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

Inwood, S.N[1]§., Skelly, J[1,2]§., Guhlin, J[3], Harrop, T[4]., Goldson, S[5]., Dearden, P.K[1,3]*.

[1] Bioprotection Aotearoa and Biochemistry Department, University of Otago, Dunedin, Aotearoa-New Zealand

[2] Humble Bee Bio, Wellington, Aotearoa-New Zealand

[3] Genomics Aotearoa, University of Otago, Dunedin, Aotearoa-New Zealand.

[4] Melbourne Bioinformatics, The University of Melbourne, Parkville, VIC, 3010, Australia

[5] Biocontrol and Biosecurity Group, AgResearch Limited, Lincoln, Aotearoa-New Zealand.

*To whom correspondence should be addressed.

§ These first authors contributed equally to this work and are listed alphabetically.

## Keywords

Genome assembly, metagenome assembly, *Microctonus*, endoparasitoid wasp, biocontrol, parthenogenesis, virus

# 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 used for biocontrol in Aotearoa New Zealand, which 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.

# Background

Species in the genus of *Microtonus* (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 (Figure 1). *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. (Ferguson et al., 2019), while strains of *M. aethiopoides* Loan are effective against the clover root weevil *Sitona obsoletus* (Gmelin) (Coleoptera: Curculionidae), which causes an estimated NZD$235 in damage p.a. (Ferguson et al., 2019), and the lucerne weevil *Sitona discoideus* (Gyllenhal) (Figure 1).

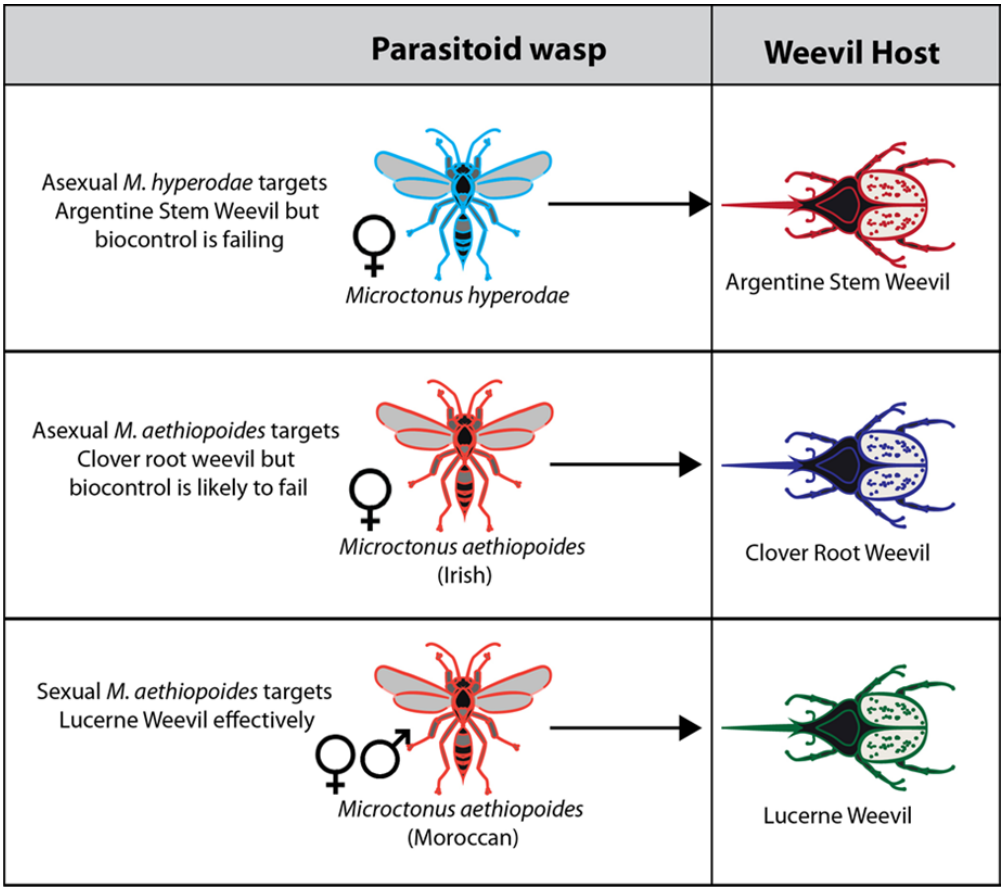


**Figure 1** Host target and reproductive mechanisms of *Microctonus* wasps used as biocontrol agents in New Zealand.

behaviours by the weevil (Shields, Wratten, Phillips, et al., 2022; Shields, Wratten, Van Koten, et al., 2022).

*S. discoideus* is controlled by a sexually reproducing strain of *M. aethiopoides* from the Mediterranean area (Stufkens et al., 1987) and is often referred to as the Moroccan strain (e.g. Gerard et al., 2006), released for biocontrol in 1982. This *M. aethiopoides* strain was subsequently found to have no appreciable effect on the invasive populations of *S. obsoletus* (Barratt et al., 1997), resulting in a widespread European search for a suitable parasitoid. A sexually reproducing French strain of *M. aethiopoides* (Goldson, McNeill, et al., 2004) 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 (Goldson et al., 2003). Continued searching led to the discovery of an asexual European strain of *M. aethiopoides* in Ireland, which reproduces asexually via thelytokous parthenogenesis which would not hybridise with the Moroccan strain (McNeill et al., 2006). This permitted its widespread rearing and release and the eventual suppression of the weevil (Gerard et al., 2006, 2011).

Presence of two reproductive strategies within the *M. aethiopoides* 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 (Goldson et al., 1990), indicating some ability to reproduce sexually. Preliminary examination of the biological underpinnings of reproductive mode in

105 *M. hyperodae* using allozymes revealed an absence of recombination, inferred from a lack of
106 certain homozygous genotypes despite prevalent heterozygotes in populations (Iline &
107 Phillips, 2004). From this lack of recombination, it was suggested that *M. hyperodae*
108 parthenogenesis might be apomictic, using mitosis rather than meiosis. However, more
109 recently the conservation and expression of core meiosis genes in *M. hyperodae* ovaries has
110 been demonstrated, indicating that *M. hyperodae* parthenogenesis may involve meiosis
111 thereby retaining the potential for sexual reproduction (Inwood et al., 2023). No such
112 investigation of the parthenogenesis mechanism of the *M. aethiopoides* Irish strain has been
113 performed to date.
114
115 Another difference between *Microctonus* wasps is in the presence and transmission of virus
116 particles during parasitism. Endogenous viral elements (EVEs) in the form of polydnaviruses
117 (PDVs) or virus-like particles (VLPs) as well as exogenous viruses, often play a role in the
118 parasitism processes of koinobiont endoparasitoids, particularly in host immune-suppression
119 (Coffman et al., 2022; Di Giovanni et al., 2020; Drezen et al., 2017; Martinez et al., 2012; Ye
120 et al., 2018). Barratt et al. (1999, 2006) detected the presence of viral particles in the ovarian
121 epithelial cells of the sexual Moroccan strain of *M. aethiopoides*, with no viral particles found
122 in other *Microctonus* strains or species. There has been no further investigation into the
123 presence of EVEs or infectious viruses associated with the *Microctonus* wasps, and no
124 investigation using genomic data.
125
126 A more comprehensive investigation into the potential presence of viruses or EVEs in *M.*
127 *hyperodae* is required due to a phenomenon of premature mortality observed in *L.*

128 *bonariensis,* whereby greater weevil mortality was observed in the presence of an adult *M.*
129 *hyperodae* than could be explained by parasitism alone (Goldson, McNeill, & Proffitt, 1993;
130 Goldson, Proffitt, et al., 2004; Vereijssen et al., 2011). The current hypothesis is that there
131 may be a toxin-antitoxin system acting during parasitism, whereby *M. hyperodae* transmits
132 something toxic during an unsuccessful ovipositional attempt, which is offset by an ovarian
133 extract during successful parasitism (Vereijssen et al., 2011). This toxin-antitoxin
134 phenomenon has also been observed with the parasitoid *Asobara japonica* Forster
135 (Hymenoptera: Braconidae) and its *Drosophila* host, with the source of toxicity revealed to
136 be virus particles (Furihata et al., 2016; Furihata & Kimura, 2009). The variance in virus
137 particle detection between *Microctonus* species, and the premature mortality phenomenon
138 associated with *M. hyperodae* therefore necessitate an investigation into the viral gene
139 content of the *Microctonus* spp. genome assemblies, and further investigation into the
140 virome of *M. hyperodae*.
141
142 Given the range of host targets and reproductive modes, and the importance of biocontrol
143 to a pastoral economy, the genomes of *M. hyperodae* and *M. aethiopoides* have the
144 potential to determine factors that influence biocontrol efficacy, such as genomic correlates
145 of the reproductive mode and host preference, offering a basis for a better understanding
146 the biology of these biocontrol systems. Here we present the scaffolded genomes of two
147 *Microctonus* species, *M. hyperodae* and *M. aethiopoides*, with additional assemblies for a
148 further two *M. aethiopoides* strains with divergent biology, and that of a previously
149 undetected virus found to be infecting *M. hyperodae*, which provide a unique insight into
150 parasitoid reproduction and commensal viruses that play key roles in parasitic life history.

151 These are valuable genomic resources for understanding the biology of *Microctonus*
152 biocontrol and ongoing investigations into its success or decline in NZ, which is crucial if we
153 are to maintain its efficacy.
154

# 155 Results and Discussion

156

## 157 Wasp Genome assembly

158

159 Bacterial endosymbionts, particularly *Rickettsia, Wolbachia*, and *Cardinium* can induce
160 parthenogenesis in insects (Ma & Schwander, 2017), while maintaining the potential for
161 sexual reproduction (Arakaki et al., 2000; Stouthamer et al., 1990). As *M. hyperodae* and the
162 Irish strain of *M. aethiopoides* reproduce asexually, the presence of such endosymbionts was
163 investigated using Kraken2 read classification. Kraken2 analysis resulted in the classification
164 of 17.7%, 17.3%, 17.4% and 7.25% of reads from *M. aethiopoides* Irish, French, and
165 Moroccan strains, and *M. hyperodae*, with most remaining unclassified (as the database
166 does not contain insects). Classification of reads as known parthenogenesis-inducing
167 endosymbiont genera was low, with 0.05% or less as *Rickettsia*, 0.01% or less of reads
168 classified as *Wolbachia*, and 0.00% as *Cardinium*. There was no correlation between the
169 percentage of reads assigned to the three genera and the reproductive mechanism of the
170 parasitoids. These results are consistent with previous RNA-seq read classification and PCR

171 results from *M. hyperodae* (Inwood et al., 2023), and with antibiotic and heat treatment
172 failing to revert the asexual reproduction mechanism of *Microctonus* wasps (Phillips, 1995).
173
174 Using short-read Illumina sequencing, draft genomes were produced for M. hyperodae, and
175 the Irish, French and Moroccan strains of *M. aethiopoides*, containing 105-128 Mb of total
176 sequence, with a BUSCO completeness of 86.8-93.2% (Table 1). Hi-C scaffolding of the *M.*
177 *hyperodae* and Irish *M. aethiopoides* assemblies improved these assemblies substantially,
178 with N50 lengths increasing from 15 Kb to 9 Mb for *M. hyperodae* and from 64 Kb to 23 Mb
179 for *M. aethiopoides* Irish (Table 1). The contiguity of these scaffolded *Microctonus*
180 assemblies (Table 1) are comparable to the model insects *D. melanogaster*, *A. mellifera* and
181 *N. vitripennis*, which have N50 lengths of 25.2 Mb, 13.6 Mb and 24.7 Mb, and L50 values of
182 3, 5 and 7. Compared to other scaffolded Hymenopteran genomes available on NCBI
183 (https://www.ncbi.nlm.nih.gov/data-hub/genome/?taxon=7399, accessed 14/09/22,
184 excluding contig assemblies) the *M. aethiopoides* Irish Hi-C assembly has an N50 length
185 higher than 90% of assemblies, and *M. hyperodae* higher than 75% of assemblies, indicating
186 that these genomes are more contiguous than most scaffolded Hymenopteran genomes,
187 with a high level of BUSCO completeness (Table 1).
188
189 **Table 1** Assembly and annotation statistics for Meraculous and Hi-C scaffolded assemblies
190 for *M. hyperodae* and *M. aethiopoides* strains. BUSCO completeness (%) refers to the

191 number of BUSCO genes found complete (whether single-copy or duplicated) in the genome
192 assemblies.

| | Hi-C assemblies | | Meraculous assemblies | | | |
|---|---|---|---|---|---|---|
| | M. hyperodae | M. aethiopoides Irish | M. hyperodae | M. aethiopoides Irish | M. aethiopoides French | M. aethiopoides Moroccan |
| Total length (Mb) | 106.7 | 129.2 | 105.8 | 128.8 | 118.9 | 120.3 |
| Number of scaffolds | 3663 | 2844 | 14207 | 6322 | 10395 | 13735 |
| Scaffold N50 (bp) | 9364176 | 23025277 | 15392 | 64081 | 30183 | 17930 |
| L50 | 5 | 3 | 1758 | 484 | 881 | 1634 |
| GC (%) | 29.5 | 29.4 | 29.5 | 29.4 | 29.3 | 29.4 |
| Genome BUSCO completeness (%) | 89.8 | 93.9 | 86.8 | 93.2 | 90.7 | 89.1 |
| Predicted genes | 11744 | 12474 | 13667 | 13521 | 14229 | 14475 |
| Annotation BUSCO completeness (%) | 87.9 | 94.0 | 85.2 | 90.7 | 88.4 | 85.8 |

193

194 The Hi-C data suggests variance of chromosome number between *M. hyperodae* and *M.*
195 *aethiopoides*, with 12 and eight Hi-C scaffolds respectively. Investigation of chromosome
196 synteny suggests that chromosomes 1-4 in *M. aethiopoides* are each represented by two
197 chromosomes in *M. hyperodae* (Figure 2), which may represent fusion of chromosomes in

198 *M. aethiopoides*, or fragmentation of chromosomes in *M. hyperodae* in the time since they

199 have diverged.

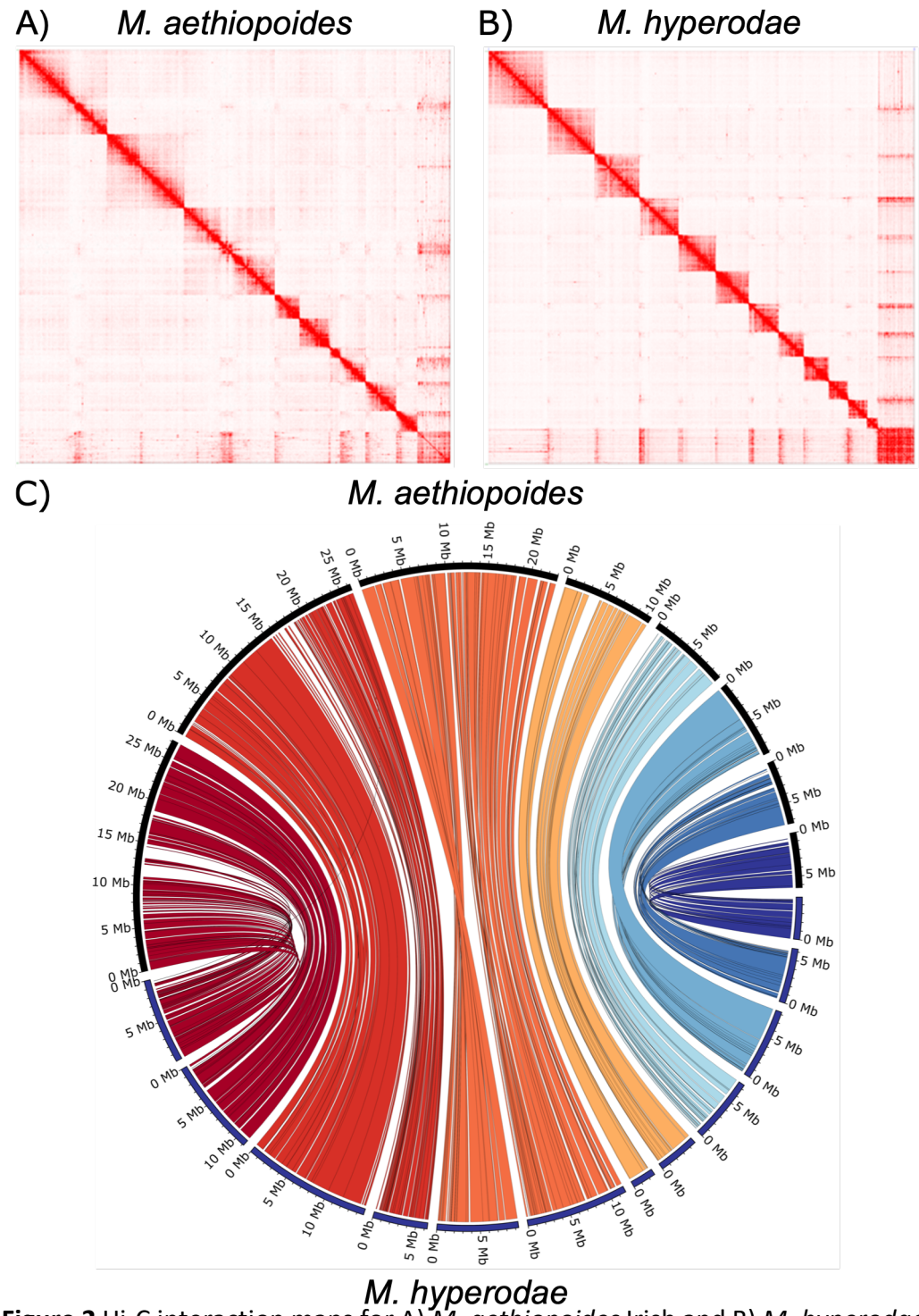


**Figure 2** Hi-C interaction maps for A) *M. aethiopoides* Irish and B) *M. hyperodae* displaying chromosome scaffolds, and C) a Circos plot displaying chromosomal synteny between both Hi-C assemblies

200

201 ## Wasp Genome annotation

202

203 Gene prediction of *Microctonus* genomes was performed with Funannotate, resulting in
204 12,982 to 14,475 gene predictions for each assembly, with a BUSCO completeness of 85.2 to
205 94.1% (Table 1). The high gene prediction BUSCO completeness is comparable to other
206 parasitoid wasps (Dalla Benetta et al., 2020; Gauthier et al., 2021) implying that these are
207 high-quality genomes and gene predictions. While Hi-C scaffolding improved assembly
208 contiguity, there was little difference in BUSCO gene prediction results, indicating the short-
209 read only assemblies are still contiguous enough for gene prediction. Only the gene
210 prediction performed for *M. hyperodae* used RNA-seq data (two ovarian samples), and while
211 this may have resulted in improved gene prediction for the *M. hyperodae* assembly,
212 particularly for genes with high expression in the ovaries, the *M. hyperodae* gene
213 predictions, for both the unscaffolded and scaffolded assemblies, had lower BUSCO
214 completeness scores than those for the *M. aethiopoides* assemblies.

215

216 Using these high-quality annotations, the divergence between *M. hyperodae* and *M.*
217 *aethiopoides* strains was estimated, indicating that the French and Moroccan strains (both
218 sexually reproducing) are more closely related than either is to the asexual Irish strain, with
219 the Irish strain estimated to have diverged from the sexually reproducing strains 2 million
220 years ago (MYA), and the two sexually reproducing strains diverging 1 MYA (Supplementary
221 figure 1). The divergence between *M. hyperodae* and *M. aethiopoides* is estimated to be 17

222 MYA (Supplementary figure 1). This is a larger divergence estimate than was suggested by
223 previous morphological and genetic analysis (Vink et al., 2003).
224
225
## 226 Meiosis gene inventory in sexual and asexual strains
227
228 A 'meiosis detection toolkit' has been described for investigation of asexual reproduction
229 mechanisms in Hymenoptera, cataloguing the presence or absence of genes with known
230 roles in mitosis or meiosis (Schurko & Logsdon, 2008). In an organism that does not use
231 meiosis for gamete production there is no evolutionary constraint on core meiosis genes, so
232 it is expected that their sequence and function would not be conserved (Schurko et al., 2010;
233 Schurko & Logsdon, 2008). Loss of these core meiosis genes would lead to obligate
234 parthenogenesis, where reversion to sexual reproduction is not be possible. This toolkit was
235 used to clarify whether asexual reproduction in *M. hyperodae* and Irish *M. aethiopoides* is
236 automictic, involving meiosis, or apomictic, relying solely on mitosis for egg production
237 which cannot be reverted to sexual reproduction.
238
239 This analysis found 36 of 40 meiosis toolkit genes present in single copies in all *Microctonus*
240 *aethiopoides* assemblies, with the same 36 found in *M. hyperodae* with the addition of
241 RECQ2 (Figure 3). The presence/absence of these genes in these *Microctonus* species is
242 segregated by phylogenetic relatedness, rather than their reproductive mechanism (Figure
243 3). All but one of the core meiosis genes were detected in *Microctonus* genomes, with only
244 DMC1 being absent (Figure 3), though DMC1 is also absent in *Drosophila melanogaster* and

245 the majority of Hymenopteran species assayed regardless of their reproductive mechanism,

246 implying it is dispensable in meiosis in these species (Tvedte et al., 2017).

247

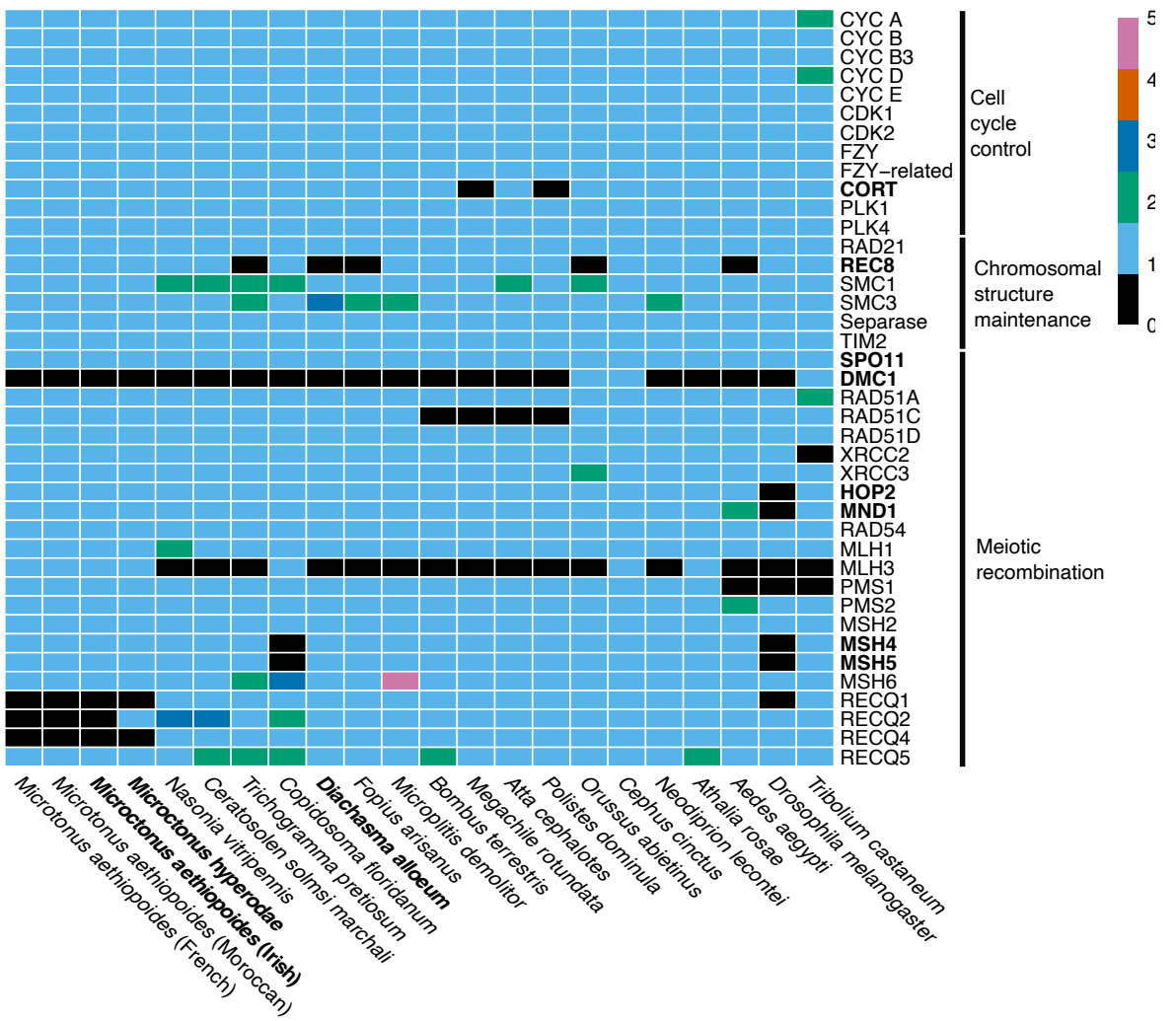


**Figure 3** A heatmap displaying the meiosis gene inventory of *Microctonus*, compared to results in other Hymenoptera, Diptera and Coleoptera from Tvedte et al (2017). Colours in heatmap indicate the absence or presence (and gene number) of each meiosis gene. Core meiosis genes, which are specific to meiosis, and asexual Hymenoptera species are indicated in bold.

248

249 There is variation in the presence and absence of RECQ genes in the *Microctonus*. RECQ1

250 and RECQ4 were not identified in the peptide databases of any *Microctonus*, and RECQ2 was

251 only detected in *M. hyperodae* (Figure 3), though this could be due to gene prediction for *M.*

252 *hyperodae* using RNA-seq data from ovaries while *M. aethiopoides* prediction did not. The
253 RECQ group of proteins are DNA helicases that unwind double-stranded DNA for DNA repair,
254 recombination, and transcription. Some of the RECQ orthologs are required to protect the
255 genome against deleterious mutations (Rezazadeh, 2012). RECQ1 interacts with parts of the
256 DNA mismatch repair pathway during recombination in humans (Doherty et al., 2005).
257 RECQ4 co-localises with Rad51 after the induction of DNA double-strand breaks, with a
258 possible role in DNA double-strand breakthrough homologous recombination (Petkovic et
259 al., 2005). RECQ2 mutants in *S. cerevisiae* have suppressed non-crossover recombinants
260 indicating a role in mediating recombination product formation (Petkovic et al., 2005). The
261 absence of RECQ genes in *Microctonus* genomes may influence rates of recombination
262 leading to lower genetic diversity, particularly in those that reproduce asexually, and may
263 explain the lack of recombination previously observed in *Microctonus hyperodae* at two
264 allozyme loci (Iline & Phillips, 2004).
265
266 Having surveyed the meiosis toolkit genes and examined reads for the presence of
267 endosymbionts that manipulate reproductive mechanisms, we have found no clear evidence
268 for a cause of asexual reproduction in *M. hyperodae* or *M. aethiopoides* Irish. The
269 conservation of most meiosis toolkit genes, particularly all core meiosis genes aside from
270 DMC1 (Figure 3), indicates that asexual reproduction in *M. aethiopoides* Irish and *M.*
271 *hyperodae* likely uses automixis involving meiosis. This is supported by the detection of
272 MND1 expression in the asexual Irish *M. aethiopoides* ovaries (Supplementary Figure 2), and
273 is consistent with the detection of meiosis core gene expression in *M. hyperodae* ovaries

274 (Inwood et al., 2023), implying that these parasitoids have retained the potential to
275 reproduce sexually.
276
277 Heterozygosity rates in these *Microctonus* also do not segregate based on their reproductive
278 mechanisms, with Genomescope Kmer estimates of 0.59%, 0.25%, 0.34% and 0.41% for *M.*
279 *hyperodae* and the Irish, French, and Moroccan *M. aethiopoides*. The formation of gametes
280 via automixis can involve various strategies to restore diploidy, which have varied effects on
281 rates of heterozygosity, from elimination to retention of all heterozygosity (Pearcy et al.,
282 2006). Previous detection of heterozygosity in *M. hyperodae* (Iline & Phillips, 2004) is
283 supported by these results, with heterozygosity also detected in the asexual Irish *M.*
284 *aethiopoides*, indicating that if parthenogenesis is automictic in these parasitoids it relies on
285 a strategy that retains at least some heterozygosity. One such mechanism for this is
286 premeiotic duplication which retains all heterozygosity, with observed upregulation of genes
287 involved in endoreduplication in *M. hyperodae* ovaries providing a putative mechanism for
288 this (Inwood et al., 2023). Investigation of where heterozygous sites occur along
289 chromosomes in the Hi-C scaffolded *Microctonus* genome assemblies could assist in
290 determining this mechanism, as was used in investigation of parthenogenesis in the clonal
291 raider ant *Cerapachys biroi* (Oxley et al., 2014).
292

## 293 Identification of an infectious virus involved in biocontrol

294
295 A reciprocal BlastP search of predicted proteins from the *Microctonus* genomes was
296 performed to investigate viral gene content in *Microctonus* genomes, which identified

297 significant viral hits in all assemblies (Supplementary table 2). The number of RNA virus hits
298 was low as expected, given that DNA sequencing should not detect RNA viruses unless their
299 genes are integrated into the genome. A variable number of DNA virus hits were detected in
300 *M. hyperodae* (across 20 contigs) as well in the Moroccan and French strains of *M.*
301 *aethiopoides* (across five and three contigs) (Supplementary Figure 3). All hits came from
302 viral families known to infect insects, with most *M. aethiopoides* hits to *Polydnaviridae* and
303 most *M. hyperodae* hits to *Baculoviridae* and unclassified DNA viruses (Supplementary
304 Figure 4).
305
306 To investigate whether viral genes detected in these assemblies are endogenous (and
307 putatively involved in PDV particle/VLP production) or exogenous, several approaches were
308 taken. First the presence of eukaryotic genes on the same contigs as viral genes was
309 investigated using BlastP and BUSCO. This revealed no strong evidence of eukaryotic genes
310 on viral contigs in *M. hyperodae*, in which no virus-like particles had previously been
311 detected (Barratt, et al., 1999), and recent transcriptome analysis revealed no clear signs of
312 PDVs or other endogenous viral elements with high expression in venom or ovaries (Inwood
313 et al., 2023). *M. hyperodae* was the only parasitoid with a Hi-C genome and DNA virus hit,
314 with none present on the Hi-C scaffolds. This is despite Hi-C data having coverage for the
315 viral contigs, with 0.15% of Hi-C interactions involving viral contigs, and 83.8% of these viral

316 interactions being between two viral contigs, indicating that their exclusion from the main
317 Hi-C scaffolds was not due to these contigs not being sequenced.
318
319 Presence of eukaryotic genes on contigs with DNA virus hits was also examined in French
320 and Moroccan *M. aethiopoides*, with no DNA virus hits for the Irish strain. For French *M.*
321 *aethiopoides* 10 out of 16 genes on these contigs had eukaryotic hits, with all of these
322 belonging to the Braconidae. Similarly, for Moroccan *M. aethiopoides* 11 of 21 genes had
323 eukaryotic hits, seven of which were to Braconidae, providing some evidence that these viral
324 contigs may be endogenous. No BUSCO genes were present on the contigs with DNA virus
325 hits for the Moroccan or French assemblies, which would have provided stronger evidence
326 of whether DNA virus genes are endogenous or otherwise. By contrast, BUSCO genes were
327 present on all contigs with RNA virus hits, indicating they are endogenous as expected. VLPs
328 had previously been detected in the Moroccan *M. aethiopoides*, but not in the limited
329 number of French or Irish samples collected (Barratt et al., 2006; Barratt, Evans, et al., 1999).
330
331 The GC sequence content and sequencing depth of viral contigs was compared to Hi-C
332 chromosome scaffolds for *M. hyperodae* and to contigs with BUSCO genes for *M.*
333 *aethiopoides* strains to provide further evidence of whether viral contigs were exogenous or
334 endogenous (Supplementary Figure 4). There is a significant difference between both the GC
335 content and sequencing depth of viral contigs and Hi-C scaffolds in *M. hyperodae* (GC: mean
336 34.0% compared to 29.5%, p = 6.7E-08, depth: mean 258 compared to 175, p = 1.6E-03),
337 with this depth result still significant when two high outlier contigs (scaffolds 90 and 995,
338 mean depth of 879) are removed (mean 189 compared to 175, p = 1.3E-02) (Supplementary

339 Figure 4). There was also a significant difference in GC content and depth between BUSCO
340 and viral contigs in *M. aethiopoides* French (GC: mean 35.1% compared to 28.9%, p = 1.9E-
341 02, depth: mean 172 compared to 281, p = 7.4E- 03) while in *M. aethiopoides* Moroccan
342 there was no significant difference in either metric (GC: mean 32.3% compared to 29.0%, p =
343 0.12, depth: mean 548 compared to 194, p = 0.88) even when the high depth outlying contig
344 in *M. aethiopoides Moroccan* (depth = 1822), which contains the only *Baculoviridae*,
345 unclassified DNA virus and hemoflagellate parasite hits, was analysed in a separate group (p
346 ≥ 0.50) (Supplementary Figure 4).
347
348 Only *M. aethiopoides Moroccan* returned non-significant results for both comparisons
349 (Supplementary Figure 4) providing more evidence supporting that these contigs may be
350 endogenous, with this strain previously demonstrated to have VLPs in their ovaries (Barratt
351 et al., 1999, 2006). No analysis has been done on what these particles contain beyond an
352 inability to extract DNA from them, so it is unknown whether they have a viral origin, though
353 these identified viral genes in *M. aethiopoides* Moroccan are candidates for involvement in
354 VLP production. A more contiguous genome assembly would be required to confirm
355 whether they are exogenous or endogenous. The Hi-C data, GC and sequencing depth

356 analyses do however provide strong evidence that the viral contigs in *M. hyperodae* are
357 derived from an exogenous viral infection.
358

## 359 *M. hyperodae* metagenome assembly.

360

361 Given these results, efforts were therefore made to assemble a complete genome for the
362 exogenous virus in *M. hyperodae*. Meta-genomic assembly was performed using MinION
363 long reads, which were generated from a sample that preliminary Illumina sequencing had
364 confirmed infection status with high viral coverage (mean viral depth of 19x, compared to
365 mean Hi-C scaffold depth of 9x). This generated 354,749 MinION reads with a mean 65x
366 depth of the viral contigs and 19x depth of the Hi-C scaffolds. All reads that mapped to the
367 *M. hyperodae* Hi-C scaffolds were removed, retaining 44,388 reads with a mean length of
368 7305 bp.
369

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

379 number of BlastP hits to LbFV and later phylogenetic analysis, we provisionally name this
380 virus *Microctonus hyperodae* filamentous virus (MhFV).
381

## 382 Complete genome assembly of *M. hyperodae* filamentous virus.

383

384 Illumina read coverage for this complete MhFV assembly was variable (Figure 4), with a
385 maximum of 494x, and a mean of 147x. Illumina short read coverage dropped to 0x in 44
386 regions, ranging in length from 1 bp to 370 bp, with a mean of 88 bp and total length of 3860
387 bp. Nanopore read coverage was less variable (with a maximum of 168x and a mean of 69x)
388 and didn't drop below 4x (Figure 4). The MhFV genome assembly is 163 Kb long, with a GC
389 content of 37.8% and has comparable characteristics compared to other nuclear arthropod-
390 specific large dsDNA viruses (NALDVs) (Supplementary Table 3).
391

392 There are 158 predicted genes in the MhFV genome, all of which are complete and 88.6% of
393 which start with a methionine (minimum amino acid length = 34, maximum = 1652, mean =
394 286.7). Of the 158 predicted genes, there were significant BlastP hits or Pfam protein
395 domains for 86, leaving 72 genes without any significant homology to indicate potential
396 function (Supplementary Table 4). The most common virus hits were to *Leptopilina boulardi*
397 filamentous virus (LBFV) with 23 hits, followed by *Drosophila*-associated filamentous virus
398 (DaFV, eight hits) and *Glossina pallidipes* salivary gland hypertrophy virus (five hits)
399 (Supplementary Table 4). The predicted genes contain several known viral gene families,

400 such as *per os* infectivity factors, late expression factors, five KilA-N domain-containing

401 genes, and 17 BRO genes (Supplementary Table 4).

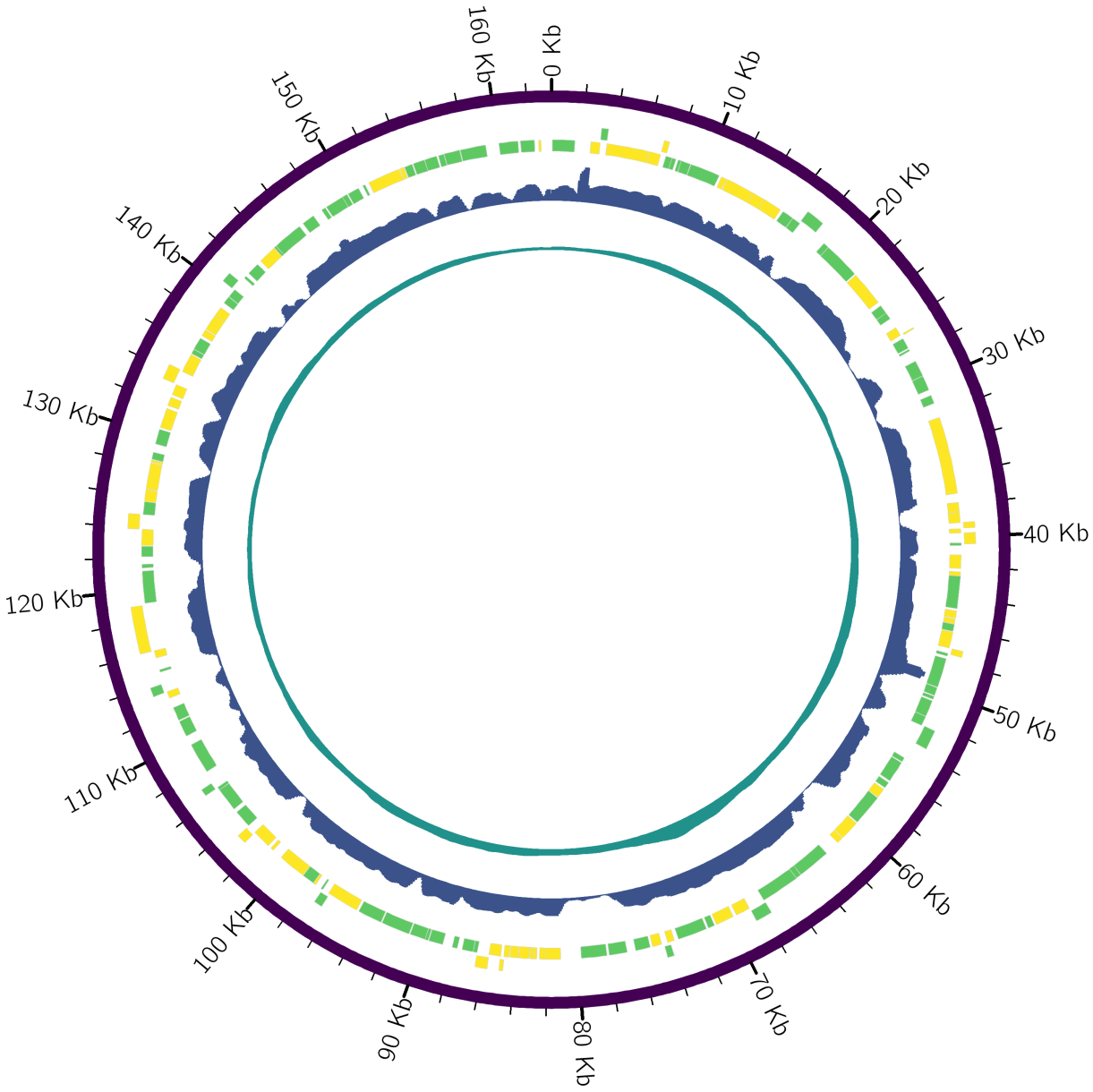


**Figure 4** A Circos plot of the MhFV genome assembly. Predicted genes are indicated in the yellow and green blocks, on the positive and negative strand respectively. Relative read depth for Illumina and ONT MinION sequencing is indicated in blue and teal respectively, calculated with a sliding window with a size of 500bp and slide of 100bp.

402

403 While there were 27 hits to bacteria and eukaryotes, an investigation of the full list of Blast
404 hits revealed 11 of these genes also had other viral hits, suggesting potential viral
405 contamination in the assemblies that the best hits came from, rather than there being a
406 putative non-viral origin for all 27 genes. Three genes had only non-viral BlastP hits,
407 including two inhibitor of apoptosis (IAP) genes and a lytic polysaccharide monooxygenase
408 (Supplementary Table 4). ORF32 and ORF152 are both annotated as IAPs (Supplementary
409 Table 4), which are often found in NALDVs and thought to manipulate the host immune
410 system by suppressing apoptosis of infected cells (Crook et al., 1993; Lu & Miller, 1995). IAP
411 genes were also present in the LbFV genome and had a eukaryotic origin (Lepetit et al.,
412 2016). ORF116 and ORF133 are both annotated as lytic polysaccharide monooxygenases
413 (Supplementary Table 4), with at least 97% of hits for both genes from bacterial species.
414 While none of these hits were viral, many viruses in the families Poxviridae and
415 Baculoviridae contain genes named Fusolin and GP37, which belong to the lytic
416 polysaccharide monooxygenase family (Levasseur et al., 2013). These act to disrupt the
417 peritrophic matrix of insect hosts and facilitate viral infection (Chiu et al., 2015; Phanis et al.,
418 1999). ORF67 was annotated as containing a JmJc-domain (Supplementary Table 4), and
419 while it did have viral hits with one from DaFV and two from different LbFV JmJc-domain
420 containing proteins, there were no other viral hits and 497 non-viral hits. Phylogenetic
421 analysis of the JmJc-domain genes in LbFV indicated they were likely acquired via horizontal

422 transfer from an ancestral host parasitoid (Lepetit et al., 2016). JmJc-containing proteins are
423 a class of demethylase enzymes involved in transcription regulation (Klose et al., 2006).
424

## 425 Phylogenetic position of MhFV.

426

427 Recent work identifying endogenous viral elements (EVEs) in parasitoid genomes has
428 identified 12 core genes (*DNApol*, *helicase*, *lef-5*, *lef-8*, *lef-9*, *p33*, *pif-0*, *pif-1*, *pif-2*, *pif- 3*, *pif-*
429 *5* and *ac81*) conserved in many NALDVs, and EVEs derived from NALDVs (Burke et al., 2021).
430 Of these core genes, ten were identified in the LbFV genome, missing both *pif-3* and *lef-5*. A
431 BlastP search of gene predictions from the incomplete assembly of DaFV only identified four
432 of the core genes. The MhFV genome has predicted genes with significant BlastP and/or
433 protein domain hits to 11 of these core genes. MhFV is missing only *lef-5*, which is the core
434 gene found in the least number of NALDVs due to its short length (Burke et al., 2021).
435

436 Phylogenetic analysis was carried out using protein sequences of the 12 core genes, from all
437 NALDVs used in the analysis by Burke et al., (2021). EVEs derived from NALDVs were
438 excluded, as were the LbFV-like viral contigs identified in that analysis that may have an
439 endogenous origin, as MhFV is an active viral infection rather than an EVE. The resulting
440 phylogeny is globally well resolved, and all characterized viral families were monophyletic
441 displaying expected relationships (Figure 5) (Burke et al., 2021). Despite sharing similar
442 names, *Apis mellifera* filamentous virus and the LbFV-like viruses are separated on the
443 phylogeny, which is the expected result. The placement of LbFV outside of characterized
444 NALDV families is the same result found in other analyses (Burke et al., 2021; Kawato et al.,

445 2019; Lepetit et al., 2016), with the closest family in all previous analyses being

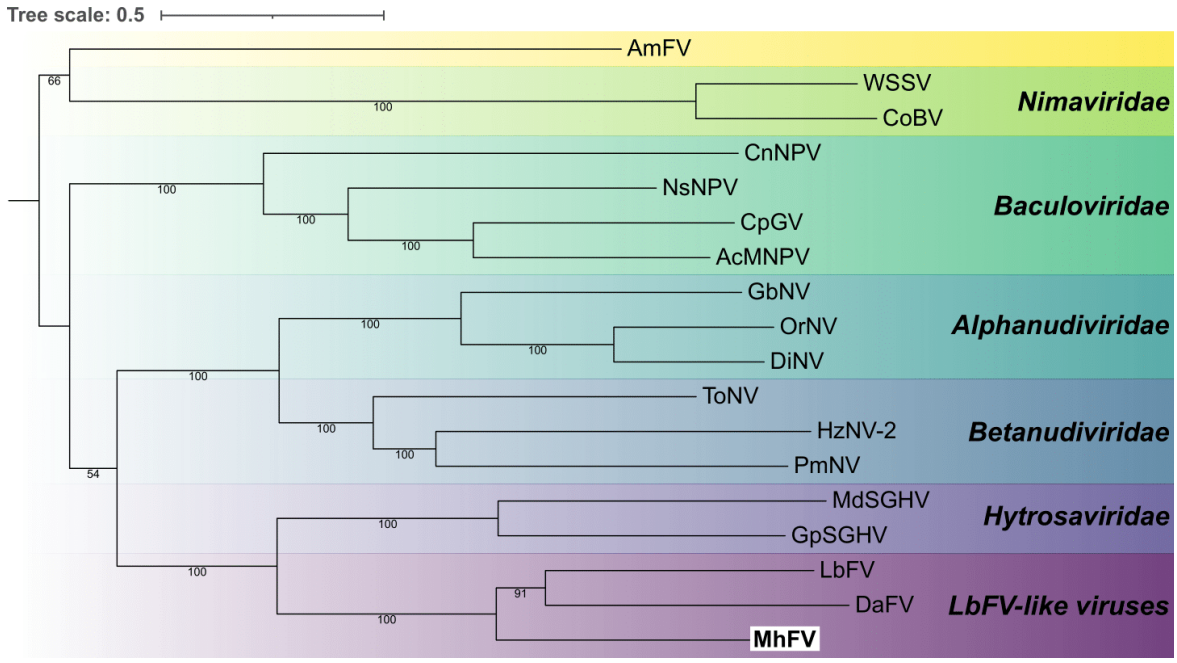


**Figure 5** Phylogenetic analysis of nuclear arthropod-specific large double-stranded DNA viruses (NALDVs). Relationships were derived using a maximum likelihood analysis with RAxML-NG, from 12 core NALDV genes, as defined by Burke et al., (2021), with a total of 6818 characters from concatenated amino 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), Penaeus 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).*

446

447 Hytrosaviridae with which a monophyletic group is formed (Figure 5). This analysis places *M.*

448 *hyperodae* filamentous virus in the same clade as LbFV and DaFV (Figure 5), supporting that

449 MhFV is likely a member of this uncharacterized LbFV-like virus family, though electron
450 microscopy would be required to confirm whether MhFV also has filamentous viral particles.
451
452 Manual inspection of mapped Illumina reads revealed the presence of many single
453 nucleotide polymorphisms (SNPs) in the MhFV genome. This variation is likely a result of
454 Illumina sequencing having been performed on a sample with five pooled *M. hyperodae*
455 individuals, as the variability in the MinION reads derived from a single individual is much
456 lower. Variant calling and filtering revealed 1674 SNPs in the MhFV genome, compared to
457 only two in the 111 kb LbFV genome assembly, which was also assembled from a pooled
458 sample (Lepetit et al., 2016). This indicates a relatively high level of variation in the MhFV
459 genome that future population genetics research on *M. hyperodae* should also consider.
460

## 461 MhFV expression in *M. hyperodae* tissues

462 A BlastN search of the MhFV gene predictions against a previous *de novo* transcriptome
463 assembly for *M. hyperodae* (Inwood et al., 2023) revealed 79 transcripts with significant hits
464 to 89 MhFV predicted genes. When this RNA-seq dataset is re-examined with the *M.*
465 *hyperodae* and MhFV genomes, mean TPM values for each predicted MhFV gene in adult
466 tissue samples range from 1.23 (ORF80) to 0.00, while mean TPM in pupa samples ranges
467 from 16.48 (ORF125) to 0.00. The highest mean TPM value for each adult tissue is 3.12 in the
468 abdomen (ORF156), 1.44 in the ovary (ORF88), 1.09 in the thorax (ORF80), 0.64 in the venom
469 gland (ORF125) and 0.31 in the head (ORF156). This data provides transcriptomic support for
470 MhFV infection in *M. hyperodae*, revealing levels of expression are relatively low in adult
471 tissue samples, though this may have been impacted by the use of poly(A) enrichment

472 during library preparation. Differential gene expression analysis revealed 12 MhFV genes
473 with expression patterns significantly influenced by tissue type. These genes had the highest
474 expression in abdomen samples (Supplementary Figure 5), which suggests future efforts to
475 image MhFV particles should focus on tissues in the abdomen.
476

## 477 Implications of MhFV Infection

478

479 MhFV has 23 genes with significant BlastP hits to LbFV, and phylogenetic analysis indicates it
480 is in the same uncharacterized viral family. LbFV infects *Leptopilina boulardi* parasitoids with
481 prevalence varying between 0% to 95% depending on parasitoid density (Patot et al., 2010),
482 and manipulates *L. boulardi* oviposition behaviour, significantly increasing superparasitism of
483 their *Drosophila* host to facilitate horizontal transmission of the virus (Varaldi et al., 2005,
484 2006). This superparasitism causes decreased overall parasitism rates due to egg wastage,
485 and infected *L. boulardi* are outcompeted by the related parasitoid *L. heterotoma* (Patot et
486 al., 2012). LbFV infection also causes an increase in developmental time and egg load
487 (thought to offset egg wastage from superparasitism), and a decrease in female tibia length.
488 *L. boulardi* eggs infected with LbFV are also encapsulated significantly less by *Drosophila*,
489 and so this infection may also assist in the avoidance of the host immune system during
490 parasitism (Martinez et al., 2012). This host manipulation could provide another mechanism
491 by which *M. hyperodae* avoids encapsulation of their eggs by *L. bonariensis* (Tomasetto et
492 al., 2017), alongside identified venom components (Inwood et al., 2023), and therefore the
493 potential for MhFV transmission during parasitism should be investigated in future. The

494 possibility for behavioural manipulation of *M. hyperodae* by MhFV should also be
495 considered, but not assumed.
496
497 Future research into this biocontrol system should consider the significant impact that the
498 MhFV infection may play in *M. hyperodae* and biocontrol of ASW. MhFV may play a role in
499 the premature mortality phenomenon observed in ASW exposed to *M. hyperodae* (Goldson,
500 et al. 1993; Goldson, et al 2004; Vereijssen et al., 2011), a hypothesis supported by the cause
501 of toxicity during interrupted parasitism attempts with the parasitoid *A. tabida* being
502 identified as viral particles (Furihata et al., 2016). Future work should therefore investigate
503 whether MhFV is transmitted to ASW during successful parasitism and failed parasitism
504 attempts.
505
506 Determining the prevalence of MhFV in *M. hyperodae* populations around NZ is also of
507 interest, to see if it is varied between locations with different declines in parasitism rates. If
508 populations without MhFV were identified, these could then be used to determine the
509 physiological effects of MhFV infection on *M. hyperodae*. Examining historical *M. hyperodae*
510 samples, reared and stored from lines imported into NZ from South America before release
511 for biocontrol, should provide evidence of whether MhFV was brought into NZ from their
512 home range, as well as allowing for investigation of whether the prevalence and/or viral load
513 differs in the historical and contemporary samples. Finally, given the large number of
514 variants detected in the MhFV genome with pooled sequencing of five individuals, genetic
515 variation in the virus genome should be considered alongside that in the parasitoid genome,

516 both in comparisons between contemporary locations with different rates of biocontrol
517 decline, and compared to historical samples.
518

# 519 Conclusions

520

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

530 genomic factors that influence success or failure of *Microctonus*-based biocontrol in New
531 Zealand, and for ongoing investigation of the *M. hyperodae* biocontrol decline.
532

# 533 Methods

534

## 535 *Microctonus* spp. samples & sequencing

536

537 *Microctonus* spp. samples were supplied by AgResearch, Lincoln,consisting of pools of five
538 *M. hyperodae*, five Moroccan and 10 French *M. aethiopoides* frozen in ethanol as well as 23
539 live Irish *M. aethiopoides*. DNA was extracted from these samples using a DNeasy kit
540 (Qiagen), and prepared for sequencing using the TruSeq DNA PCR-Free platform. Samples
541 were sequenced on an Illumina HiSeq 2500 by the Otago University Genomics Facility
542 (https://www.otago.ac.nz/genomics/index.html), generating 26, 28, 24, and 35 Gb of 250 bp
543 paired-end reads respectively.
544

## 545 Genome assembly and scaffolding

546

547 The BBTools v37.57 suite BBDuk program (http://jgi.doe.gov/data-and-tools/bb-tools/) was
548 used to trim and decontaminate reads, using default settings for adapter and quality
549 trimming, while also removing the last 5 bp from each read. Kraken2 v2.0.7 was used to
550 taxonomically classify reads against the Kraken standard database (downloaded 17th
551 September 2018) (Wood et al., 2019). Kmer counting was performed using KMC v3.1.1

552 (Kokot et al., 2017) with a kmer length of 21, for Genomescope2 analysis on the web-based
553 interface (Ranallo-Benavidez et al., 2020).
554
555 Meraculous v2.2.5 was used for genome assembly (Chapman et al., 2011), trialling a range of
556 Kmer values, with both trimmed reads and reads normalised by BBNorm to a maximum
557 coverage of 50 and minimum of 5. A range of parameters (including N50 and L50, as
558 determined by the BBTools v38.0 stats wrapper), comparison to estimated genome sizes,
559 and BUSCO v5.4.3 analysis (Simão et al., 2015) were used to determine the best parameters
560 for each genome. Final Kmer values for assembly were as follows: 79 for *M. aethiopoides*
561 French and Irish, 71 for *M. aethiopoides* Moroccan, and 41 for *M. hyperodae*. Normalised
562 reads were used for the French and Moroccan assemblies, and non-normalised for French
563 and *M. hyperodae*. Assembly was run in haploid mode for Irish, and diploid mode for French,
564 Moroccan and *M. hyperodae* assemblies.
565
566 Draft genomes for *M aethiopoides* Irish and *M. hyperodae* were scaffolded by Phase
567 Genomics using Hi-C data generated from pools of 10 individuals. Phase Genomics Proximo
568 Hi-C 2.0 kit, a commercially available version of the Hi-C protocol (Lieberman-Aiden et al.,
569 2009) was used to generate chromatin conformation capture data. These data were then
570 used with the Phase Genomics Proximo Hi-C genome scaffolding platform to create
571 chromosome scale scaffolds, as described in Bickhart et al. (2017). BUSCO v5.4.3, and the
572 BBTools v38.0 stats wrapper were then run on Hi-C scaffolded assemblies, with exogenous
573 viral contigs (identification detailed below) removed from the *M. hyperodae* assembly. To
574 investigate genome synteny between the *M. hyperodae* and *M. aethiopoides* Hi-C

575 assemblies, MCScanX was used to identify the 2000 best linkage bundles (Wang et al., 2012),
576 which were then visualised using Circos v0.69-9 (Krzywinski et al., 2009).
577

## 578 Genome Annotations

579

580 Gene prediction was performed using Funannotate v1.5.0-12dd8c7 (Palmer & Stajich, 2019).
581 Repeats were identified using RepeatModeler v1.0.11 (github.com/Dfam-
582 consortium/RepeatModeler) and RepeatMasker v1.5.0
583 (repeatmasker.org/RMDownload.html) via the Funannotate pipeline. The Funannotate
584 model was then trained using Funannotate train on the masked genome assembly using
585 *Microctonus hyperodae* or *Microctonus aethiopoides* sequences depending on the species of
586 interest. Augustus v 3.3.1 was iteratively trained (Stanke et al., 2008) and used as input for
587 Funannotate predict to predict genes, using the Hymenoptera BUSCO and optimized
588 Augustus settings. Funannotate update was subsequently used to upgrade the annotation.
589 Further annotation was performed with Funannotate using InterProScan 5.32-71.0 (Mitchell
590 et al., 2019). To provide additional data for *M. hyperodae* gene prediction, RNA was
591 extracted from two pools of ovaries using a hybrid of Trizol (Ambion) and RNeasy mini kit
592 (Qiagen) methods. RNA was prepared for using the TruSeq Stranded mRNA platform, and
593 sequenced on an Illumina HiSeq 2500 by Otago Genomics Facility

594 (https://www.otago.ac.nz/genomics/index.html) to generate 250 bp paired-end reads. RNA-
595 seq data was not available for *M. aethiopoides* gene prediction.
596
597 To estimate the divergence between *Microctonus* species, Orthofinder v2.3.12 (Emms &
598 Kelly, 2015, 2019) was run in multiple sequence alignment mode, using the diamond search
599 engine and maft, on the genomes of 25 Hymenoptera species and *Drosophila melanogaster*,
600 to produce an ultrametric species tree from peptide databases. The divergence of
601 *Acromyrmex echinatior* and *Camponotus floridanus* of 62 MYA (Peters et al., 2017) was used
602 to calibrate the molecular clock, with a branch length sum of 0.1101, which divided by 62
603 MYA gave a scaling factor of 563.12. Scaling the ultrametric Orthofinder species tree in
604 Figtree resulted in a phylogenetic tree with inferred divergence estimates, which allowed for
605 divergence between *M. hyperodae* and *M. aethiopoides* to be estimated.
606

## 607 Identification of mitosis and meiosis-related Genes.

608
609 To investigate the parthenogenesis mechanism of asexual *Microctonus*, 40 genes known to
610 play a role in meiosis and mitosis (including 8 genes with roles specific to meiosis) were
611 collected for 15 Hymenopteran species, as well as *Aedes aegypti*, *D. melanogaster* and
612 *Tribolium castaneum*, from the NCBI database corresponding to accessions used in a
613 previous analysis by Tvedte et al. (2017). Fasta files were made for the genes and ortholog
614 groups the cyclins (CYC A, CYC B, CYC B3, CYC D and CYC E), the cyclin-dependent kinases
615 (CDK1 and CDK2), CDC20 homologs (CORT, CDC20/FZY and CDC20-like/FZY-related), the
616 Polo-like kinases (PLK1 and PLK4), RAD21 and REC8, structural maintenance of the

617 chromosomes orthologs (SMC1 and SMC3), RAD51 orthologs (RAD51A, RAD51C, RAD51D,
618 XRCC2 and XRCC3), heterodimers HOP2 and MND1, MutL orthologs (MLH1, MLH3, PMS1
619 and PMS2), MutS homologs (MSH2, MSH4, MSH5 and MSH6), RECQ helicase orthologs
620 (RECQ1, RECQ2, RECQ4 and RECQ5), RAD54, SPO11, TIM2, DMC1, and Separase.
621
622 Sequences for each gene and ortholog group were aligned using Muscle v3.8.13 and
623 trimmed using TrimAl v1.2 with the strictplus trimming parameters. Gene-specific hidden
624 Markov models (HMMs) were generated using hmmbuild from Hmmer v3.2.1 (Eddy, 2011).
625 These HMMs were used to search the *Microctonus* peptide databases with hmmsearch,
626 using an E-value of 1E-15. Top hits were selected and peptides retrieved using esl-sfetch
627 from Hmmer v3.2.1. Peptide sequences for each gene were aligned using Muscle v3.8.13
628 and trimmed using TrimAL v1.2 with the "strictplus" parameters. Phylogenetic trees were
629 constructed for individual genes and ortholog groups using rapidnj v2.3.2 (Simonsen et al.,
630 2008) with a bootstrap value of 1000 for genes and 1000000 for ortholog groups.
631

## 632 Hybridisation chain reaction in *Microctonus aethiopoides* Irish.

633
634 Parthenogenetic *Microctonus aethiopoides* (Irish) were reared from CRW collected from
635 pasture in Timaru, New Zealand, with ovaries dissected from 10 wasps three days after
636 eclosion. Ovaries were fixed, and hybridisation chain reaction (HCR) performed with an
637 MND1 probe (with fluorophore 488) as per Inwood et al, (2023). Ovaries were stained with

638 DAPI at a 1:1000 concentration, and mounted on slides in glycerol, for imaging on an
639 Olympus BX61 Fluoview FV100 confocal microscope with FV10-ASW 3.0 imaging software.
640

## 641 Identification of viral genes in *Microctonus* genome assemblies.

642

643 In order to identify genes with a viral origin in the *Microctonus* genome assemblies, the
644 predicted peptide databases were searched for peptides with significant homology to viral
645 genes using a reciprocal BlastP v2.9.0 search approach. First a BlastX v2.12.0 search was
646 performed (with an E value of 1E-05 to minimise false-positive results) of all transcripts
647 against the nr database downloaded on May 16$^{th}$, 2021, restricted to only viral entries in the
648 database by using TaxonKit 0.8.0 (Shen & Ren, 2021) to produce a list of all viral taxonomy
649 identifiers at a species level or below and restricting the BlastP search to this list using the -
650 taxidlist option. Any gene with a significant viral BlastX result was then used in a subsequent
651 BlastP search against the whole nr database to remove those genes with better non-viral
652 hits. Finally, any remaining genes on contigs with viral hits that did not have a viral hit were
653 also subject to a BlastP search against the whole nr database to identify potentially
654 eukaryotic genes on the viral contigs. The scripts used for this viral gene identification are
655 available on Zenodo at doi: 10.5281/zenodo.7939016 using a Snakemake-managed
656 workflow (Köster & Rahmann, 2012).
657

658 GC content and sequencing depth were compared between Hi-C scaffolds and contigs with
659 DNA virus hits for *M. hyperodae*, and between contigs containing BUSCO hits and contigs
660 with DNA virus hits for *M. aethiopoides* French and Moroccan. The GC content for all contigs

661 was calculated using BBStats v38.0. To determine read depth of contigs,reads were trimmed
662 using BBDuk v38.0 using default settings and trimq=15, and then mapped to the appropriate
663 genome assembly using BWA v0.7.15 (Li & Durbin, 2009). SAMtools coverage v1.10-98 (Li et
664 al., 2009) was then used to determine read mean sequencing depth of contigs. To test for
665 statistical differences between viral and Hi-C/BUSCO contigs, a Shapiro-Wilk test was first
666 carried out to determine whether data were normally distributed, which indicated non-
667 parametric tests should be used. A Kruskal-Wallis rank sum test was used to identify
668 whether there was a significant difference between groups, followed by a pairwise Wilcoxon
669 rank sum test. All statistical tests used an alpha threshold of 0.05 as indication of a
670 significant difference. The scripts used for these sequence characteristic comparisons are
671 available on Zenodo at doi: 10.5281/zenodo.7939021 using a Snakemake-managed
672 workflow.
673

## 674 Viral genome sequencing and assembly

675

676 To completely assemble a genome for the virus detected in *M. hyperodae*, some preliminary
677 Illumina sequence data were examined to identify an *M. hyperodae* sample with high viral
678 coverage. A library was prepared from this sample for long-read Nanopore sequencing on a
679 MinION (Oxford Nanopore Technologies, ONT) using whole-genome amplification and
680 genomic DNA by ligation (SQK-LSK110, ONT), and sequenced on a FLO-MIN106 flowcell
681 (ONT). MinION reads were basecalled using Guppy v6.0.0 with the high-accuracy mode, and
682 Porechop v.0.2.4 used to trim adapters off reads (https://github.com/rrwick/Porechop).
683 Reads were mapped to the *M. hyperodae* assembly using Minimap2 v2.24 (Li, 2018), to

684 remove reads that aligned to the *M. hyperodae* Hi-C scaffolds. The scripts used for the initial
685 base-calling and read filtering are available on Zenodo at doi: 10.5281/zenodo.7939025
686 using a Snakemake-managed workflow.
687
688 Remaining reads were then used for assembly using Flye v2.9 with the metagenome mode
689 (Kolmogorov et al., 2019, 2020), with output then subject to viralFlye v0.2 analysis (Antipov
690 et al., 2022) to identify which contigs represented full viral genomes. A BlastN search of the
691 only complete circular viral genome assembled revealed significant hits to all *M. hyperodae*
692 viral contigs suspected to belong to the same virus, indicating the complete viral genome
693 had been assembled. Medaka v1.7 (https://github.com/nanoporetech/medaka) and
694 nextPolish v1.4.1 (Hu et al., 2020) were used to polish the assembled genome with long then
695 short reads.The filtered MinION and Illumina reads were then mapped back onto the
696 assembled, polished genome using BWA v2.24, and depth of sequencing at each loci
697 determined using Samtools coverage v1.10-98. Gene prediction was performed using
698 Prodigal v2.6.3 (Hyatt et al., 2010) using the metagenome mode. Predicted protein
699 sequences were subject to a BlastP search against the nr database, retaining all significant
700 hits and filtering results for the lowest E-value and highest bit-score. Hmmscan v3.2.1 was
701 used against the Pfam database (Finn et al., 2014) downloaded 3rd February 2020, to identify
702 protein domains in predicted genes, with results filtered to retain domains with an
703 independent E-value below 5e-04. The scripts used for genome assembly and annotation are

704 available on Zenodo at doi: 10.5281/zenodo.7939027 using a Snakemake-managed
705 workflow.
706
707 To determine the phylogenetic placement of *M. hyperodae* virus 1, an approach similar to
708 that used by Burke et al., (2021) was used. This work had identified 12 core genes (*DNApol,*
709 *helicase, lef-5, lef-8, lef-9, p33, pif-0, pif-1, pif-2, pif-3, pif-5 and ac81*) conserved in NALDVs.
710 All accessions for exogenous viruses used in their phylogenetic analysis were used, and
711 annotations from BlastP and HMMscan were used to identify these genes in the assembled
712 *M. hyperodae* virus genome. To include *Drosophila*-associated filamentous virus (DaFV) in
713 this phylogeny, a BlastP search was performed using the gene predictions available to
714 identify DaFV genes to include in the analysis. Genes and their accessions are detailed in
715 supplementary table 1.
716
717 Muscle v3.7 was used to align gene sequences separately, and alignments then
718 concatenated with FASconCAT-G v1.05 (Kück & Longo, 2014). The concatenated alignment
719 was trimmed with default parameters and a gap threshold of 0.6 by TrimAl v1.2. A maximum
720 likelihood phylogeny was constructed for the concatenated alignment using RAxML-NG
721 v1.0.2, with a Blosum62 model and 1000 bootstrap replicates. The phylogenetic tree was
722 drawn using Interactive Tree of Life (Letunic & Bork, 2021), rooted at the midpoint,

723 displaying bootstrap values above 50%. The scripts used for these analyses are available on
724 Zenodo at doi: 10.5281/zenodo.7939029 using a Snakemake-managed workflow.
725
726 To investigate the presence of single nucleotide polymorphisms (SNPs) in the assembled
727 virus genome, trimmed reads were mapped to the *M. hyperodae* and virus genomes
728 concatenated into a single file, using BWA v2.24. Bam files were sorted using Samtools
729 v1.10.2, duplicate reads marked with GATK v4.3.0 MarkDuplicates (McKenna et al., 2010)
730 read groups added using AddOrReplaceReadGroups from the Picard tool v2.27.5
731 (https://github.com/broadinstitute/picard), and resulting bam files indexed with samtools
732 v1.10.2. The bam file was subset to retain only the virus genome, before haplotype calling
733 was performed with GATK v4.3.0 HaplotypeCaller, with ploidy set to five (for a haploid
734 genome with five pooled samples), and variant calling then performed with GenotypeGVCFs.
735 Variants were filtered to retain only SNPs which passed GATK hard-filtering metric
736 recommendations of QD < 2, QUAL < 30, SOR > 3, FS > 60, MQ < 40, MQRankSum < −12.5,
737 ReadPosRankSum < −8, with a minor allele frequency of 0.1. The scripts for virus variant
738 calling are available on Zenodo at doi: 10.5281/zenodo.7939033 using a Snakemake-
739 managed workflow.
740
741 To investigate gene expression support for the virus infection in *M. hyperodae*, RNA-seq data
742 from a previous publication (Inwood et al., 2023) was used (NCBI SRA accession
743 PRJNA841753). Reads were trimmed and quasi-mapped as per Inwood et al. (2023) against
744 the *M. hyperodae* and virus predicted genes concatenated into one file. DESeq2 v1.34 (Love
745 et al., 2014) was used to create the DESeqDataSet (DDS) object, by importing Salmon output

746 files using tximport v1.22.0 (Soneson et al., 2016) in R v4.1.3 (R Development Core Team,
747 2020), with size factors estimated on the concatenated DDS file. The DDS file was then split
748 into separate objects for *M. hyperodae* and MhFV, and DESeq2 then used to perform a
749 likelihood ratio test (LRT) on the virus DDS file with the design ~Flowcell+Tissue, to control
750 for sequencing runs on different flow cells and test for the influence of tissue on viral gene
751 expression. Differentially expressed genes (DEGs) were identified by filtering DESeq2 results
752 with the arbitrary alpha threshold value of 0.05 for all analyses. A heatmap of viral gene
753 expression was generated using VST normalized data with pheatmap v1.0.12 (Kolde, 2019).
754 The scripts used for this analysis are available on Zenodo at doi: 10.5281/zenodo.7939037
755 using a Snakemake managed workflow.

# 756 Declarations

757

## 758 Ethics approval and consent to participate

759 Not applicable.
760

## 761 Consent for publication

762 Not applicable.
763

## 764 Availability of data and materials

765 Raw Illumina sequence data, genome assemblies and gene predictions for *Microctonus*
766 wasps are available at the National Center for Biotechnology Information (NCBI) Sequence
767 Read Archive (SRA) and Whole Genome Shotgun (WGS) databases, under accession
768 PRJNA930586, with the Hi-C scaffolded assemblies submitted for *M. aethiopoides* Irish and
769 *M. hyperodae*, with the latter having identified MhFV contigs removed prior to submission.
770 The genome and annotation for *Microctonus hyperodae* filamentous virus is available at
771 Genbank with the accession number OQ439926. Scripts used for analyses are available as
772 detailed in the methods section.
773

## Competing Interests

The authors report no competing interests.

## 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.

## Author's contributions

SNI; Data generation, bioinformatics, manuscript drafting,
JS; Data generation, bioinformatics, manuscript drafting,
JG; bioinformatics
TWH; bioinformatics
SG; Sample collection and curation
PKD; Funding, project conception, supervision, manuscript drafting.

## Acknowledgements

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

# Legends

**Figure 1** Host target and reproductive mechanisms of *Microctonus* wasps used as biocontrol agents in New Zealand.

**Figure 2** Hi-C interaction maps for A) *M. aethiopoides* Irish and B) *M. hyperodae* displaying chromosome scaffolds, and C) a Circos plot displaying chromosomal synteny between both Hi-C assemblies.

**Figure 3** A heatmap displaying the meiosis gene inventory of *Microctonus*, compared to results in other Hymenoptera, Diptera and Coleoptera from Tvedte et al (2017). Colours in heatmap indicate the absence or presence (and gene number) of each meiosis gene. Core meiosis genes, which are specific to meiosis, and asexual Hymenoptera species are indicated in bold.

**Figure 4** 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. Relative read

813 depth for Illumina and ONT MinION sequencing is indicated in blue and teal respectively,
814 calculated with a sliding window with a size of 500bp and slide of 100bp.
815
816 **Figure 5** Phylogenetic analysis of nuclear arthropod-specific large double-stranded DNA
817 viruses (NALDVs). Relationships were derived using a maximum likelihood analysis with
818 RAxML-NG, from 12 core NALDV genes, as defined by Burke et al., (2021), with a total of
819 6818 characters from concatenated amino acid sequences. Bootstrap branch support values
820 over 50% are indicated on relevant branches. Species names are abbreviated as follows; *Apis*
821 *mellifera* filamentous virus *(AmFV),* White spot syndrome virus *(WSSV), Chionoecetes opilio*
822 *bacilliform* virus *(CoBV), Culex nigripalpus* nucleopolyhedrovirus *(CnNPV), Neodiprion sertifer*
823 nucleopolyhedrovirus *(NsNPV), Cydia pomonella* granulovirus *(CpGV), Autographa*
824 *californica* multiple nucleopolyhedrovirus *(AcMNPV), Gryllus bimaculatus* nudivirus *(GbNV),*
825 *Oryctes rhinoceros* nudivirus *(OrNV), Drosophila innubila* nudivirus *(DiNV), Tipula oleracea*
826 nudivirus *(ToNV), Helicoverpa zea* nudivirus 2 *(HzNV-2), Penaeus monodon* nudivirus
827 *(PmNV), Musca* domestica salivary gland hypertrophy virus *(MdSGHV), Glossina* pallidipes
828 salivary gland hypertrophy virus *(GpSGHV), Leptopilina boulardi* filamentous virus *(LbFV),*
829 *Drosophila-associated* filamentous virus *(DaFV), and Microctonus hyperodae* filamentous
830 virus *(MhFV).*
831
832 **Supplementary Table 1** Accessions for virus genes used to construct nuclear arthropod-
833 specific large double-stranded DNA virus phylogeny.
834
835 **Supplementary Table 2** Viral hits in *Microctonus* genomes from the reciprocal BlastP
836 analysis.
837
838 **Supplementary Table 3** A comparison of MhFV virus genome characteristics to other nuclear
839 arthropod-specific large double-stranded DNA viruses.
840
841 **Supplementary Table 4** BlastP and HMMscan annotations for MhFV ORFs.
842
843 **Supplementary Table 5** Significant differentially expressed genes (DEGs) from DESeq2 MhFV
844 tissue LRT analysis.
845
846 **Supplementary Figure 1** Rudimentary estimation of *Microctonus* evolutionary origin based
847 on an ultrametric Orthofinder species tree, using a branch length scaling factor of 563.12.
848 The *Microctonus* clade is highlighted in the blue dashed box. The “Irish”, “French” and
849 “Moroccan” strains of *Microctonus aethiopoides* are Maeth IR, Maeth FR and Maeth MO
850 respectively. *Drosophila melanogaster* is the outgroup. The divergence of *Acromyrmex*
851 *echinatior* (*) and *Camponotus floridanus* (*) from Peters et al. (2017) was used to calibrate
852 the molecular clock and is indicated by an arrow. The Hymenopteptera abbreviations are as

853 follows: Nvit (*Nasonia vitripennis*), Tpre (*Trichogramma pretiosum*), Vvul (*Vespula vulgaris*),
854 Vpen (*Vespula pensylvanica*), Vger (*Vespula germanica*), Dnov (*Dufourea novaeangliae*),
855 Mrot (*Megachile rotundata*), Ccal (*Ceratina calcarata*), Amel (*Apis mellifera*), Acer (*Apis*
856 *cerana*), Bter (*Bombus terrestris*), Bimp (*Bombus impatiens*), Hsal (*Harpegnathos saltator*),
857 Cbir (*Cerapachys biroi*), Lhum (*Linepithema humile*), Fexs (*Formica exsecta*), Cflo
858 (*Camponotus floridanus*), Pbar (*Pogonomyrmex barbatus*), Cobs (*Cardiocondyla obscurior*),
859 Veme (*Vollenhovia emeryi*), Mpha (*Monomorium pharaonic*), Sinv (*Solenopsis invicta*), Waur
860 (*Wasmannia auropunctata*), Aech (*Acromyrmex echinator*) and Acep (*Atta cephalotes*).
861
862 **Supplementary Figure 2** HCR in situ hybridisation of MND1 (red) and DAPI staining (cyan) in
863 the ovaries of asexual "Irish" *M. aethiopoides,* with MND1 expression highlighted by the
864 dashed red box.
865
866 **Supplementary Figure 3** Stacked bar graphs showing the number of significant BlastP hits to
867 viral families for predicted proteins from *Microctonus* genome assemblies. Significant BlastP
868 hits were those with E-values below the E-value threshold of 1E-05. Viral hits are divided
869 into two panels, the upper containing hits to DNA viruses, and the lower to RNA viruses. *M.*
870 *aethiopoides* Irish is not shown in the upper panel due to having no DNA virus hits.
871
872 **Supplementary Figure 4** Scatterplots and boxplots of mean sequencing depth and GC
873 content of contigs in *Microctonus* genomes. Contigs with viral genes are in purple, and
874 contigs with BUSCO orthologs (for *M. aethiopoides* strains) or Hi-C scaffolds (for *M.*
875 *hyperodae*), in yellow. Reported P-values were determined using a Kruskal-Wallis test,
876 followed by a pairwise Wilcox test. The number of contigs in each group is reported in the
877 GC content boxplot.
878
879 **Supplementary Figure 5** A clustered heatmap showing expression for all MhFV genes, across
880 the pupa, head, thorax, abdomen, ovaries and venom of *M. hyperodae*, normalized by VST.
881 Genes detected as significantly differentially expressed in the tissue LRT analysis are
882 indicated in bold.
883
884

885 # References

886
887 Antipov, D., Rayko, M., Kolmogorov, M., & Pevzner, P. A. (2022). viralFlye: assembling viruses
888 and identifying their hosts from long-read metagenomics data. *Genome Biol.*, *23*(1), 1–
889 21.
890 Arakaki, N., Noda, H., & Yamagishi, K. (2000). Wolbachia-induced parthenogenesis in the egg
891 parasitoid Telenomus nawai. *Entomol. Exp. Appl.*, *96*(2), 177–184.

https://doi.org/10.1046/j.1570-7458.2000.00693.x
Barker, G. M. (2013). Biology of the introduced biocontrol agent Microctonus hyperodae (Hymenoptera: Braconidae) and its host Listronotus bonariensis (Coleoptera: Curculionidae) in Northern New Zealand. *Environ. Entomol.*, *42*(5), 902–914. https://doi.org/10.1603/EN11248
Barker, G. M., & Addison, P. J. (2006). 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.*, *99*(2), 273–287. https://doi.org/10.1093/jee/99.2.273
Barratt, B. I. P., Evans, A. A., Ferguson, C. M., Barker, G. M., McNeill, M. R., & Phillips, C. B. (1997). Laboratory Nontarget Host Range of the Introduced Parasitoids Microctonus aethiopoides and M. hyperodae (Hymenoptera: Braconidae) Compared with Field Parasitism in New Zealand. *Environ. Entomol.*, *26*(3), 694–702. https://doi.org/10.1093/ee/26.3.694
Barratt, B. I. P., Evans, A. A., Stoltz, D. B., Vinson, S. B., & Easingwood, R. (1999). Virus-like Particles in the Ovaries of Microctonus aethiopoides Loan (Hymenoptera: Braconidae), a Parasitoid of Adult Weevils (Coleoptera: Curculionidae). In *Journal of Invertebrate Pathology* (Vol. 73, Issue 2). https://doi.org/10.1006/jipa.1998.4826
Barratt, B. I. P., Murney, R., Easingwood, R., & Ward, V. K. (2006). Virus-like particles in the ovaries of Microctonus aethiopoides Loan (Hymenoptera: Braconidae): Comparison of biotypes from Morocco and Europe. *J. Invertebr. Pathol.*, *91*(1), 13–18. https://doi.org/10.1016/j.jip.2005.10.008
Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., Lee, J., Lam, E. T., Liachko, I., & Sullivan, S. T. (2017). Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.*, *49*(4), 643–650.
Burke, G. R., Hines, H. M., & Sharanowski, B. J. (2021). The presence of ancient core genes reveals endogenization from diverse viral ancestors in parasitoid wasps. *Genome Biol. Evol.* https://doi.org/10.1093/gbe/evab105
Casanovas, P., Goldson, S. L., & Tylianakis, J. M. (2018). Asymmetry in reproduction strategies drives evolution of resistance in biological control systems. *PLoS One*, *13*(12), e0207610. https://doi.org/10.1371/journal.pone.0207610
Chapman, J. A., Ho, I., Sunkara, S., Luo, S., Schroth, G. P., & Rokhsar, D. S. (2011). Meraculous: de novo genome assembly with short paired-end reads. *PLoS One*, *6*(8), e23501.
Chiu, E., Hijnen, M., Bunker, R. D., Boudes, M., Rajendran, C., Aizel, K., Oliéric, V., Schulze-Briese, C., Mitsuhashi, W., Young, V., Ward, V. K., Bergoin, M., Metcal, P., & Coulibaly, F. (2015). Structural basis for the enhancement of virulence by viral spindles and their in vivo crystallization. *Proc. Natl. Acad. Sci. U. S. A.*, *112*(13), 3973–3978. https://doi.org/10.1073/pnas.1418798112
Coffman, K. A., Hankinson, Q. M., & Burke, G. R. (2022). A viral mutualist employs posthatch transmission for vertical and horizontal spread among parasitoid wasps. *Proc. Natl. Acad. Sci.*, *119*(16), e2120048119.
Crook, N. E., Clem, R. J., & Miller, L. K. (1993). An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif. *J. Virol.*, *67*(4), 2168–2174.

https://doi.org/10.1128/jvi.67.4.2168-2174.1993
Dalla Benetta, E., Antoshechkin, I., Yang, T., Nguyen, H. Q. M., Ferree, P. M., & Akbari, O. S. (2020). Genome elimination mediated by gene expression from a selfish chromosome. *Sci. Adv.*, *6*(14), eaaz9808.
Di Giovanni, D., Lepetit, D., Guinet, B., Bennetot, B., Boulesteix, M., Couté, Y., Bouchez, O., Ravallec, M., & Varaldi, J. (2020). A behavior-manipulating virus relative as a source of adaptive genes for drosophila parasitoids. *Mol. Biol. Evol.*, *37*(10), 2791–2807. https://doi.org/10.1093/molbev/msaa030
Doherty, K. M., Sharma, S., Uzdilla, L. A., Wilson, T. M., Cui, S., Vindigni, A., & Brosh, R. M. (2005). RECQ1 Helicase Interacts with Human Mismatch Repair Factors That Regulate Genetic Recombination*[boxs]. *J. Biol. Chem.*, *280*(30), 28085–28094.
Drezen, J. M., Leobold, M., Bézier, A., Huguet, E., Volkoff, A. N., & Herniou, E. A. (2017). Endogenous viruses of parasitic wasps: variations on a common theme. *Curr. Opin. Virol.*, *25*, 41–48. https://doi.org/10.1016/j.coviro.2017.07.002
Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.*, *7*(10), e1002195.
Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.*, *16*(1), 1–14.
Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.*, *20*, 1–14.
Ferguson, C. M., Barratt, B. I. P. B. I. P., Bell, N., Goldson, S. L. L., Hardwick, S., Jackson, M., Jackson, T. A. A., Phillips, C. B. C. B., Popay, A. J. J., Rennie, G., Sinclair, S., Townsend, R., & Wilson, M. (2019). Quantifying the economic cost of invertebrate pests to New Zealand's pastoral industry. *New Zeal. J. Agric. Res.*, *62*(3), 255–315. https://doi.org/10.1080/00288233.2018.1478860
Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., & Punta, M. (2014). Pfam: The protein families database. In *Nucleic Acids Research*. https://doi.org/10.1109/TCSVT.2017.2671899
Furihata, S. X., & Kimura, M. T. (2009). Effects of Asobara japonica venom on larval survival of host and nonhost Drosophila species. *Physiol. Entomol.*, *34*(3), 292–295. https://doi.org/10.1111/j.1365-3032.2009.00676.x
Furihata, S. X., Matsumura, T., Hirata, M., Mizutani, T., Nagata, N., Kataoka, M., Katayama, Y., Omatsu, T., Matsumoto, H., & Hayakawa, Y. (2016). Characterization of venom and oviduct components of parasitoid wasp asobara japonica. *PLoS One*, *11*(7). https://doi.org/10.1371/journal.pone.0160210
Gauthier, J., Boulain, H., van Vugt, J. J. F. A., Baudry, L., Persyn, E., Aury, J. M., Noel, B., Bretaudeau, A., Legeai, F., Warris, S., Chebbi, M. A., Dubreuil, G., Duvic, B., Kremer, N., Gayral, P., Musset, K., Josse, T., Bigot, D., Bressac, C., … Drezen, J. M. (2021). Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. *Commun. Biol.*, *4*(1), 1–15. https://doi.org/10.1038/s42003-020-01623-8
Gerard, P. J., McNeill, M. R., Barratt, B. I. P., & Whiteman, S. A. (2006). Rationale for release of the Irish strain of *Microctonus aethiopoides* for biocontrol of clover root weevil. *New Zeal. Plant Prot.*, *59*, 285–289. https://doi.org/10.30843/nzpp.2006.59.4471
Gerard, P. J., Wilson, D. J., & Eden, T. M. (2011). Field release, establishment and initial

982 dispersal of Irish Microctonus aethiopoides in Sitona lepidus populations in northern
983 New Zealand pastures. *BioControl*, *56*(6), 861–870. https://doi.org/10.1007/s10526-
984 011-9367-5
985 Goldson, S. L., McNeill, M. R., & Proffitt, J. R. (2003). Negative effects of strain hybridisation
986 on the biocontrol agent Microctonus aethiopoides. *New Zeal. Plant Prot.*,
987 *56*(November), 138–142.
988 Goldson, S. L., McNeill, M. R., & Proffitt, J. R. (1993). Unexplained mortality amongst
989 parasitized Listronotus bonariensis in the presence of the parasitoid Microctonus
990 hyperodae under caging conditions. *Proc. 6th Australas. Conf. Grassl. Invertebr. Ecol.*,
991 355–362.
992 Goldson, S. L., McNeill, M. R., Proffitt, J. R., Barker, G. M., Addison, P. J., Barratt, B. I. P., &
993 Ferguson, C. M. (1993). Systematic mass rearing and release of Microctonus hyperodae
994 (Hym.: Braconidae, Euphorinae), a parasitoid of the argentine stem weevil Listronotus
995 bonariensis (Col.: Curculionidae) and records of its establishment in New Zealand.
996 *Entomophaga*, *38*(4), 527–536. https://doi.org/10.1007/BF02373087
997 Goldson, S. L., McNeill, M. R., Proffitt, J. R., Phillips, C. B., Gerard, P. J., Murray, P. J., Cane, R.
998 P., Proffitt, J. R., Phillips, C. B., Cane, R. P., & Murray, P. J. (2004). British-based search
999 for natural enemies of the clover root weevil, Sitona lepidus in Europe. *New Zeal. J.*
1000 *Zool.*, *31*(3), 233–240. https://doi.org/10.1080/03014223.2004.9518375
1001 Goldson, S. L., McNeill, M. R., Stufkens, M. W., Proffitt, J. R., Pottinger, R. P., & Farrell, J. A.
1002 (1990). Importation and quarantine of Microctonus hyperodae, a South American
1003 parasitoid of Argentine stem weevil. In *Proceedings of the Forty Third New Zealand*
1004 *Weed and Pest Control Conference* (pp. 334–338). New Zealand Weed and Pest Control
1005 Society Inc.
1006 Goldson, S. L., Proffitt, J. R., & Baird, D. B. (1998). Establishment and Phenology of the
1007 Parasitoid Microctonus hyperodae (Hymenoptera: Braconidae) in New Zealand.
1008 *Environ. Entomol.*, *27*(6), 1386–1392. https://doi.org/10.1093/ee/27.6.1386
1009 Goldson, S. L., Proffitt, J. R., McNeill, M. R., Phillips, C. B., Barlow, N. D., & Baird, D. B. (2004).
1010 Unexpected Listronotus bonariensis (Coleoptera: Curculionidae) mortality in the
1011 presence of parasitoids . *Bull. Entomol. Res.*, *94*(5), 411–417.
1012 https://doi.org/10.1079/ber2004314
1013 Goldson, S. L., Wratten, S. D., Ferguson, C. M., Gerard, P. J., Barratt, B. I. P., Hardwick, S.,
1014 McNeill, M. R., Phillips, C. B., Popay, A. J., Tylianakis, J. M., & Tomasetto, F. (2014). If
1015 and when successful classical biological control fails. *Biol. Control*, *72*, 76–79.
1016 https://doi.org/10.1016/j.biocontrol.2014.02.012
1017 Harrop, T. W. R., Le Lec, M. F., Jauregui, R., Taylor, S. E., Inwood, S. N., van Stijn, T., Henry,
1018 H., Skelly, J. G., Ganesh, S., Ashby, R. L., Jacobs, J. M. E., Goldson, S. L., & Dearden, P. K.
1019 (2020). Genetic diversity in invasive populations of argentine stem weevil associated
1020 with adaptation to biocontrol. *Insects*, *11*(7), 1–14.
1021 https://doi.org/10.3390/insects11070441
1022 Hu, J., Fan, J., Sun, Z., & Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool
1023 for long-read assembly. *Bioinformatics*.
1024 Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010).
1025 Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC*

1026 *Bioinformatics*, *11*(1), 119. https://doi.org/10.1186/1471-2105-11-119
1027 Iline, I. I., & Phillips, C. B. (2004). Allozyme markers to help define the South American origins
1028 of Microctonus hyperodae (Hymenoptera: Braconidae) established in New Zealand for
1029 biological control of Argentine stem weevil . *Bull. Entomol. Res.*, *94*(3), 229–234.
1030 https://doi.org/10.1079/ber2004303
1031 Inwood, S. N., Harrop, T. W. R. R., & Dearden, P. K. (2023). The venom composition and
1032 parthenogenesis mechanism of the parasitoid wasp Microctonus hyperodae, a declining
1033 biocontrol agent. *Insect Biochem. Mol. Biol.*, *153*, 103897.
1034 https://doi.org/10.1016/j.ibmb.2022.103897
1035 Kawato, S., Shitara, A., Wang, Y., Nozaki, R., Kondo, H., & Hirono, I. (2019). Crustacean
1036 Genome Exploration Reveals the Evolutionary Origin of White Spot Syndrome Virus. *J.*
1037 *Virol.*, *93*(3), e01144-18. https://doi.org/10.1128/jvi.01144-18
1038 Klose, R. J., Kallin, E. M., & Zhang, Y. (2006). JmjC-domain-containing proteins and histone
1039 demethylation. *Nat. Rev. Genet.*, *7*(9), 715–727. https://doi.org/10.1038/nrg1945
1040 Kolde, R. (2019). *Pretty Heatmaps* (R package version 1.0.12; pp. 1–8). https://cran.r-
1041 project.org/package=pheatmap
1042 Kolmogorov, M., Bickhart, D. M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S. B., Kuhn, K.,
1043 Yuan, J., Polevikov, E., & Smith, T. P. L. (2020). metaFlye: scalable long-read
1044 metagenome assembly using repeat graphs. *Nat. Methods*, *17*(11), 1103–1110.
1045 Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone
1046 reads using repeat graphs. *Nat. Biotechnol.*, *37*(5), 540–546.
1047 Köster, J., & Rahmann, S. (2012). Snakemake-a scalable bioinformatics workflow engine.
1048 *Bioinformatics*, *28*(19), 2520–2522. https://doi.org/10.1093/bioinformatics/bts480
1049 Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., &
1050 Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics.
1051 *Genome Res.*, *19*(9), 1639–1645.
1052 Kück, P., & Longo, G. C. (2014). FASconCAT-G: Extensive functions for multiple sequence
1053 alignment preparations concerning phylogenetic studies. *Front. Zool.*, *11*(1), 1–8.
1054 https://doi.org/10.1186/s12983-014-0081-x
1055 Lepetit, D., Gillet, B., Hughes, S., Kraaijeveld, K., & Varaldi, J. (2016). Genome sequencing of
1056 the behavior manipulating virus lbfv reveals a possible new virus family. *Genome Biol.*
1057 *Evol.*, *8*(12), 3718–3739. https://doi.org/10.1093/gbe/evw277
1058 Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic
1059 tree display and annotation. *Nucleic Acids Res.*, *49*(W1), W293–W296.
1060 https://doi.org/10.1093/nar/gkab301
1061 Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., & Henrissat, B. (2013). Expansion of
1062 the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes.
1063 *Biotechnol. Biofuels*, *6*(1), 1–14. https://doi.org/10.1186/1754-6834-6-41
1064 Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*,
1065 *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191
1066 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
1067 transform. *Bioinformatics*, *25*(14), 1754–1760.
1068 https://doi.org/10.1093/bioinformatics/btp324
1069 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., &
1070 Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*,

*25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., & Dorschner, M. O. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (80-. ).*, *326*(5950), 289–293.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, A., & Miller, L. K. (1995). The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. *J. Virol.*, *69*(2), 975–982. https://doi.org/10.1128/jvi.69.2.975-982.1995

Ma, W. J., & Schwander, T. (2017). Patterns and mechanisms in instances of endosymbiont-induced parthenogenesis. *J. Evol. Biol.*, *30*(5), 868–888. https://doi.org/10.1111/jeb.13069

Martinez, J., Fleury, F., & Varaldi, J. (2012). Heritable variation in an extended phenotype: The case of a parasitoid manipulated by a virus. *J. Evol. Biol.*, *25*(1), 54–65. https://doi.org/10.1111/j.1420-9101.2011.02405.x

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., & Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, *20*(9), 1297–1303.

McNeill, M. R., Goldson, S. L., Proffitt, J. R., Phillips, C. B., & Addison, P. J. (2002). A description of the commercial rearing and distribution of Microctonus hyperodae (Hymenoptera: Braconidae) for biological control of Listronotus bonariensis (Kuschel) (Coleoptera: Curculionidae). *Biol. Control*, *24*(2), 167–175. https://doi.org/10.1016/S1049-9644(02)00018-X

McNeill, M. R., Proffitt, J. R., Gerard, P. J., & Goldson, S. L. (2006). Collections of Microctonus aethiopoides Loan (Hymenoptera Braconidae) from Ireland. *New Zeal. Plant Prot.*, *59*, 290–296.

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H.-Y., El-Gebali, S., & Fraser, M. I. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.*, *47*(D1), D351–D360.

Oxley, P. R., Ji, L., Fetter-Pruneda, I., McKenzie, S. K., Li, C., Hu, H., Zhang, G., & Kronauer, D. J. C. (2014). The genome of the clonal raider ant Cerapachys biroi. *Curr. Biol.*, *24*(4), 451–458.

Palmer, J., & Stajich, J. (2019). *nextgenusfs/funannotate: funannotate v1.6.0*. Zenodo. https://doi.org/https://doi.org/10.5281/zenodo.3354704

Patot, S., Allemand, R., Fleury, F., & Varaldi, J. (2012). An inherited virus influences the coexistence of parasitoid species through behaviour manipulation. *Ecol. Lett.*, *15*(6), 603–610. https://doi.org/10.1111/j.1461-0248.2012.01774.x

Patot, S., Martinez, J., Allemand, R., Gandon, S., Varaldi, J., & Fleury, F. (2010). Prevalence of a virus inducing behavioural manipulation near species range border. *Mol. Ecol.*, *19*(14), 2995–3007. https://doi.org/10.1111/j.1365-294X.2010.04686.x

Pearcy, M., Hardy, O., & Aron, S. (2006). Thelytokous parthenogenesis and its consequences

on inbreeding in an ant. *Heredity (Edinb).*, *96*(5), 377–382.
Peters, R. S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., & Lanfear, R. (2017). Evolutionary history of the Hymenoptera. *Curr. Biol.*, *27*(7), 1013–1018.
Petkovic, M., Dietschy, T., Freire, R., Jiao, R., & Stagljar, I. (2005). 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.*, *118*(18), 4261–4269.
Phanis, C. G., Miller, D. P., Cassar, S. C., Tristem, M., Thiem, S. M., & O'Reilly, D. R. (1999). Identification and expression of two baculovirus gp37 genes. *J. Gen. Virol.*, *80*(7), 1823–1831. https://doi.org/10.1099/0022-1317-80-7-1823
Phillips, C. B. (1995). *Intraspecific variation in Microctonus hyperdoae and M. aethipopides (Hymenoptera: Braconidae); significance for their use of biological control agents*. Lincoln University.
Phillips, C. B., & Kean, J. M. (2017). Response of parasitoid egg load to host dynamics and implications for egg load evolution. *J Evol Biol*, *30*(7), 1313–1324. https://doi.org/10.1111/jeb.13095
Popay, A. J., McNeill, M. R., Goldson, S. L., & Ferguson, C. M. (2011). The current status of Argentine stem weevil (Listronotus bonariensis) as a pest in the North Island of New Zealand. *New Zeal. Plant Prot.*, *64*, 55–62. https://doi.org/10.30843/nzpp.2011.64.5962
R Development Core Team. (2020). A Language and Environment for Statistical Computing. *R Found. Stat. Comput.*, https://www.R-project.org. http://www.r-project.org
Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.*, *11*(1), 1432.
Rezazadeh, S. (2012). RecQ helicases; at the crossroad of genome replication, repair, and recombination. *Mol. Biol. Rep.*, *39*(4), 4527–4543.
Schurko, A. M., & Logsdon, J. M. (2008). Using a meiosis detection toolkit to investigate ancient asexual "scandals" and the evolution of sex. *BioEssays*, *30*(6), 579–589. https://doi.org/10.1002/bies.20764
Schurko, A. M., Mazur, D. J., & Logsdon Jr, J. M. (2010). Inventory and phylogenomic distribution of meiotic genes in Nasonia vitripennis and among diverse arthropods. *Insect Mol. Biol.*, *19*, 165–180.
Shen, W., & Ren, H. (2021). TaxonKit: A practical and efficient NCBI taxonomy toolkit. *J. Genet. Genomics*. https://doi.org/10.1016/j.jgg.2021.03.006
Shields, M. W., Wratten, S. D., Phillips, C. B., Van Koten, C., & Goldson, S. L. (2022). Plant-Mediated Behavioural Avoidance of a Weevil Towards Its Biological Control Agent. *Front. Plant Sci.*, *13*. https://doi.org/10.3389/fpls.2022.923237
Shields, M. W., Wratten, S. D., Van Koten, C., Phillips, C. B., Gerard, P. J., & Goldson, S. L. (2022). Behaviour drives contemporary evolution in a failing insect-parasitoid importation biological control programme. *Front. Ecol. Evol.*, *10*. https://doi.org/10.3389/FEVO.2022.923248
Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212.

https://doi.org/10.1093/bioinformatics/btv351
Simonsen, M., Mailund, T., & Pedersen, C. N. S. (2008). Rapid neighbour-joining. *Algorithms Bioinforma. 8th Int. Work. WABI 2008, Karlsruhe, Ger. Sept. 15-19, 2008. Proc. 8*, 113–122.
Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*(5), 637–644.
Stouthamer, R., Luck, R. F., & Hamilton, W. D. (1990). Antibiotics cause parthenogenetic Trichogramma (hymenoptera/trichogrammatidae) to revert to sex. *Proc. Natl. Acad. Sci. U. S. A.*, *87*(7), 2424–2427. https://doi.org/10.1073/pnas.87.7.2424
Stufkens, M. W., Farrell, J. A., & Goldson, S. L. (1987). *Establishment of Microctonus aethiopoides, a parasitoid of the sitona weevil, in New Zealand* (A. J. Popay (ed.)). Proceedings of the Forty Third New Zealand Weed and Pest Control Conference; The New Zealand Weed and Pest Control Society.
Tomasetto, F., Tylianakis, J. M., Reale, M., Wratten, S. D., & Goldson, S. L. (2017). Intensified agriculture favors evolved resistance to biological control. *Proc. Natl. Acad. Sci. U. S. A.*, *114*(15), 3885–3890. https://doi.org/10.1073/pnas.1618416114
Tvedte, E. S., Forbes, A. A., Logsdon, J. M., & Reed, F. (2017). Retention of Core Meiotic Genes Across Diverse Hymenoptera. *J. Hered.*, *108*(7), 791–806. https://doi.org/10.1093/jhered/esx062
Varaldi, J., Boulétreau, M., & Fleury, F. (2005). Cost induced by viral particles manipulating superparasitism behaviour in the parasitoid Leptopilina boulardi. *Parasitology*, *131*(2), 161–168. https://doi.org/10.1017/S0031182005007602
Varaldi, J., Petit, S., Boulétreau, M., & Fleury, F. (2006). The virus infecting the parasitoid Leptopilina boulardi exerts a specific action on superparasitism behaviour. *Parasitology*, *132*(6), 747–756. https://doi.org/10.1017/S0031182006009930
Vereijssen, J., Armstrong, K. F., Barratt, B. I. P., Crawford, A. M., Mcneill, M. R., & Goldson, S. L. (2011). Evidence for parasitoid-induced premature mortality in the Argentine stem weevil. *Physiol. Entomol.*, *36*(2), 194–199. https://doi.org/10.1111/j.1365-3032.2010.00773.x
Vink, C. J., Phillips, C. B., Mitchell, A. D., Winder, L. M., & Cane, R. P. (2003). Genetic variation in Microctonus aethiopoides (Hymenoptera: Braconidae). *Biol. Control*, *28*(2), 251–264.
Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., & Guo, H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.*, *40*(7), e49–e49.
Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.*, *20*(1), 1–13. https://doi.org/10.1186/s13059-019-1891-0
Ye, X. qian, Shi, M., Huang, J. hua, & Chen, X. xin. (2018). Parasitoid polydnaviruses and immune interaction with secondary hosts. *Dev. Comp. Immunol.*, *83*, 124–129. https://doi.org/10.1016/j.dci.2018.01.007