripennis* at 297 Mb [41–43] despite their high BUSCO completeness scores (Table 1). Genome sizes vary greatly across Hymenoptera, with this variance at least partially attributed to repeat content [44]. The repeat content of the Hi-C scaffolded *M. hyperodae* and *M. aethiopoidea* Irish was 12.85% and 21.67%, while for the *M. aethiopoidea* Moroccan and French assemblies repeat content was 17.96% and 17.63% (Table 1). This indicates that these *Microctonus* genomes are relatively compact, and when compared to other larger parasitoid wasp assemblies with *Pteromalus puparum*, *Diachasma alloenum*, *Cotesia glomerata* and *Cotesia congregata* genomes, which are 338 Mb, 389 Mb, 290 Mb and 199 Mb in length, all also have higher repeat contents of 40.11%, 49%, 54.75% and 35.04% respectively [44–46]. The Hi-C data suggests variance of chromosome number between *M. hyperodae* and *M. aethiopoidea*, with 12 and eight chromosome-scale Hi-C scaffolds respectively (Fig. 1B & C). Investigation of gene synteny suggests that chromosomes 1–4 in *M. aethiopoidea* are each represented by two chromosomes in *M. hyperodae* (Fig. 1D), which may represent fusion of chromosomes in *M. aethiopoidea*, or fragmentation

of chromosomes in *M. hyperodae* in the time since they have diverged.

Wasp Genome annotation

Gene prediction of *Microctonus* genomes was performed with Funannotate, resulting in 12,982 to 14,475 gene predictions for each assembly, with a BUSCO completeness of 85.2 to 94.1% (Table 1). The high gene prediction BUSCO completeness is comparable to other parasitoid wasps [42, 47] implying that these are high-quality genomes and gene predictions. Hi-C scaffolding improved assembly contiguity, however had minimal impact on the number of predicted genes or on BUSCO gene prediction results, as was expected. The small reduction in predicted genes and increase in BUSCO completeness scores for Hi-C assemblies are likely a result of separate contigs with incomplete predicted genes on either end being scaffolded together to form complete gene annotations, as despite the decrease in predicted gene numbers, BUSCO completeness was higher for scaffolded assemblies (Table 1). These results indicate that the short-read only assemblies are still contiguous enough for high-quality gene prediction. Only the gene prediction performed for *M. hyperodae* used RNA-seq data (two ovarian samples), and while this may have resulted in improved gene prediction for the *M. hyperodae* assembly, particularly for genes with high expression in the ovaries, the *M. hyperodae* gene predictions, for both the unscaffolded and scaffolded assemblies, had lower BUSCO completeness scores than those for the *M. aethiopoidea* assemblies.

Using these high-quality annotations, the divergence between *M. hyperodae* and *M. aethiopoidea* strains was estimated, indicating that the French and Moroccan

strains (both sexually reproducing) are more closely related than either is to the asexual Irish strain, with the Irish strain estimated to have diverged from the sexually reproducing strains 2 million years ago (MYA), and the two sexually reproducing strains diverging 1 MYA (Supplementary Fig. 1). The divergence between *M. hyperodae* and *M. aethiopoides* is estimated to be 17 MYA (Supplementary Fig. 1). This is a larger divergence estimate than was suggested by previous morphological and genetic analysis [48].

Meiosis gene inventory in sexual and asexual strains

A ‘meiosis detection toolkit’ has been described for investigation of asexual reproduction mechanisms in Hymenoptera, cataloguing the presence or absence of genes with known roles in mitosis or meiosis [49]. In an organism that does not use meiosis for gamete production there is no evolutionary constraint on core meiosis genes, so it is expected that their sequence and function would not be conserved [49, 50]. Loss of these core meiosis genes would lead to obligate parthenogenesis, where reversion to sexual reproduction is not possible. This toolkit was used to clarify whether asexual reproduction

in *M. hyperodae* and Irish *M. aethiopoides* is automictic, involving meiosis, or apomictic, relying solely on mitosis for egg production which cannot be reverted to sexual reproduction.

This analysis found 36 of 40 meiosis toolkit genes present in single copies in all *Microctonus aethiopoides* assemblies, with the same 36 found in *M. hyperodae* with the addition of RECQ2 (Fig. 2). The presence/absence of these genes in these *Microctonus* species is segregated by phylogenetic relatedness, rather than their reproductive mechanism (Fig. 2). All but one of the core meiosis genes were detected in *Microctonus* genomes, with only DMC1 being absent (Fig. 2), though DMC1 is also absent in *Drosophila melanogaster* and the majority of Hymenopteran species assayed regardless of their reproductive mechanism, implying it is dispensable in meiosis in these species [51].

There is variation in the presence and absence of RECQ genes in the *Microctonus*. RECQ1 and RECQ4 were not identified in the peptide databases of any *Microctonus*, and RECQ2 was only detected in *M. hyperodae* (Fig. 2), though this could be due to gene prediction for *M. hyperodae* using RNA-seq data from ovaries while

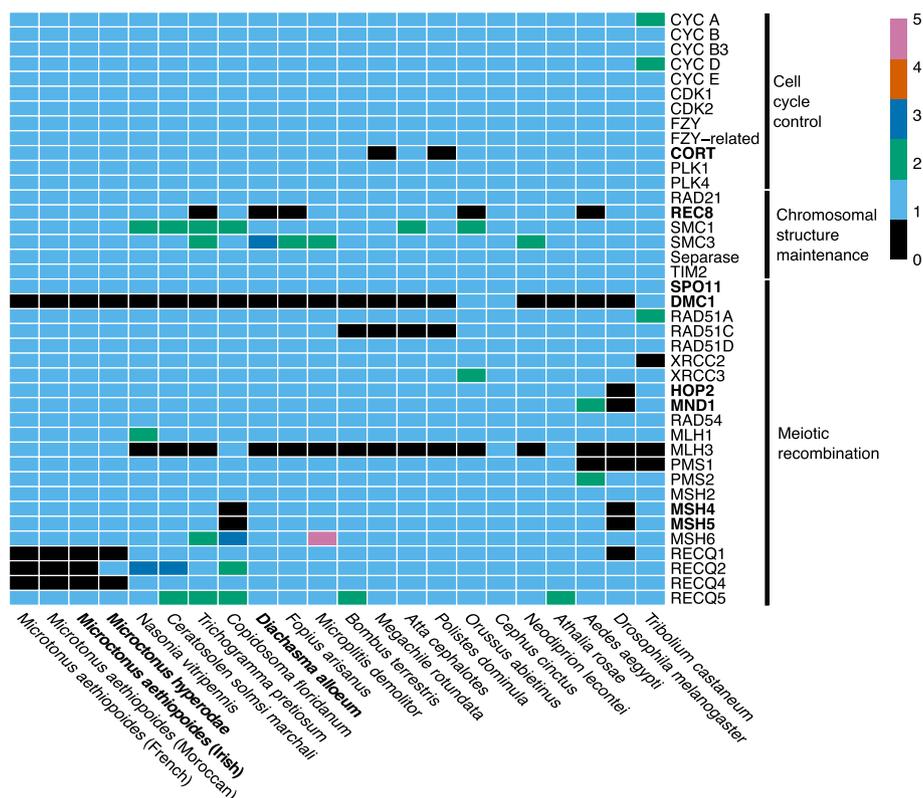


Fig. 2 A heatmap displaying the meiosis gene inventory of *Microctonus* wasps, compared to results in other Hymenoptera, Diptera and Coleoptera from Tvedte et al. [51]. Colours in the heatmap indicate the absence or presence (and gene copy number) of each meiosis gene. Core meiosis genes, which are specific to meiosis, and asexual Hymenoptera species are indicated in bold

M. aethiopoides prediction did not. The RECQ group of proteins are DNA helicases that unwind double-stranded DNA for DNA repair, recombination, and transcription. Some of the RECQ orthologs are required to protect the genome against deleterious mutations [52]. RECQ1 interacts with parts of the DNA mismatch repair pathway during recombination in humans [53]. RECQ4 co-localises with Rad51 after the induction of DNA double-strand breaks, with a possible role in DNA double-strand breakthrough homologous recombination [54]. RECQ2 mutants in *S. cerevisiae* have suppressed non-crossover recombinants indicating a role in mediating recombination product formation [54]. The absence of RECQ genes in *Microctonus* genomes may influence rates of recombination leading to lower genetic diversity, particularly in those that reproduce asexually, and may explain the lack of recombination previously observed in *Microctonus hyperodae* at two allozyme loci [23].

Having surveyed the meiosis toolkit genes and examined reads for the presence of endosymbionts that manipulate reproductive mechanisms, we have found no clear evidence for a cause of asexual reproduction in *M. hyperodae* or *M. aethiopoides* Irish. The conservation of most meiosis toolkit genes, particularly all core meiosis genes aside from DMC1 (Fig. 2), indicates that asexual reproduction in *M. aethiopoides* Irish and *M. hyperodae* likely uses automixis involving meiosis. This is supported by the detection of MND1 expression in the asexual Irish *M. aethiopoides* ovaries (Supplementary Fig. 2), and is consistent with the detection of meiosis core gene expression in *M. hyperodae* ovaries [24], implying that these parasitoids have retained the potential to reproduce sexually. To investigate whether there has been any sexual reproduction occurring since their release, comparisons of mitochondrial and nuclear gene phylogenies from population sequencing could be performed, as if sexual reproduction has occurred these phylogenies should be incongruent [55].

Females of these two asexual *Microctonus* are diploid, despite developing from unfertilised eggs. The formation of unfertilised yet still diploid gametes via automixis can involve various strategies to restore this diploidy, which have varied effects on rates of heterozygosity, from elimination to retention of all heterozygosity [56]. Heterozygosity rates in these *Microctonus* do not segregate based on their reproductive mechanisms, with Genomescope Kmer estimates of 0.59%, 0.25%, 0.34% and 0.41% for *M. hyperodae* and the Irish, French, and Moroccan *M. aethiopoides*. These results are supported by previous detection of heterozygosity in *M. hyperodae* (Iline & Phillips, 2004, and with heterozygosity also detected in the asexual Irish *M. aethiopoides*, this indicates that if parthenogenesis is automictic in these parasitoids it

relies on a strategy that retains at least some heterozygosity. One such mechanism for this is premeiotic duplication which retains all heterozygosity, with observed upregulation of genes involved in endoreduplication in *M. hyperodae* ovaries providing a putative mechanism for this [24]. Investigation of where heterozygous sites occur along chromosomes in the Hi-C scaffolded *Microctonus* genome assemblies could assist in determining this mechanism, as was used in investigation of parthenogenesis in the clonal raider ant *Cerapachys biroi* [57].

Identification of an infectious virus involved in biocontrol

A reciprocal BlastP search of predicted proteins from the *Microctonus* genomes was performed to investigate viral gene content in *Microctonus* genomes, which identified significant viral hits in all assemblies (Supplementary table 1). The number of RNA virus hits was low as expected, given that DNA sequencing should not detect RNA viruses unless their genes are integrated into the genome. A variable number of DNA virus hits were detected in *M. hyperodae* (across 20 contigs) as well in the Moroccan and French strains of *M. aethiopoides* (across five and three contigs) (Supplementary Fig. 3). All hits came from viral families known to infect insects, with most *M. aethiopoides* hits to *Polydnaviridae* and most *M. hyperodae* hits to *Baculoviridae* and unclassified DNA viruses (Supplementary Fig. 4).

To investigate whether viral genes detected in these assemblies are endogenous (and putatively involved in PDV particle/VLP production) or exogenous, several approaches were taken. First the presence of eukaryotic genes on the same contigs as viral genes was investigated using BlastP and BUSCO. This revealed no strong evidence of eukaryotic genes on viral contigs in *M. hyperodae*, in which no virus-like particles had previously been detected [30], and recent transcriptome analysis revealed no clear signs of PDVs or other endogenous viral elements with high expression in venom or ovaries [24]. *M. hyperodae* was the only parasitoid with a Hi-C genome and DNA virus hit, with none of these virus hits present on Hi-C scaffolds. This is despite Hi-C data having coverage for those viral contigs, with 0.15% of Hi-C interactions involving viral contigs, and 83.8% of these viral interactions being between two viral contigs, which indicates that their exclusion from the main Hi-C scaffolds was not due to these contigs having no sequence data.

Presence of eukaryotic genes on contigs with DNA virus hits was also examined in French and Moroccan *M. aethiopoides*, with no DNA virus hits for the Irish strain. For French *M. aethiopoides* 10 out of 16 genes on these contigs had eukaryotic hits, with all of these belonging to the Braconidae. Similarly, for Moroccan *M. aethiopoides* 11 of 21 genes had eukaryotic hits,

seven of which were to Braconidae, providing some evidence that these viral contigs may be endogenous. No BUSCO genes were present on the contigs with DNA virus hits for the Moroccan or French assemblies, which would have provided stronger evidence of whether DNA virus genes are endogenous or otherwise. By contrast, BUSCO genes were present on all contigs with RNA virus hits, indicating they are endogenous as expected. VLPs had previously been detected in the Moroccan *M. aethioides*, but not in the limited number of French or Irish samples collected [30, 31].

The GC sequence content and sequencing depth of viral contigs was compared to Hi-C chromosome scaffolds for *M. hyperodae* and to contigs with BUSCO genes for *M. aethioides* strains to provide further evidence of whether viral contigs were exogenous or endogenous (Supplementary Fig. 4). There is a significant difference between both the GC content and sequencing depth of viral contigs and Hi-C scaffolds in *M. hyperodae* (GC: mean 34.0% compared to 29.5%, $p=6.7E-08$, depth: mean 258 compared to 175, $p=1.6E-03$), with this depth result still significant when two high outlier contigs (scaffolds 90 and 995, mean depth of 879) are removed (mean 189 compared to 175, $p=1.3E-02$) (Supplementary Fig. 4). There was also a significant difference in GC content and depth between BUSCO and viral contigs in *M. aethioides* French (GC: mean 35.1% compared to 28.9%, $p=1.9E-02$, depth: mean 172 compared to 281, $p=7.4E-03$) while in *M. aethioides* Moroccan there was no significant difference in either metric (GC: mean 32.3% compared to 29.0%, $p=0.12$, depth: mean 548 compared to 194, $p=0.88$) even when the high depth outlying contig in *M. aethioides* Moroccan (depth=1822), which contains the only *Baculoviridae*, unclassified DNA virus and hemoflagellate parasite hits, was analysed in a separate group ($p \geq 0.50$) (Supplementary Fig. 4).

Only *M. aethioides* Moroccan returned non-significant results for both comparisons (Supplementary Fig. 4) providing more evidence supporting that these contigs may be endogenous, with this strain previously demonstrated to have VLPs in their ovaries [30, 31]. No analysis has been done on what these particles contain beyond an inability to extract DNA from them, so it is unknown whether they have a viral origin, though these identified viral genes in *M. aethioides* Moroccan are candidates for involvement in VLP production. A more contiguous genome assembly would be required to confirm whether they are exogenous or endogenous. The Hi-C data, GC and sequencing depth analyses do however provide strong evidence that the viral contigs in *M. hyperodae* are derived from an exogenous viral infection.

M. hyperodae metagenome assembly

Given these results, efforts were therefore made to assemble a complete genome for the exogenous virus in *M. hyperodae*. Meta-genomic assembly was performed using MinION long reads, which were generated from a sample that preliminary Illumina sequencing had confirmed infection status with high viral coverage (mean viral depth of 19x, compared to mean Hi-C scaffold depth of 9x). This generated 354,749 MinION reads with a mean 65×depth of the viral contigs and 19×depth of the Hi-C scaffolds. All reads that mapped to the *M. hyperodae* Hi-C scaffolds were removed, retaining 44,388 reads with a mean length of 7305 bp.

Metagenomic assembly from these filtered reads generated 933 contigs, 41 of which were circular assemblies. viralFlye classified one circular contig, six linear contigs and three component contigs with branching paths, as complete viral assemblies. Given the viral BlastP hits to *Baculoviridae* and an unclassified DNA virus, *Leptopilina boulardi filamentous virus* (LbFV), both of which have a circular dsDNA genome, it was expected that the MhFV genome would also be circular. A BlastN search of the complete circular viral contig found significant hits to 18 of the 20 viral contigs (lacking hits to the two high depth contigs excluded from analysis earlier) identified in the *M. hyperodae* genome assembly, indicating that this complete assembly is for the exogenous viral infection identified. Based on the number of BlastP hits to LbFV and later phylogenetic analysis, we provisionally name this virus *Microctonus hyperodae* filamentous virus (MhFV).

Complete genome assembly of *M. hyperodae* filamentous virus

Illumina read coverage for this complete MhFV assembly was variable (Fig. 3), with a maximum of 494x, and a mean of 147x. Illumina short read coverage dropped to 0× in 44 regions, ranging in length from 1 to 370 bp, with a mean of 88 bp and total length of 3860 bp. Nanopore read coverage was less variable (with a maximum of 168× and a mean of 69x) and didn't drop below 4x (Fig. 3). The MhFV genome assembly is 163 Kb long, with a GC content of 37.8% and has comparable characteristics compared to other nuclear arthropod-specific large dsDNA viruses (NALDVs) (Supplementary Table 2).

There are 158 predicted genes in the MhFV genome, all of which are complete (minimum amino acid length=34, maximum=1652, mean=286.7). Of the 158 predicted genes, there were significant BlastP hits and/or Pfam protein domains for 86, leaving 72 genes without any significant homology to indicate potential function (Supplementary Table 3). The most common virus hits were to *Leptopilina boulardi* filamentous virus (LbFV) with 23 hits, followed by *Drosophila*-associated filamentous virus

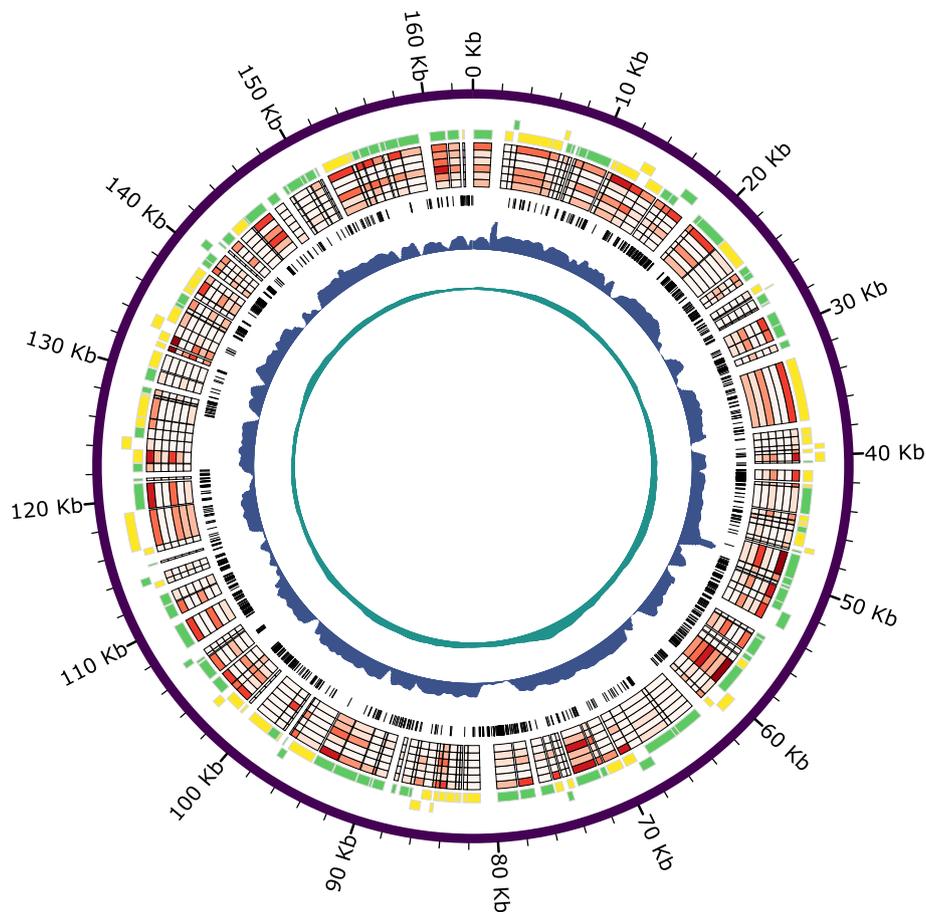


Fig. 3 A Circos plot of the MhFV genome assembly. Predicted genes are indicated in the yellow and green blocks, on the positive and negative strand respectively. Normalised gene expression is displayed in the heatmap, where higher expression is indicated by darker red, and with the six rings representing tissues from the outside inwards as follows; pupa, head, thorax, abdomen, ovaries and venom gland. Dashed black lines indicate SNP variants detected throughout the MhFV genome. Relative read depth for Illumina and ONT MinION sequencing is indicated in blue and teal respectively, calculated with a sliding window with a size of 500 bp and slide of 100 bp

(DaFV, eight hits) and *Glossina pallidipes* salivary gland hypertrophy virus (five hits) (Supplementary Table 3). The predicted genes contain several known viral gene families, such as *per os* infectivity factors, late expression factors, five KIL-N domain-containing genes, and 17 BRO genes (Supplementary Table 3).

While there were 27 hits to bacteria and eukaryotes, an investigation of the full list of Blast hits revealed 11 of these genes also had other viral hits, suggesting potential viral contamination in the assemblies that the best hits came from, rather than there being a putative non-viral origin for all 27 genes. Three genes had only non-viral BlastP hits, including two inhibitor of apoptosis (IAP) genes and a lytic polysaccharide monoxygenase (Supplementary Table 3). ORF32 and ORF152 are both annotated as IAPs (Supplementary Table 3), which are often found in NALDVs and thought to manipulate the host immune system by suppressing

apoptosis of infected cells [58, 59]. IAP genes were also present in the LbFV genome and had a eukaryotic origin [60]. ORF116 and ORF133 are both annotated as lytic polysaccharide monoxygenases (Supplementary Table 3), with at least 97% of hits for both genes from bacterial species. While none of these hits were viral, many viruses in the families Poxviridae and Baculoviridae contain genes named Fusolin and GP37, which belong to the lytic polysaccharide monoxygenase family [61]. These act to disrupt the peritrophic matrix of insect hosts and facilitate viral infection [62, 63]. ORF67 was annotated as containing a JmJc-domain (Supplementary Table 3), and while it did have viral hits with one from DaFV and two from different LbFV JmJc-domain containing proteins, there were no other viral hits and 497 non-viral hits. Phylogenetic analysis of the JmJc-domain genes in LbFV indicated they were likely acquired via horizontal transfer from an ancestral

host parasitoid [60]. JmJc-containing proteins are a class of demethylase enzymes involved in transcription regulation [64].

Phylogenetic position of MhFV

Recent work identifying endogenous viral elements (EVEs) in parasitoid genomes has identified 12 core genes (*DNApol*, *helicase*, *lef-5*, *lef-8*, *lef-9*, *p33*, *pif-0*, *pif-1*, *pif-2*, *pif-3*, *pif-5* and *ac81*) conserved in many NALDVs, and EVEs derived from NALDVs [65]. Of these core genes, ten were identified in the LbFV genome, missing both *pif-3* and *lef-5*. A BlastP search of gene predictions from the incomplete assembly of DaFV only identified four of the core genes. The MhFV genome has predicted genes with significant BlastP and/or protein domain hits to 11 of these core genes. MhFV is missing only *lef-5*, which is the core gene found in the least number of NALDVs due to its short length [65].

Phylogenetic analysis was carried out using protein sequences of the 12 core genes, from all NALDVs used in the analysis by Burke et al., [65]. EVEs derived from NALDVs were excluded, as were the LbFV-like viral

contigs identified in that analysis that may have an endogenous origin, as MhFV is an active viral infection rather than an EVE. The resulting phylogeny is globally well resolved, and all characterized viral families were monophyletic displaying expected relationships (Fig. 4) [65]. Despite sharing similar names, *Apis mellifera* filamentous virus and the LbFV-like viruses are separated on the phylogeny, which is the expected result. The placement of LbFV outside of characterized NALDV families is the same result found in other analyses [60, 65, 66], with the closest family in all previous analyses being Hytrosaviridae with which a monophyletic group is formed (Fig. 4). This analysis places *M. hyperodae* filamentous virus in the same clade as LbFV and DaFV (Fig. 4), supporting that MhFV is likely a member of this uncharacterized LbFV-like virus family, though electron microscopy would be required to confirm whether MhFV also has filamentous viral particles.

Manual inspection of mapped Illumina reads revealed the presence of many single nucleotide polymorphisms (SNPs) in the MhFV genome. This variation is likely a result of Illumina sequencing having been performed on

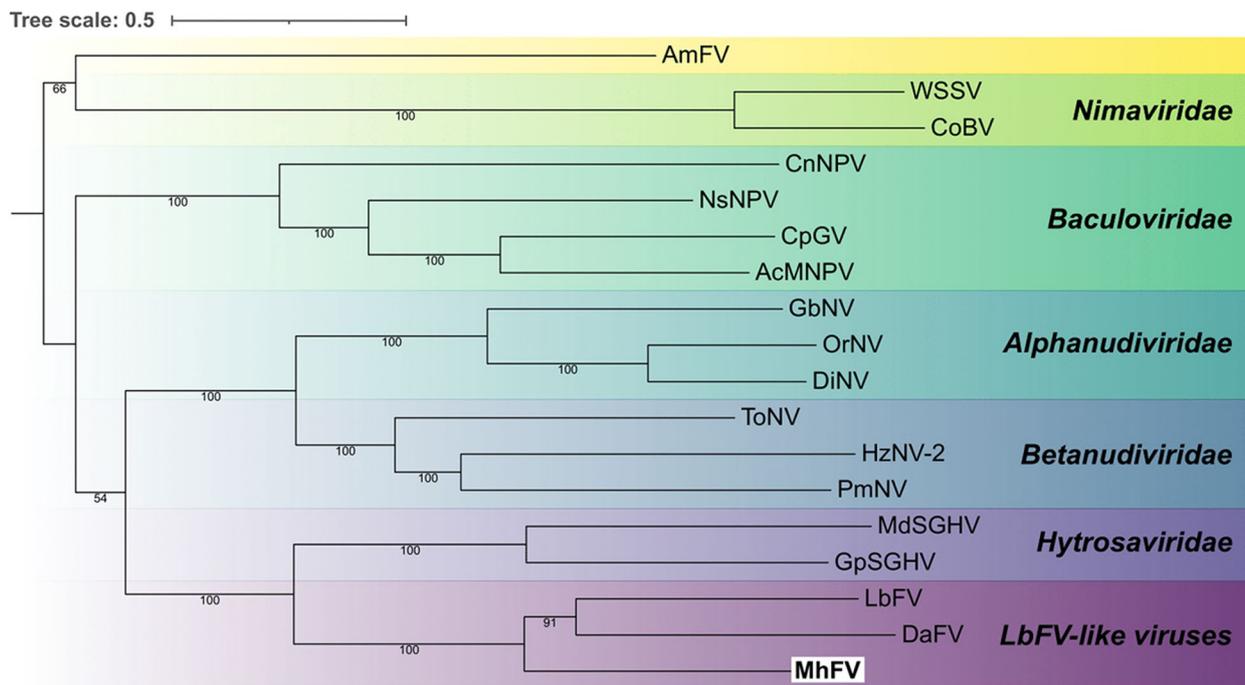


Fig. 4 Phylogenetic analysis of nuclear arthropod-specific large double-stranded DNA viruses (NALDVs). Relationships were derived using a maximum likelihood analysis with RAxML-NG, from 12 core NALDV genes, as defined by Burke et al., [65], with a total of 6818 characters from concatenated amino acid sequences. Bootstrap branch support values over 50% are indicated on relevant branches. Species names are abbreviated as follows; *Apis mellifera* filamentous virus (*AmFV*), White spot syndrome virus (*WSSV*), *Chionoecetes opilio* bacilliform virus (*CoBV*), *Culex nigripalpus* nucleopolyhedrovirus (*CnNPV*), *Neodiprion sertifer* nucleopolyhedrovirus (*NsNPV*), *Cydia pomonella* granulovirus (*CpGV*), *Autographa californica* multiple nucleopolyhedrovirus (*AcMNPV*), *Gryllus bimaculatus* nudivirus (*GbNV*), *Oryctes rhinoceros* nudivirus (*OrNV*), *Drosophila innubila* nudivirus (*DiNV*), *Tipula oleracea* nudivirus (*ToNV*), *Helicoverpa zea* nudivirus 2 (*HzNV-2*), *Panaeus monodon* nudivirus (*PmNV*), *Musca domestica* salivary gland hypertrophy virus (*MdSGHV*), *Glossina pallidipes* salivary gland hypertrophy virus (*GpSGHV*), *Leptopilina boulardi* filamentous virus (*LbFV*), *Drosophila-associated* filamentous virus (*DaFV*), and *Microctonus hyperodae* filamentous virus (*MhFV*)

a sample with five pooled *M. hyperodae* individuals, as the variability in the MinION reads derived from a single individual is much lower. Variant calling and filtering revealed 1674 SNPs in the MhFV genome (Fig. 3), compared to only two in the 111 kb LbFV genome assembly, which was also assembled from a pooled sample [60]. This indicates a relatively high level of variation in the MhFV genome that future population genetics research on *M. hyperodae* should also consider.

MhFV expression in *M. hyperodae* tissues

A BlastN search of the MhFV gene predictions against a previous de novo transcriptome assembly for *M. hyperodae* [24] revealed 79 transcripts with significant hits to 89 MhFV predicted genes. When this RNA-seq dataset is re-examined with the *M. hyperodae* and MhFV genomes, mean TPM values for each predicted MhFV gene in adult tissue samples range from 1.23 (ORF80) to 0.00, while mean TPM in pupa samples ranges from 16.48 (ORF125) to 0.00. The highest mean TPM value for each adult tissue is 3.12 in the abdomen (ORF156), 1.44 in the ovary (ORF88), 1.09 in the thorax (ORF80), 0.64 in the venom gland (ORF125) and 0.31 in the head (ORF156). This data provides transcriptomic support for MhFV infection in *M. hyperodae*, revealing levels of expression are relatively low in adult tissue samples (Fig. 3), though this may have been impacted by the use of poly(A) enrichment during library preparation. Differential gene expression analysis revealed 12 MhFV genes with expression patterns significantly influenced by tissue type. These genes had the highest expression in abdomen samples (Supplementary Fig. 5, Supplementary Table 4), which suggests future efforts to image MhFV particles should focus on tissues in the abdomen.

Implications of MhFV Infection

MhFV has 23 genes with significant BlastP hits to LbFV, and phylogenetic analysis indicates it is in the same uncharacterized viral family. LbFV infects *Leptopilina boulardi* parasitoids with prevalence varying between 0 to 95% depending on parasitoid density [67], and manipulates *L. boulardi* oviposition behaviour, significantly increasing superparasitism of their *Drosophila* host to facilitate horizontal transmission of the virus [68, 69]. This superparasitism causes decreased overall parasitism rates due to egg wastage, and infected *L. boulardi* are outcompeted by the related parasitoid *L. heterotoma* [70]. LbFV infection also causes an increase in developmental time and egg load (thought to offset egg wastage from superparasitism), and a decrease in female tibia length. *L. boulardi* eggs infected with LbFV are also encapsulated significantly less by *Drosophila*, and so this infection may also assist in the avoidance of the host immune system

during parasitism [29]. This host manipulation could provide another mechanism by which *M. hyperodae* avoids encapsulation of their eggs by *L. bonariensis* [10], alongside identified venom components [24], and therefore the potential for MhFV transmission during parasitism should be investigated in future. The possibility for behavioural manipulation of *M. hyperodae* by MhFV should also be considered, but not assumed.

Future research into this biocontrol system should consider the significant impact that the MhFV infection may play in *M. hyperodae* and biocontrol of ASW. MhFV may play a role in the premature mortality phenomenon observed in ASW exposed to *M. hyperodae* [3, 19, 34], a hypothesis supported by the cause of toxicity during interrupted parasitism attempts with the parasitoid *A. tabida* being identified as viral particles [35]. Future work should therefore investigate whether MhFV is transmitted to ASW during successful parasitism and failed parasitism attempts.

Determining the prevalence of MhFV in *M. hyperodae* populations around NZ is also of interest, to see if it is varied between locations with different declines in parasitism rates. If populations without MhFV were identified, these could then be used to determine the physiological effects of MhFV infection on *M. hyperodae*. Examining historical *M. hyperodae* samples, reared and stored from lines imported into NZ from South America before release for biocontrol, should provide evidence of whether MhFV was brought into NZ from their home range, as well as allowing for investigation of whether the prevalence and/or viral load differs in the historical and contemporary samples. Finally, given the large number of variants detected in the MhFV genome with pooled sequencing of five individuals, genetic variation in the virus genome should be considered alongside that in the parasitoid genome, both in comparisons between contemporary locations with different rates of biocontrol decline, and compared to historical samples.

Conclusions

We generated high quality genomes for the parasitoid wasps *M. hyperodae* and *M. aethiopoidea*, which are used as biocontrol agents in NZ, and have investigated aspects of their divergent biology. We have shown that core meiosis genes are conserved in both sexual and asexual *Microctonus* genomes, consistent with previous work implying that asexual reproduction is automictic, involving meiosis, with the potential for sexual reproduction retained. By investigating viral gene content in the genomes, we have also identified candidate genes in the *M. aethiopoidea* Moroccan genome that could be involved in VLP production, as well as a novel virus infecting *M. hyperodae*, for which a complete genome

was assembled. These resources will be invaluable for future work investigating genomic factors that influence success or failure of *Microctonus*-based biocontrol in New Zealand, and for ongoing investigation of the *M. hyperodae* biocontrol decline.

Methods

Microctonus spp. samples & sequencing

Microctonus spp. samples were supplied by AgResearch, Lincoln, consisting of pools of five *M. hyperodae*, five Moroccan and 10 French *M. aethiopoidea*s frozen in ethanol as well as 23 live Irish *M. aethiopoidea*s. DNA was extracted from these samples using a DNeasy kit (Qiagen), and prepared for sequencing using the TruSeq DNA PCR-Free platform. Samples were sequenced on an Illumina HiSeq 2500 by the Otago University Genomics Facility (<https://www.otago.ac.nz/genomics/index.html>), generating 26, 28, 24, and 35 Gb of 250 bp paired-end reads respectively.

Genome assembly and scaffolding

The BBTools v37.57 suite BBDuk program (<http://jgi.doe.gov/data-and-tools/bb-tools/>) was used to trim and decontaminate reads, using default settings for adapter and quality trimming, while also removing the last 5 bp from each read. Kraken2 v2.0.7 was used to taxonomically classify reads against the Kraken standard database (downloaded 17th September 2018) [71]. Kmer counting was performed using KMC v3.1.1 [72] with a kmer length of 21, for Genomescope2 analysis on the web-based interface [73].

Meraculous v2.2.5 was used for genome assembly [74], trialling a range of Kmer values, with both trimmed reads and reads normalised by BBNorm to a maximum coverage of 50 and minimum of 5. A range of parameters (including N50 and L50, as determined by the BBTools v38.0 stats wrapper), comparison to estimated genome sizes, and BUSCO v5.4.3 analysis [75] were used to determine the best parameters for each genome. Final Kmer values for assembly were as follows: 79 for *M. aethiopoidea*s French and Irish, 71 for *M. aethiopoidea*s Moroccan, and 41 for *M. hyperodae*. Normalised reads from BBNorm were used for the French and Moroccan assemblies, and non-normalised for French and *M. hyperodae*. Assembly was run with Meraculous in haploid mode for Irish, and diploid mode for French, Moroccan and *M. hyperodae* assemblies.

Draft genomes for *M. aethiopoidea*s Irish and *M. hyperodae* were then scaffolded by Phase Genomics using Hi-C data generated from pools of 10 individuals. Phase Genomics Proximo Hi-C 2.0 kit, a commercially available version of the Hi-C protocol [76] was used to generate chromatin conformation capture data. These data

were then used with the Phase Genomics Proximo Hi-C genome scaffolding platform to create chromosome scale scaffolds, as described in Bickhart et al. [77]. BUSCO v5.4.3 and the BBTools v38.0 stats wrapper were then run on Hi-C scaffolded assemblies, with exogenous viral contigs (identification detailed below) removed from the *M. hyperodae* assembly. To investigate genome synteny between the Hi-C assemblies for *M. hyperodae* and *M. aethiopoidea*s on a gene level, MCScanX was used to identify the 2000 best linkage bundles [78], which were then visualised using Circos v0.69–9 [79].

Genome annotations

Gene prediction was performed using Funannotate v1.5.0-12dd8c7 [80]. Repeats were identified and masked using RepeatModeler v1.0.11 (github.com/Dfam-consortium/RepeatModeler) and RepeatMasker v1.5.0 ([repeatmasker.org/RMDownload.html](https://www.repeatmasker.org/RMDownload.html)) via the Funannotate pipeline. The Funannotate model was then trained using Funannotate train on the repeat masked genome assembly using *Microctonus hyperodae* or *Microctonus aethiopoidea*s sequences depending on the species of interest. Augustus v 3.3.1 was iteratively trained [81] and used as input for Funannotate predict to predict genes, using the Hymenoptera BUSCO and optimized Augustus settings. Funannotate update was subsequently used to upgrade the annotation. Further annotation was performed with Funannotate using InterProScan 5.32–71.0 [82]. To provide additional data for *M. hyperodae* gene prediction, RNA was extracted from two pools of ovaries using a hybrid of Trizol (Ambion) and RNeasy mini kit (Qiagen) methods. RNA was prepared for using the TruSeq Stranded mRNA platform, and sequenced on an Illumina HiSeq 2500 by Otago Genomics Facility (<https://www.otago.ac.nz/genomics/index.html>) to generate 250 bp paired-end reads. RNA-seq data was not available for *M. aethiopoidea*s gene prediction.

To estimate the divergence between *Microctonus* species, Orthofinder v2.3.12 [83, 84] was run in multiple sequence alignment mode, using the diamond search engine and maft, on the predicted proteins from 25 Hymenoptera species and *Drosophila melanogaster* genomes (accessions in Supplementary Table 5). These peptide databases were used to produce an ultrametric species tree from. The previously determined divergence of *Acromyrmex echinator* and *Camponotus floridanus* is 62 MYA [85], which with the sum of branch lengths between these two species of 0.1101, gives a scaling factor of 563.12. This factor was used to scale the ultrametric Orthofinder species tree in Figtree, resulting in a phylogenetic tree with inferred divergence estimates, allowing for the divergence between *M. hyperodae* and *M. aethiopoidea*s to be estimated.

Identification of mitosis and meiosis-related genes

To investigate the parthenogenesis mechanism of asexual *Microctonus*, 40 genes known to play a role in meiosis and mitosis (including 8 genes with roles specific to meiosis) were collected for 15 Hymenopteran species, as well as *Aedes aegypti*, *D. melanogaster* and *Tribolium castaneum*, with the same accessions as were used in a previous analysis by Tvedte et al. [51]. Fasta files were made for the genes and ortholog groups the cyclins (CYC A, CYC B, CYC B3, CYC D and CYC E), the cyclin-dependent kinases (CDK1 and CDK2), CDC20 homologs (CORT, CDC20/FZY and CDC20-like/FZY-related), the Polo-like kinases (PLK1 and PLK4), RAD21 and REC8, structural maintenance of the chromosomes orthologs (SMC1 and SMC3), RAD51 orthologs (RAD51A, RAD51C, RAD51D, XRCC2 and XRCC3), heterodimers HOP2 and MND1, MutL orthologs (MLH1, MLH3, PMS1 and PMS2), MutS homologs (MSH2, MSH4, MSH5 and MSH6), RECQ helicase orthologs (RECQ1, RECQ2, RECQ4 and RECQ5), RAD54, SPO11, TIM2, DMC1, and Separase.

Sequences for each gene and ortholog group were aligned using Muscle v3.8.13 and trimmed using TrimAl v1.2 with the strictplus trimming parameters. Gene-specific hidden Markov models (HMMs) were generated using hmmbuild from Hmmer v3.2.1 [86]. These HMMs were used to search the *Microctonus* peptide databases for these genes of interest with hmmsearch, using an E-value of 1E-15. Top hits were selected and peptides retrieved using esl-sfetch from Hmmer v3.2.1. Peptide sequences for each gene were aligned using Muscle v3.8.13 and trimmed using TrimAl v1.2 with the “strict-plus” parameters. Phylogenetic trees were constructed for individual genes and ortholog groups using rapidnj v2.3.2 [87] with a bootstrap value of 1000 for genes and 1,000,000 for ortholog groups.

Hybridisation chain reaction in *Microctonus aethiopoides* Irish

Parthenogenetic *Microctonus aethiopoides* (Irish) were reared from CRW collected from pasture in Timaru, New Zealand, with ovaries dissected from 10 wasps three days after eclosion. Ovaries were fixed, and hybridisation chain reaction (HCR) performed with an MND1 probe (with fluorophore 488) as per Inwood et al., [24]. Ovaries were stained with DAPI at a 1:1000 concentration, and mounted on slides in glycerol, for imaging on an Olympus BX61 Fluoview FV100 confocal microscope with FV10-ASW 3.0 imaging software.

Identification of viral genes in *Microctonus* genome assemblies

In order to identify genes with a viral origin in the *Microctonus* genome assemblies, the predicted peptide

databases were searched for peptides with significant homology to viral genes using a reciprocal BlastP search approach. First a BlastP v2.12.0 search was performed (with an E value of 1E-05 to minimise false-positive results) of all *Microctonus* predicted genes against the non-redundant (nr) database downloaded on May 16th, 2021. This initial BlastP search was restricted to only viral entries in the database by using TaxonKit 0.8.0 [88] to produce a list of all viral taxonomy identifiers at a species level or below and restricting the BlastP search to this list using the -taxidlist option. Any gene with a significant viral BlastP result was then used in a subsequent BlastP v2.12.0 search against the whole nr database to remove those genes with better non-viral hits. Finally, any remaining genes on contigs with viral hits that did not have a viral hit were also subject to a BlastP search against the whole nr database to identify potentially eukaryotic genes on the viral contigs. The scripts used for this viral gene identification are available on Zenodo at <https://doi.org/10.5281/zenodo.7939016> using a Snake-make-managed workflow [89].

GC content and sequencing depth were compared between Hi-C scaffolds and contigs with DNA virus hits for *M. hyperodae*, and between contigs containing BUSCO hits and contigs with DNA virus hits for *M. aethiopoides* French and Moroccan. The GC content for all contigs was calculated using BBStats v38.0. To determine read depth of contigs, reads from genome assembly were trimmed using BBDuk v38.0 using default settings and trimq=15, and then mapped to the appropriate genome assembly using BWA v0.7.15 [90]. SAMtools coverage v1.10–98 [91] was then used to determine read mean sequencing depth of contigs. To test for statistical differences between viral and Hi-C/BUSCO contigs, a Shapiro–Wilk test was first carried out to determine whether data were normally distributed, which indicated non-parametric tests should be used. A Kruskal–Wallis rank sum test was used to identify whether there was a significant difference between groups, followed by a pairwise Wilcoxon rank sum test. All statistical tests used an alpha threshold of 0.05 as indication of a significant difference. The scripts used for these sequence characteristic comparisons are available on Zenodo at <https://doi.org/10.5281/zenodo.7939021> using a Snakemake-managed workflow.

Viral genome sequencing and assembly

To completely assemble a genome for the virus detected in *M. hyperodae*, some preliminary Illumina sequence data were examined to identify an *M. hyperodae* sample with high viral coverage. A library was prepared from this sample for long-read Nanopore sequencing on a MinION (Oxford Nanopore Technologies, ONT) using

whole-genome amplification and genomic DNA by ligation (SQK-LSK110, ONT), and sequenced on a FLO-MIN106 flowcell (ONT). MinION reads were basecalled using Guppy v6.0.0 with the high-accuracy mode, and Porechop v.0.2.4 used to trim adapters off reads (<https://github.com/rrwick/Porechop>). Reads were mapped to the *M. hyperodae* assembly using Minimap2 v2.24 [92], to remove any reads that aligned to the *M. hyperodae* Hi-C scaffolds. The scripts used for the initial base-calling and read filtering are available on Zenodo at <https://doi.org/10.5281/zenodo.7939025> using a Snakemake-managed workflow.

Remaining Nanopore reads were then used for assembly using Flye v2.9 with the metagenome mode [93, 94], with output then subject to viralFlye v0.2 analysis [95] to identify which contigs were likely to represent full viral genomes. A BlastN search of the only complete circular viral genome assembled revealed significant hits to all *M. hyperodae* viral contigs suspected to belong to the same virus, indicating the complete viral genome had been assembled. Medaka v1.7 (<https://github.com/nanoporetech/medaka>) and nextPolish v1.4.1 [96] were used to polish the assembled genome with long then short reads. The filtered MinION and Illumina reads were then mapped back onto the assembled, polished genome using BWA v2.24, and depth of sequencing at each loci determined using Samtools coverage v1.10–98.

Gene prediction was performed using Prodigal v2.6.3 [97] using the metagenome mode. Any predicted genes that had alternate start codons predicted, due to an inflexibility in this option with Prodigal, had annotations reviewed manually and ATG start codons identified using ORFfinder. Predicted protein sequences were then subject to a BlastP search against the nr database, retaining all significant hits and filtering results for the lowest E-value and highest bit-score. Hmmscan v3.2.1 was used against the Pfam database [98] downloaded 3rd February 2020, to identify protein domains in predicted genes, with results filtered to retain domains with an independent E-value below 5e-04. The scripts used for genome assembly and annotation are available on Zenodo at <https://doi.org/10.5281/zenodo.7939027> using a Snake-make-managed workflow.

To determine the phylogenetic placement of *M. hyperodae* filamentous virus, an approach similar to that used by Burke et al., [65] was used. This work had identified 12 core genes (*DNApol*, *helicase*, *lef-5*, *lef-8*, *lef-9*, *p33*, *pif-0*, *pif-1*, *pif-2*, *pif-3*, *pif-5* and *ac81*) conserved in NALDVs. All accessions for exogenous viruses used in their phylogenetic analysis were used, and annotations from BlastP and HMMscan were used to identify these genes in the assembled *M. hyperodae* virus genome. To include

Drosophila-associated filamentous virus (DaFV) in this phylogeny, a BlastP search was performed using the gene predictions available to identify DaFV genes to include in the analysis. Genes and their accessions are detailed in supplementary table 6.

Muscle v3.7 was used to align gene sequences separately, and alignments then concatenated with FAS-conCAT-G v1.05 [99]. The concatenated alignment was trimmed with default parameters and a gap threshold of 0.6 by TrimAl v1.2. A maximum likelihood phylogeny was constructed for the concatenated alignment using RAxML-NG v1.0.2, with a Blosum62 model and 1000 bootstrap replicates. The phylogenetic tree was drawn using Interactive Tree of Life [100], rooted at the midpoint, displaying bootstrap values above 50%. The scripts used for these analyses are available on Zenodo at <https://doi.org/10.5281/zenodo.7939029> using a Snakemake-managed workflow.

To investigate the presence of single nucleotide polymorphisms (SNPs) in the assembled virus genome, trimmed reads were mapped to the *M. hyperodae* and virus genomes concatenated into a single file, using BWA v2.24. Bam files were sorted using Samtools v1.10.2, duplicate reads marked with GATK v4.3.0 MarkDuplicates [101] read groups added using AddOrReplaceReadGroups from the Picard tool v2.27.5 (<https://github.com/broadinstitute/picard>), and resulting bam files indexed with samtools v1.10.2. The bam file was subset to retain only the virus genome, before haplotype calling was performed with GATK v4.3.0 HaplotypeCaller, with ploidy set to five (for a haploid genome with five pooled samples), and variant calling then performed with GenotypeGVCFs. Variants were filtered to retain only SNPs which passed GATK hard-filtering metric recommendations of QD < 2, QUAL < 30, SOR > 3, FS > 60, MQ < 40, MQRankSum < -12.5, ReadPosRankSum < -8, with a minor allele frequency of 0.1. The scripts for virus variant calling are available on Zenodo at <https://doi.org/10.5281/zenodo.7939033> using a Snakemake-managed workflow.

To investigate gene expression support for the virus infection in *M. hyperodae*, RNA-seq data from a previous publication [24] was used (NCBI SRA accession PRJNA841753). Reads were trimmed and quasi-mapped as per Inwood et al. [24] against the *M. hyperodae* and virus predicted genes concatenated into one file. DESeq2 v1.34 [102] was used to create the DESeqDataSet (DDS) object, by importing Salmon output files using tximport v1.22.0 [103] in R v4.1.3 [104], with size factors estimated on the concatenated DDS file. The DDS file was then split into separate objects for *M. hyperodae* and MhFV, and DESeq2 then used to

perform a likelihood ratio test (LRT) on the virus DDS file with the design ~Flowcell+Tissue, to control for sequencing runs on different flow cells and test for the influence of tissue on viral gene expression. Differentially expressed genes (DEGs) were identified by filtering DESeq2 results with the arbitrary alpha threshold value of 0.05 for all analyses. A heatmap of viral gene expression was generated using VST normalized data with pheatmap v1.0.12 [105]. The scripts used for this analysis are available on Zenodo at <https://doi.org/10.5281/zenodo.7939037> using a Snakemake managed workflow.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09538-4>.

Additional file 1.

Acknowledgements

The authors would like to thank Mark McNeill for providing samples and Petra Dearden for proofreading this manuscript.

Author's contributions

SNi; Data generation, bioinformatics, manuscript drafting. JS; Data generation, bioinformatics, manuscript drafting. JGG; bioinformatics. TWRH; bioinformatics. SLG; sample collection and curation. PKD; Funding, project conception, supervision, manuscript drafting.

Funding

This work was funded by a New Zealand Ministry of Business Innovation and Employment Endeavour grant to PKD, funding from the Bioprotection Research Centre (Project 2), and Bioprotection Aotearoa (Project 2.2), both Royal Society of New Zealand funded Centres of Research Excellence.

Availability of data and materials

Raw Illumina sequence data, genome assemblies and gene predictions for *Microctonus* wasps are available at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) and Whole Genome Shotgun (WGS) databases, under accession PRJNA930586, with the Hi-C scaffolded assemblies submitted for *M. aethioides* Irish and *M. hyperodae*, with the latter having identified MhFV contigs removed prior to submission. The genome and annotation for *Microctonus hyperodae* filamentous virus is available at Genbank with the accession number OQ439926. Scripts used for analyses are available as detailed in the methods section.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing Interests

The authors report no competing interests.

Author details

¹Bioprotection Aotearoa and Biochemistry Department, University of Otago, Dunedin, Aotearoa, New Zealand. ²Humble Bee Bio, Wellington, Aotearoa, New Zealand. ³Genomics Aotearoa, University of Otago, Dunedin, Aotearoa, New Zealand. ⁴Melbourne Bioinformatics, The University of Melbourne, Parkville, VIC 3010, Australia. ⁵Biocontrol and Biosecurity Group, AgResearch Limited, Lincoln, Aotearoa, New Zealand.

Received: 22 May 2023 Accepted: 27 July 2023

Published online: 05 August 2023

References

- Ferguson CM, Barratt BIP, Bell N, Goldson SL, Hardwick S, Jackson M, et al. Quantifying the economic cost of invertebrate pests to New Zealand's pastoral industry. *New Zeal J Agric Res.* 2019;62(3):255–315.
- McNeill MR, Goldson SL, Proffitt JR, Phillips CB, Addison PJ. A description of the commercial rearing and distribution of *Microctonus hyperodae* (Hymenoptera: Braconidae) for biological control of *Listronotus bonariensis* (Kuschel) (Coleoptera: Curculionidae). *Biol Control.* 2002;24(2):167–75.
- Goldson SL, McNeill MR, Proffitt JR, Barker GM, Addison PJ, Barratt BIP, et al. Systematic mass rearing and release of *Microctonus hyperodae* (Hym.: Braconidae, Euphorinae), a parasitoid of the argentine stem weevil *Listronotus bonariensis* (Col.: Curculionidae) and records of its establishment in New Zealand. *Entomophaga.* 1993;38(4):527–36.
- Goldson SL, McNeill MR, Stufkens MW, Proffitt JR, Pottinger RP, Farrell JA. Importation and quarantine of *Microctonus hyperodae*, a South American parasitoid of Argentine stem weevil. *Proceedings of the Forty Third New Zealand Weed and Pest Control Conference.* Palmerston North: New Zealand Weed and Pest Control Society Inc; 1990. p. 334–8.
- Barker GM, Addison PJ. Early impact of endoparasitoid *Microctonus hyperodae* (Hymenoptera: Braconidae) after its establishment in *Listronotus bonariensis* (Coleoptera: Curculionidae) populations of northern New Zealand pastures. *J Econ Entomol.* 2006;99(2):273–87.
- Goldson SL, Proffitt JR, Baird DB. Establishment and Phenology of the Parasitoid *Microctonus hyperodae* (Hymenoptera: Braconidae) in New Zealand. *Environ Entomol.* 1998;27(6):1386–92.
- Phillips CB, Kean JM. Response of parasitoid egg load to host dynamics and implications for egg load evolution. *J Evol Biol.* 2017;30(7):1313–24.
- Barker GM. Biology of the introduced biocontrol agent *Microctonus hyperodae* (Hymenoptera: Braconidae) and its host *Listronotus bonariensis* (Coleoptera: Curculionidae) in Northern New Zealand. *Environ Entomol.* 2013;42(5):902–14.
- Goldson SL, Wratten SD, Ferguson CM, Gerard PJ, Barratt BIP, Hardwick S, et al. If and when successful classical biological control fails. *Biol Control.* 2014;72:76–9.
- Tomasetto F, Tylianakis JM, Reale M, Wratten SD, Goldson SL. Intensified agriculture favors evolved resistance to biological control. *Proc Natl Acad Sci U S A.* 2017;114(15):3885–90.
- Popay AJ, McNeill MR, Goldson SL, Ferguson CM. The current status of Argentine stem weevil (*Listronotus bonariensis*) as a pest in the North Island of New Zealand. *New Zeal Plant Prot.* 2011;64:55–62.
- Harrop TWR, Le Lec MF, Jauregui R, Taylor SE, Inwood SN, van Stijn T, et al. Genetic diversity in invasive populations of Argentine stem weevil associated with adaptation to biocontrol. *Insects.* 2020;11(7):1–14.
- Casanovas P, Goldson SL, Tylianakis JM. Asymmetry in reproduction strategies drives evolution of resistance in biological control systems. *PLoS One.* 2018;13(12):e0207610. <https://doi.org/10.1371/journal.pone.0207610>.
- Shields MW, Wratten SD, Phillips CB, Van Koten C, Goldson SL. Plant-Mediated Behavioural Avoidance of a Weevil Towards Its Biological Control Agent. *Front Plant Sci.* 2022;13(923237).
- Shields MW, Wratten SD, Van Koten C, Phillips CB, Gerard PJ, Goldson SL. Behaviour drives contemporary evolution in a failing insect-parasitoid importation biological control programme. *Front Ecol Evol.* 2022;10:923248.
- Stufkens MW, Farrell JA, Goldson SL. Establishment of *Microctonus aethioides*, a parasitoid of the sitona weevil in New Zealand. In *Proceedings of the New Zealand Weed and Pest Control Conference.* New Zealand Weed and Pest Control Society Inc. 1987;40:31–5.
- Gerard PJ, McNeill MR, Barratt BIP, Whiteman SA. Rationale for release of the Irish strain of *Microctonus aethioides* for biocontrol of clover root weevil. *New Zeal Plant Prot.* 2006;59:285–9.
- Barratt BIP, Evans AA, Ferguson CM, Barker GM, McNeill MR, Phillips CB. Laboratory Nontarget Host Range of the Introduced Parasitoids *Microctonus aethioides* and *M. hyperodae* (Hymenoptera:

- Braconidae) Compared with Field Parasitism in New Zealand. *Environ Entomol.* 1997;26(3):694–702.
19. Goldson SL, McNeill MR, Proffitt JR, Phillips CB, Gerard PJ, Murray PJ, et al. British-based search for natural enemies of the clover root weevil, *Sitona lepidus* in Europe. *New Zeal J Zool.* 2004;31(3):233–40.
 20. Goldson SL, McNeill MR, Proffitt JR. Negative effects of strain hybridisation on the biocontrol agent *Microctonus aethiopoidea*. *New Zeal Plant Prot.* 2003;56(November):138–42.
 21. McNeill MR, Proffitt JR, Gerard PJ, Goldson SL. Collections of *Microctonus aethiopoidea* (Hymenoptera: Braconidae) from Ireland. *New Zeal Plant Prot.* 2006;59:290–6.
 22. Gerard PJ, Wilson DJ, Eden TM. Field release, establishment and initial dispersal of Irish *Microctonus aethiopoidea* in *Sitona lepidus* populations in northern New Zealand pastures. *Biocontrol.* 2011;56(6):861–70.
 23. Iline II, Phillips CB. Allozyme markers to help define the South American origins of *Microctonus hyperodae* (Hymenoptera: Braconidae) established in New Zealand for biological control of Argentine stem weevil. *Bull Entomol Res.* 2004;94(3):229–34.
 24. Inwood SN, Harrop TWR, Dearden PK. The venom composition and parthenogenesis mechanism of the parasitoid wasp *Microctonus hyperodae*, a declining biocontrol agent. *Insect Biochem Mol Biol.* 2023;153:103897. Available from: <https://doi.org/10.1016/j.ibmb.2022.103897>.
 25. Coffman KA, Hankinson QM, Burke GR. A viral mutualist employs posthatch transmission for vertical and horizontal spread among parasitoid wasps. *Proc Natl Acad Sci.* 2022;119(16): e2120048119.
 26. Di Giovanni D, Lepetit D, Guinet B, Bennetot B, Boulesteix M, Couté Y, et al. A behavior-manipulating virus relative as a source of adaptive genes for *Drosophila* parasitoids. *Mol Biol Evol.* 2020;37(10):2791–807.
 27. Drezen JM, Leobold M, Bézier A, Huguet E, Volkoff AN, Herniou EA. Endogenous viruses of parasitic wasps: variations on a common theme. *Curr Opin Virol.* 2017;25:41–8.
 28. Ye X qian, Shi M, Huang J hua, Chen X xin. Parasitoid polydnviruses and immune interaction with secondary hosts. *Dev Comp Immunol.* 2018;83:124–9.
 29. Martinez J, Fleury F, Valardi J. Heritable variation in an extended phenotype: The case of a parasitoid manipulated by a virus. *J Evol Biol.* 2012;25(1):54–65.
 30. Barratt BIP, Evans AA, Stoltz DB, Vinson SB, Easingwood R. Virus-like Particles in the Ovaries of *Microctonus aethiopoidea* Loan (Hymenoptera: Braconidae), a Parasitoid of Adult Weevils (Coleoptera: Curculionidae). *J Invertebr Pathol.* 1999;73:182–8.
 31. Barratt BIP, Murney R, Easingwood R, Ward VK. Virus-like particles in the ovaries of *Microctonus aethiopoidea* Loan (Hymenoptera: Braconidae): Comparison of biotypes from Morocco and Europe. *J Invertebr Pathol.* 2006;91(1):13–8.
 32. Goldson SL, McNeill MR, Proffitt JR. Unexplained mortality amongst parasitized *Listronotus bonariensis* in the presence of the parasitoid *Microctonus hyperodae* under caging conditions. In: Proceedings of the 6th Australasian Conference on Grassland Invertebrate Ecology. Ruakura Agricultural Centre, Hamilton; 1993. p. 355–62.
 33. Goldson SL, Proffitt JR, McNeill MR, Phillips CB, Barlow ND, Baird DB. Unexpected *Listronotus bonariensis* (Coleoptera: Curculionidae) mortality in the presence of parasitoids. *Bull Entomol Res.* 2004;94(5):411–7.
 34. Vereijssen J, Armstrong KF, Barratt BIP, Crawford AM, McNeill MR, Goldson SL. Evidence for parasitoid-induced premature mortality in the Argentine stem weevil. *Physiol Entomol.* 2011;36(2):194–9.
 35. Furihata S, Matsumura T, Hirata M, Mizutani T, Nagata N, Kataoka M, et al. Characterization of venom and oviduct components of parasitoid wasp *Asobara japonica*. *PLoS One.* 2016;11(7).
 36. Furihata S, Kimura MT. Effects of *Asobara japonica* venom on larval survival of host and nonhost *Drosophila* species. *Physiol Entomol.* 2009;34(3):292–5.
 37. Ma WJ, Schwander T. Patterns and mechanisms in instances of endosymbiont-induced parthenogenesis. *J Evol Biol.* 2017;30(5):868–88.
 38. Arakaki N, Noda H, Yamagishi K. Wolbachia-induced parthenogenesis in the egg parasitoid *Telenomus nawai*. *Entomol Exp Appl.* 2000;96(2):177–84.
 39. Stouthamer R, Luck RF, Hamilton WD. Antibiotics cause parthenogenetic *Trichogramma* (Hymenoptera:Trichogrammatidae) to revert to sex. *Proc Natl Acad Sci U S A.* 1990;87(7):2424–7.
 40. Phillips CB. Intraspecific variation in *Microctonus hyperodae* and *M. aethiopoidea* (Hymenoptera: Braconidae); significance for their use of biological control agents. (Doctoral dissertation, Lincoln University); 1995.
 41. Wallberg A, Bunikis I, Pettersson OV, Mosbech MB, Childers AK, Evans JD, et al. A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. *BMC Genomics.* 2019;20(1):1–19.
 42. Dalla Benetta E, Antoshechkin I, Yang T, Nguyen HQM, Ferree PM, Akbari OS. Genome elimination mediated by gene expression from a selfish chromosome. *Sci Adv.* 2020;6(14):eaz9808.
 43. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. The genome sequence of *Drosophila melanogaster*. *Science (80-).* 2000;287(5461):2185–95.
 44. Pinto BJ, Weis JJ, Gamble T, Ode PJ, Paul R, Zaspel JM. A Chromosome-Level Genome Assembly of the Parasitoid Wasp, *Cotesia glomerata* (Hymenoptera: Braconidae). *J Hered.* 2021;112(6):558–64.
 45. Tvedte ES, Walden KKO, McElroy KE, Werren JH, Forbes AA, Hood GR, et al. Genome of the parasitoid wasp *Diachasma alloeum*, an emerging model for ecological speciation and transitions to asexual reproduction. *Genome Biol Evol.* 2019;11(10):2767–73.
 46. Ye X, Yan Z, Yang Y, Xiao S, Chen L, Wang J, et al. A chromosome-level genome assembly of the parasitoid wasp *Pteromalus puparum*. *Mol Ecol Resour.* 2020;20(5):1384–402.
 47. Gauthier J, Boulain H, van Vugt JFA, Baudry L, Persyn E, Aury JM, et al. Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. *Commun Biol.* 2021;4(1):1–15.
 48. Vink CJ, Phillips CB, Mitchell AD, Winder LM, Cane RP. Genetic variation in *Microctonus aethiopoidea* (Hymenoptera: Braconidae). *Biol Control.* 2003;28(2):251–64.
 49. Schurko AM, Logsdon JM. Using a meiosis detection toolkit to investigate ancient asexual “scandals” and the evolution of sex. *BioEssays.* 2008;30(6):579–89.
 50. Schurko AM, Mazur DJ, Logsdon JM Jr. Inventory and phylogenomic distribution of meiotic genes in *Nasonia vitripennis* and among diverse arthropods. *Insect Mol Biol.* 2010;19:165–80.
 51. Tvedte ES, Forbes AA, Logsdon JM, Reed F. Retention of Core Meiotic Genes Across Diverse Hymenoptera. *J Hered.* 2017;108(7):791–806.
 52. Rezazadeh S. RecQ helicases; at the crossroad of genome replication, repair, and recombination. *Mol Biol Rep.* 2012;39(4):4527–43.
 53. Doherty KM, Sharma S, Uzdilla LA, Wilson TM, Cui S, Vindigni A, et al. RECQ1 Helicase Interacts with Human Mismatch Repair Factors That Regulate Genetic Recombination*[bboxes]. *J Biol Chem.* 2005;280(30):28085–94.
 54. Petkovic M, Dietschy T, Freire R, Jiao R, Staglar I. The human Rothmund-Thomson syndrome gene product, RECQL4, localizes to distinct nuclear foci that coincide with proteins involved in the maintenance of genome stability. *J Cell Sci.* 2005;118(18):4261–9.
 55. Schurko AM, Neiman M, Logsdon JM. Signs of sex: what we know and how we know it. *Trends Ecol Evol.* 2009;24(4):208–17.
 56. Percy M, Hardy O, Aron S. Thelytokous parthenogenesis and its consequences on inbreeding in an ant. *Heredity (Edinb).* 2006;96(5):377–82.
 57. Oxley PR, Ji L, Fetter-Prunedo I, McKenzie SK, Li C, Hu H, et al. The genome of the clonal raider ant *Cerapachys biroi*. *Curr Biol.* 2014;24(4):451–8.
 58. Crook NE, Clem RJ, Miller LK. An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif. *J Virol.* 1993;67(4):2168–74.
 59. Lu A, Miller LK. The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. *J Virol.* 1995;69(2):975–82.
 60. Lepetit D, Gillet B, Hughes S, Kraaijeveld K, Valardi J. Genome sequencing of the behavior manipulating virus LBFV reveals a possible new virus family. *Genome Biol Evol.* 2016;8(12):3718–39.
 61. Levasseur A, Druela E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels.* 2013;6(1):1–14.
 62. Chiu E, Hijnen M, Bunker RD, Boudes M, Rajendran C, Aizel K, et al. Structural basis for the enhancement of virulence by viral spindles and their in vivo crystallization. *Proc Natl Acad Sci U S A.* 2015;112(13):3973–8.

63. Phanis CG, Miller DP, Cassar SC, Tristem M, Thiem SM, O'Reilly DR. Identification and expression of two baculovirus gp37 genes. *J Gen Virol*. 1999;80(7):1823–31.
64. Klose RJ, Kallin EM, Zhang Y. JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet*. 2006;7(9):715–27.
65. Burke GR, Hines HM, Sharanowski BJ. The presence of ancient core genes reveals endogenization from diverse viral ancestors in parasitoid wasps. *Genome Biol Evol*. 2021;13(7).
66. Kawato S, Shitara A, Wang Y, Nozaki R, Kondo H, Hirono I. Crustacean Genome Exploration Reveals the Evolutionary Origin of White Spot Syndrome Virus. *J Virol*. 2019;93(3):e01144–e1218.
67. Patot S, Martinez J, Allemand R, Gandon S, Varaldi J, Fleury F. Prevalence of a virus inducing behavioural manipulation near species range border. *Mol Ecol*. 2010;19(14):2995–3007.
68. Varaldi J, Boulétreau M, Fleury F. Cost induced by viral particles manipulating superparasitism behaviour in the parasitoid *Leptopilina boulardi*. *Parasitology*. 2005;131(2):161–8.
69. Varaldi J, Petit S, Boulétreau M, Fleury F. The virus infecting the parasitoid *Leptopilina boulardi* exerts a specific action on superparasitism behaviour. *Parasitology*. 2006;132(6):747–56.
70. Patot S, Allemand R, Fleury F, Varaldi J. An inherited virus influences the coexistence of parasitoid species through behaviour manipulation. *Ecol Lett*. 2012;15(6):603–10.
71. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019;20(1):1–13.
72. Kokot M, Długosz M, Deorowicz S. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics*. 2017;33(17):2759–61.
73. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11(1):1432.
74. Chapman JA, Ho I, Sunkara S, Luo S, Schroth GP, Rokhsar DS. Meraculous: de novo genome assembly with short paired-end reads. *PLoS ONE*. 2011;6(8): e23501.
75. Simão 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.
76. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* (80-). 2009;326(5950):289–93.
77. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat Genet*. 2017;49(4):643–50.
78. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40(7):e49–e49.
79. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639–45.
80. Palmer J, Stajich J. Zenodo. 2019. nextgenusfs/funannotate: funannotate v1.6.0. <https://doi.org/10.5281/zenodo.1134477>.
81. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*. 2008;24(5):637–44.
82. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*. 2019;47(D1):D351–60.
83. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. 2015;16(1):1–14.
84. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:1–14.
85. Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, et al. Evolutionary history of the Hymenoptera. *Curr Biol*. 2017;27(7):1013–8.
86. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol*. 2011;7(10): e1002195.
87. Simonsen M, Mailund T, Pedersen CNS. "Rapid neighbour-joining." In: Algorithms in Bioinformatics: 8th International Workshop, WABI 2008, Karlsruhe, Germany, September 15–19, 2008. Proceedings 8. Springer Berlin Heidelberg; 2008. p. 113–22.
88. Shen W, Ren H. TaxonKit: A practical and efficient NCBI taxonomy toolkit. *J Genet Genomics*. 2021;48(9):844–50.
89. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*. 2012;28(19):2520–2.
90. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
91. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
92. Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094–100.
93. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019;37(5):540–6.
94. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods*. 2020;17(11):1103–10.
95. Antipov D, Rayko M, Kolmogorov M, Pevzner PA. viralFlye: assembling viruses and identifying their hosts from long-read metagenomics data. *Genome Biol*. 2022;23(1):1–21.
96. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long read assembly. *Bioinformatics*. 2020;36(7):2253–5.
97. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11(1):119. <https://doi.org/10.1186/1471-2105-11-119>
98. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: The protein families database. *Nucleic Acids Research*. 2014;42(D1):D222–D230.
99. Kück P, Longo GC. FASconCAT-G: Extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool*. 2014;11(1):1–8.
100. Letunic I, Bork P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49(W1):W293–6.
101. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
102. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>
103. Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research*. 2015;4:1521. <https://doi.org/10.12688/F1000RESEARCH.7563.2>.
104. R Development Core Team. A Language and Environment for Statistical Computing. R Found Stat Comput. 2020; <https://www.R-project.org>. Available from: <http://www.R-project.org>
105. Kolde R. Pretty Heatmaps. 2019. p. 1–8. <https://cran.r-project.org/package=heatmap>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

