



**HAL**  
open science

## Functional insights from the GC-poor genomes of two aphid parasitoids, *Aphidius ervi* and *Lysiphlebus fabarum*

Alice B. Dennis, Gabriel I. Ballesteros, Stéphanie Robin, Lukas Schrader, Jens Bast, Jan Berghöfer, Leo Beukeboom, Maya Belghazi, Anthony Bretaudeau, Jan Büllesbach, et al.

### ► To cite this version:

Alice B. Dennis, Gabriel I. Ballesteros, Stéphanie Robin, Lukas Schrader, Jens Bast, et al.. Functional insights from the GC-poor genomes of two aphid parasitoids, *Aphidius ervi* and *Lysiphlebus fabarum*. 2019. hal-02788571

HAL Id: hal-02788571

<https://hal.inrae.fr/hal-02788571>

Preprint submitted on 5 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

1 Functional insights from the GC-poor genomes of two aphid  
2 parasitoids, *Aphidius ervi* and *Lysiphlebus fabarum*

3

4 Alice B. Dennis<sup>§1,2,3\*</sup>, Gabriel I. Ballesteros<sup>§4,5,6</sup>, Stéphanie Robin<sup>7,8</sup>, Lukas Schrader<sup>9</sup>, Jens  
5 Bast<sup>10,11</sup>, Jan Berghöfer<sup>9</sup>, Leo Beukeboom<sup>12</sup>, Maya Belghazi<sup>13</sup>, Anthony Bretaudeau<sup>7,8</sup>, Jan  
6 Büllsbach<sup>9</sup>, Elizabeth Cash<sup>14</sup>, Dominique Colinet<sup>15</sup>, Zoé Dumas<sup>10</sup>, Patrizia Falabella<sup>16</sup>, Jean-  
7 Luc Gatti<sup>15</sup>, Elzemie Geuverink<sup>12</sup>, Joshua D. Gibson<sup>14,17</sup>, Corinne Hertäg<sup>18,1</sup>, Stefanie  
8 Hartmann<sup>3</sup>, Emmanuelle Jacquin-Joly<sup>19</sup>, Mark Lammers<sup>9</sup>, Blas I. Lavandero<sup>6</sup>, Ina  
9 Lindenbaum<sup>9</sup>, Lauriane Massardier-Galata<sup>15</sup>, Camille Meslin<sup>19</sup>, Nicolas Montagné<sup>19</sup>, Nina  
10 Pak<sup>14</sup>, Marylène Poirié<sup>15</sup>, Rosanna Salvia<sup>16</sup>, Chris R. Smith<sup>20</sup>, Denis Tagu<sup>7</sup>, Sophie Tares<sup>15</sup>,  
11 Heiko Vogel<sup>21</sup>, Tanja Schwander<sup>10</sup>, Jean-Christophe Simon<sup>7</sup>, Christian C. Figueroa<sup>4,5</sup>,  
12 Christoph Vorburger<sup>1,2</sup>, Fabrice Legeai<sup>7,8</sup>, and Jürgen Gadau<sup>9</sup>

13

14 § Joint first authors

15 \*Author for correspondence: [alicebdennis@gmail.com](mailto:alicebdennis@gmail.com)

16

17

18

19

20

21

22

---

<sup>1</sup> Department of Aquatic Ecology, Eawag, 8600 Dübendorf, Switzerland

<sup>2</sup> Institute of Integrative Biology, ETH Zürich, 8092 Zürich, Switzerland

<sup>3</sup> Institute of Biochemistry and Biology, University of Potsdam, 14476 Potsdam, Germany

<sup>4</sup> Instituto de Ciencias Biológicas, Universidad de Talca, Talca, Chile

<sup>5</sup> Centre for Molecular and Functional Ecology in Agroecosystems, Universidad de Talca, Talca, Chile

<sup>6</sup> Laboratorio de Control Biológico, Instituto de Ciencias Biológicas, Universidad de Talca, Talca, Chile

<sup>7</sup> IGEPP, Agrocampus Ouest, INRA, Université de Rennes, 35650 Le Rheu, France

<sup>8</sup> Université de Rennes 1, INRIA, CNRS, IRISA, 35000, Rennes, France

<sup>9</sup> Institute for Evolution and Biodiversity, Universität Münster, Münster, Germany

<sup>10</sup> Department of Ecology and Evolution, Université de Lausanne, 1015 Lausanne

<sup>11</sup> Institute of Zoology, Universität zu Köln, 50674 Köln

<sup>12</sup> Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

<sup>13</sup> Aix-Marseille Univ, CNRS, INP, Inst Neurophysiopathol, PINT, PFNT, Marseille, France

<sup>14</sup> Department of Environmental Science, Policy, & Management, University of California, Berkeley, Berkeley, CA 94720, USA

<sup>15</sup> Université Côte d'Azur, INRA, CNRS, ISA, Sophia Antipolis, France

<sup>16</sup> University of Basilicata, Department of Sciences, 85100 Potenza, Italy

<sup>17</sup> Department of Biology, Georgia Southern University, Statesboro, GA 30460, USA

<sup>18</sup> D-USYS, Department of Environmental Systems Sciences, ETH Zürich, Switzerland

<sup>19</sup> INRA, Sorbonne Université, CNRS, IRD, UPEC, Université Paris Diderot, Institute of Ecology and Environmental Sciences of Paris, iEES-Paris, F-78000 Versailles, France

<sup>20</sup> Department of Biology, Earlham College, Richmond, IN USA 47374

<sup>21</sup> Max Planck Institute for Chemical Ecology, Department of Entomology, Jena, Germany

23 **Abstract**

24

25

26 **Background**

27 Parasitoid wasps have fascinating life cycles and play an important role in trophic  
28 networks, yet little is known about their genome content and function. Parasitoids that  
29 infect aphids are an important group with the potential for biocontrol, and infecting  
30 aphids requires overcoming both aphid defenses and their defensive endosymbionts.

31

32 **Results**

33 We present the *de novo* genome assemblies, detailed annotation, and comparative  
34 analysis of two closely related parasitoid wasps that target pest aphids: *Aphidius ervi*  
35 and *Lysiphlebus fabarum* (Hymenoptera: Braconidae: Aphidiinae). The genomes are  
36 small (139 and 141 Mbp), highly syntenic, and the most AT-rich reported thus far for  
37 any arthropod (GC content: 25.8% and 23.8%). This nucleotide bias is accompanied by  
38 skewed codon usage, and is stronger in genes with adult-biased expression. AT-richness  
39 may be the consequence of reduced genome size, a near absence of DNA methylation,  
40 and age-specific energy demands. We identify expansions of F-box/Leucine-rich-repeat  
41 proteins, suggesting that diversification in this gene family may be associated with their  
42 broad host range or with countering defenses from aphids' endosymbionts. The  
43 absence of some immune genes (Toll and Imd pathways) resembles similar losses in  
44 their aphid hosts, highlighting the potential impact of symbiosis on both aphids and  
45 their parasitoids.

46

47 **Conclusions**

48 These findings are of fundamental interest for insect evolution and beyond. This will  
49 provide a strong foundation for further functional studies including coevolution with  
50 respect to their hosts, the basis of successful infection, and biocontrol. Both genomes  
51 are available at <https://bipaa.genouest.org>.

52

53

54 **Keywords:** Parasitoid wasp, aphid host, *Aphidius ervi*, *Lysiphlebus fabarum*, GC content,  
55 *de novo* genome assembly, DNA methylation loss, chemosensory genes, venom  
56 proteins, Toll and Imd pathways

57

58

59

60

61

62

63

64

65

66

67

68

69

70

## 71 Background

72 Parasites are ubiquitously present across all of life (Poulin 2007; Windsor 1998). Their  
73 negative impact on host fitness can impose strong selection on hosts to resist, tolerate,  
74 or escape potential parasites. Parasitoids are a special group of parasites whose  
75 successful reproduction is fatal to the host (Godfray 1994; Quicke 2014). The  
76 overwhelming majority of parasitoid insects are hymenopterans that parasitize other  
77 terrestrial arthropods, and they are estimated to comprise up to 75% of the species-  
78 rich insect order Hymenoptera (Forbes *et al.* 2018; Godfray 1994; Heraty 2009;  
79 Pennacchio & Strand 2006). Parasitoid wasps target virtually all insects and  
80 developmental stages (eggs, larvae, pupae, and adults), including other parasitoids  
81 (Chen & van Achterberg 2018; Godfray 1994; Müller *et al.* 2004; Poelman *et al.* 2012).  
82 Parasitoid radiations appear to have coincided with those of their hosts (Peters *et al.*  
83 2017), and there is ample evidence that host-parasitoid relationships impose strong  
84 reciprocal selection, promoting a dynamic process of antagonistic coevolution (Dupas  
85 *et al.* 2003; Kraaijeveld *et al.* 1998; Vorburger & Perlman 2018).

86 Parasitoids of aphids play an economically important role in biological pest  
87 control (Boivin *et al.* 2012; Heimpel & Mills 2017), and aphid-parasitoid interactions are  
88 an excellent model to study antagonistic coevolution, specialization, and speciation  
89 (Henter & Via 1995; Herzog *et al.* 2007). While parasitoids that target aphids have  
90 evolved convergently several times, their largest radiation is found in the braconid  
91 subfamily Aphidiinae, which contains at least 400 described species across 50 genera  
92 (Chen & van Achterberg 2018; Shi & Chen 2005). As koinobiont parasitoids, their  
93 development progresses initially in still living, feeding, and developing hosts, and ends  
94 with the aphids' death and the emergence of adult parasitoids. Parasitoids increase

95 their success with a variety of strategies, including host choice (Chau & Mackauer 2000;  
96 Łukasik *et al.* 2013), altering larval development timing (Martinez *et al.* 2016), injecting  
97 venom during stinging and oviposition, and developing special cells called teratocytes  
98 (Burke & Strand 2014; Colinet *et al.* 2014; Falabella *et al.* 2003; Poirié *et al.* 2014; Strand  
99 2014). In response to strong selection imposed by parasitoids, aphids have evolved  
100 numerous defenses, including behavioral strategies (Gross 1993), immune defenses  
101 (Schmitz *et al.* 2012), and symbioses with heritable endosymbiotic bacteria whose  
102 integrated phages can produce toxins to hinder parasitoid success (Oliver *et al.* 2010;  
103 Oliver & Higashi 2018; Vorburger & Perlman 2018).

104 The parasitoid wasps *Lysiphlebus fabarum* and *Aphidius ervi* (*Braconidae*:  
105 Aphidiinae) are closely related endoparasitoids (Figure 1). In the wild both species are  
106 found infecting a wide range of aphid species although their host ranges differ, with *A.*  
107 *ervi* more specialized on aphids in the Macrosiphini tribe and *L. fabarum* on the Aphidini  
108 tribe (Kavallieratos *et al.* 2004; Monticelli *et al.* 2019). In both taxa, there is evidence  
109 that parasitoid success is hindered by the presence of defensive symbionts in the aphid  
110 haemocoel, including the bacteria *Hamiltonella*, *Regiella*, and *Serratia* (Oliver *et al.*  
111 2003; Vorburger *et al.* 2010). Studies employing experimental evolution in both species  
112 have shown that wild-caught populations can counter-adapt to cope with aphids and  
113 the defenses of their endosymbionts, and that the coevolutionary relationships  
114 between parasitoids and the aphids' symbionts likely fuel diversification of both  
115 parasitoids and their hosts (Dennis *et al.* 2017; Dion *et al.* 2011; Rouchet & Vorburger  
116 2014). While a number of parasitoid taxa are known to inject viruses and virus-like  
117 particles into their hosts, there is thus far no evidence that this occurs in parasitoids  
118 that target aphids; emerging studies have identified abundant RNA viruses in *L.*

119 *fabarum* (Lüthi *et al.* submitted; Obbard *et al.* in revision), but whether this impacts  
120 their ability to parasitize is not yet fully understood.

121         These two closely related parasitoids differ in several important life history  
122 traits, and are expected to have experienced different selective regimes as a result.  
123 *Aphidius ervi* is has successfully been introduced widely (Nearctic, Neotropics) as a  
124 biological control agent (far more than *L. fabarum*). Studies on both native and  
125 introduced populations of *A. ervi* have shown ongoing evolutionary processes with  
126 regard to host preferences, gene flow, and other life history components (Henry *et al.*  
127 2008; Hufbauer *et al.* 2004; Zepeda-Paulo *et al.* 2015; Zepeda-Paulo *et al.* 2013). *A. ervi*  
128 is known to reproduce only sexually, whereas *L. fabarum* is capable of both sexual and  
129 asexual reproduction. In fact, wild *L. fabarum* populations are more commonly  
130 composed of asexually reproducing (thelytokous) individuals (Sandrock *et al.* 2011). In  
131 asexual populations, diploid *L. fabarum* females produce diploid female offspring via  
132 central fusion automixis (Belshaw & Quicke 2003). While they are genetically  
133 differentiated, sexual and asexual populations appear to maintain gene flow and thus  
134 both reproductive modes and genome-wide heterozygosity are maintained in the  
135 species as a whole (Mateo Leach *et al.* 2009; Sandrock *et al.* 2011; Sandrock &  
136 Vorburger 2011). *Aphidius. ervi* and *L. fabarum* are also expected to have experienced  
137 different selective regimes with regard to their cuticular hydrocarbon profiles and  
138 chemosensory perception. *Lysiphlebus* target aphid species that are ant-tended, and  
139 ants are known to prevent parasitoid attacks on “their” aphids (Rasekh *et al.* 2010). To  
140 counter ant defenses, *L. fabarum* has evolved the ability to mimic the cuticular  
141 hydrocarbon profile of the aphid hosts (Liepert & Dettner 1993, 1996). With this, they  
142 are able to circumvent ant defenses and access this challenging ecological niche, from

143 which they also benefit nutritionally; they are the only parasitoid species thus far  
144 documented to behaviorally encourage aphid honeydew production and consume this  
145 high-sugar reward (Rasekh *et al.* 2010; Völkl 1992; Völkl 1997).

146 We present here the genomes of *A. ervi* and *L. fabarum*, assembled *de novo*  
147 using a hybrid sequencing approach. The two genomes are highly syntenic and strongly  
148 biased towards AT nucleotides. We have examined GC content in the context of host  
149 environment, nutrient limitation, and gene expression. By comparing these two  
150 genomes we identify key functional specificities in genes underlying venom  
151 composition, oxidative phosphorylation, cuticular hydrocarbon composition, and  
152 chemosensory perception. In both species, we identify losses in key immune genes and  
153 an apparent lack of key DNA methylation machinery. These are functionally important  
154 traits associated with success infecting aphids and the evolution of related traits across  
155 all of Hymenoptera.

156

## 157 **Results and Discussion**

### 158 ***Two de novo genome assemblies***

159 The genome assemblies for *A. ervi* and *L. fabarum* were constructed using hybrid  
160 approaches that incorporated high-coverage short read (Illumina) and long-read (Pac  
161 Bio) sequencing, but were assembled with different parameters (Supplementary Tables  
162 1, 2). This produced two high quality genome assemblies (*A. ervi* N50 = 581kb, *L.*  
163 *fabarum* N50 = 216kb) with similar total lengths (*A. ervi*: 139MB, *L. fabarum*: 141MB)  
164 but different ranges of scaffold-sizes (Table 1, Supplementary Table 3). These assembly  
165 lengths are within previous estimates of 110-180Mbp for braconids, including *A. ervi*  
166 (Ardila-Garcia *et al.* 2010; Hanrahan & Johnston 2011). Both assemblies are available



167 in NCBI (SAMN13190903-4) and can be accessed via the Bioinformatics Platform for  
 168 Agroecosystem Arthropods (BIPAA, <https://bipaa.genouest.org>), which contains the  
 169 full annotation reports, predicted genes, and can be searched via both keywords and  
 170 blast.

171 We constructed linkage groups for the *L. fabarum* scaffolds using phased SNPs  
 172 from the haploid (male) sons of a single female wasp from a sexually reproducing  
 173 population. This placed the 297 largest scaffolds (>50% of the nucleotides,  
 174 Supplementary Table 5, Supplementary Figure 1, Additional File 1) into the expected  
 175 six chromosomes (Belshaw & Quicke 2003). With this largely contiguous assembly, we  
 176 show that the two genomes are highly syntenic, with >60k links in alignments made by  
 177 NUCmer (Kurtz *et al.* 2004) and >350 large syntenic blocks that match the six *L. fabarum*  
 178 chromosomes to 28 *A. ervi* scaffolds (Supplementary Figures 2 and 3).

179

180 *Table 1: Assembly and draft annotation statistics*

	<i>A. ervi</i>	<i>L. fabarum</i>
Assembly statistics		
Total length (bp)	138,951,524	140,705,580
Longest scaffold (bp)	3,671,467	2,183,677
scaffolds	5,778	1,698
scaffolds $\geq$ 3,000 bp	1,503	1,698
N50 (bp)	581,355	216,143
GC %	25.8%	23.8%
Annotation statistics		
Exons	95,322	74,701
Introns	74,978	59,498
CDS	20,344	15,203
% genome covered by CDS	17.8%	14.9%
GC % in CDS	31.9%	29.8%
GC % of 3 <sup>rd</sup> position in CDS	15.5%	10.7%
CDS with transcriptomic support	77.8%	88.3%

181

182

183           Within the two assemblies, we used the Maker2 annotation pipeline to predict  
184 coding genes (CDS) for the two genomes, and these were functionally annotated  
185 against the NCBI *nr* database (NCBI), matches to gene ontology (GO) terms, and  
186 predictions for known protein motifs, signal peptides, and transmembrane domains  
187 (Supplemental Table 6). In *A. ervi* there were 20,344 predicted genes comprising  
188 27.8Mbp, while in *L. fabarum* there were 15,203 genes across 21.9 Mbp (Table 1).  
189 These numbers are on par with those predicted in other hymenopteran genomes  
190 (Table 2), and comparisons among taxa suggest that the lower number of predicted  
191 genes in *L. fabarum* are more likely due to their loss than to a gene gain in *A. ervi*.  
192 However, it is important to recognize that predictive annotation is imperfect and any  
193 missing genes should be specifically screened with more rigorous methods. In both  
194 species, there was high transcriptomic support for the predicted genes (77.8% in *A. ervi*  
195 and 88.3% in *L. fabarum*). The two genome annotations appear to be largely complete;  
196 at the nucleotide level, we could match 94.8% (*A. ervi*) and 76.3% (*L. fabarum*) of the  
197 1,658 core orthologous BUSCO genes for Insecta in both species (Supplementary Table  
198 4). Within the predicted genes, protein-level matches to the BUSCO genes were  
199 improved in *L. fabarum* (95.9%) and slightly lower for *A. ervi* (93.7%). These numbers  
200 suggest that low GC content did not negatively impact gene prediction (Supplementary  
201 Table 4).

202           A survey of transposable Elements (TEs) identified a similar overall number of  
203 putative TE elements in the two assemblies (*A. ervi*: 67,695 and *L. fabarum*: 60,306,  
204 Supplementary Table 7). Despite this similarity, the overall genomic coverage by TEs is  
205 larger in *L. fabarum* (41%, 58 Mbp) than in *A. ervi* (22%, 31 Mbp) and they differ in the  
206 TE classes that they contain (Supplementary Table 7, Supplementary Figures 4, 5). The

207 spread of reported TE coverage in arthropods is quite large, even among *Drosophila*  
 208 species (ca. 2.7% - 25%, *Drosophila* 12 Genomes *et al.* 2007). Within parasitoids,  
 209 reported TE content also varies, and relatively low coverage in the parasitoid  
 210 *Macrocentrus cingulum* in comparison to *Nasonia vitripennis* (24.9% vs 40.6% Yin *et al.*  
 211 2018) was attributed the smaller genome size of *M. cingulum* (127.9Mbp and  
 212 295.7Mbp, respectively, Table 3). However, the variation we observe here suggests  
 213 that differences in predicted TE content may be evolutionary quite labile, even within  
 214 closely related species with the same genome size.

215

216 *Table 2: Assembly summary statistics compared to other parasitoid genomes. All species are from the family*  
 217 *Braconidae, except for N. vitripennis (Pteromalidae). Protein counts from the NCBI genome deposition.*

Parasitoid species	Assembly	Total Length (Mbp)	Scaffold Count	Scaffold N50 (bp)	Predicted genes (CDS)	GC (%)	NCBI BioProject
<i>Aphidius ervi</i>	A. ervi_v3	139.0	5,778	581,355	20,344	25.8	<i>This paper</i>
<i>Lysiphlebus fabarum</i>	L. fabarum_v1	140.7	1,698	216,143	15,203	23.8	<i>This paper</i>
<i>Fopius arisanus</i>	ASM80636v1	153.6	1,042	51,867	18,906	39.4	PRJNA258104 (Geib <i>et al.</i> 2017)
<i>Diachasma alloeum</i>	Dall1.0	388.8	3,968	44,932	19,692	39.1	PRJNA284396 (Tvedte <i>et al.</i> 2019)
<i>Microplitis demolitor</i>	Mdem 2	241.2	1,794	27,508	18,586	33.1	PRJNA251518 (Burke <i>et al.</i> 2018)
<i>Cotesia vestalis</i>	ASM95615v1	186.1	9,156	46,055	-	30.4	PRJNA271135
<i>Macrocentrus cingulum</i>	MCINOGS1.0	127.9	12,056	65,089	11,993	35.6	PRJNA361069 (Yin <i>et al.</i> 2018)
<i>Nasonia vitripennis</i>	Nvit_2.1	295.7	6,169	18,840	24,891	40.6	PRJNA13660 (Werren <i>et al.</i> 2010)

218

### 219 *GC content*

220 The *L. fabarum* and *A. ervi* genomes are the most GC-poor of insect genomes  
 221 sequenced to date (GC content: 25.8% and 23.8% for *A. ervi* and *L. fabarum*,  
 222 respectively, Table 3, Supplementary Figure 6). This nucleotide bias is accompanied by  
 223 strong codon bias in the predicted genes, meaning that within the possible codons for

224 each amino acid, the two genomes are almost universally skewed towards the codon(s)  
225 with the lowest GC content (measured as Relative Synonymous Codon Usage, RSCU,  
226 Figure 2). These patterns are much more extreme than RSCU found in other  
227 hymenopterans, which are known to prefer codons that end in –A or –U (Behura &  
228 Severson 2013). This codon bias has functional consequences; work in other taxa has  
229 shown that codon usage is tied to both expression efficiency and mRNA stability  
230 (Barahimipour *et al.* 2015).

231 Low GC content could be a consequence of the relatively small size of these  
232 genomes. Genome size and GC content are positively correlated in a diverse set of taxa  
233 including bacteria (Almpanis *et al.* 2018; McCutcheon *et al.* 2009), plants (Šmarda *et al.*  
234 2014; Veleba *et al.* 2016), and vertebrates (Vinogradov 1998). This widespread pattern  
235 may be driven by GC-rich repetitive elements that are more abundant in larger  
236 genomes, stronger selection on thermal stability in larger genomes, or thermal stability  
237 associated with the environment (Šmarda *et al.* 2014; Vinogradov 1998). The apparent  
238 lack of DNA methylation in this system may also contribute to low GC content (see  
239 below and Bewick *et al.* 2017). Methylation is a stabilizing factor with regard to GC  
240 content (Mugal *et al.* 2015), so its absence could relax selection on GC content and  
241 allow it to decline. However, neither the absence of methylation nor codon bias are  
242 unique to these taxa, suggesting that some additional selective factors or genetic drift  
243 may have further shaped the composition of these two genomes.

244 We used two approaches to investigate whether environmental constraints  
245 could drive extremely low GC content, but found no evidence for such constraints.  
246 There is reason to expect that environment could contribute to the low GC content of  
247 these genomes; in taxa including bacteria (Foerstner *et al.* 2005) and plants (Šmarda *et*

248 *al.* 2014) the environment has been shown to influence GC content via limitation in  
249 elements including nitrogen. These two wasps parasitize aphids exclusively, and aphids  
250 themselves have relatively low genome-wide GC content. This includes the pea aphid  
251 (*Acyrtosiphon pisum*), which is a frequent host of *A. ervi* and also has notably low GC  
252 content (29.8%, Li *et al.* 2019). This is not limited to *A. pisum*, with other aphid  
253 genomes' GC content ranging between 26.8% - 30% (Additional File 2), perhaps related  
254 to their high-sugar, low-nitrogen, sap diet. One way to explore the restrictions imposed  
255 by nutrient limitation is to look at the expressed genes, since selective pressure should  
256 be higher for genes that are more highly expressed (Ran & Higgs 2010; Seward & Kelly  
257 2016). For our first test, we explored potential constraints in the most highly expressed  
258 genes in both genomes. In both species, the most highly expressed 5% of genes had  
259 higher GC content and higher nitrogen content, although the higher number of  
260 nitrogen molecules in G's and C's means that these two measures cannot be entirely  
261 disentangled (Additional File 3, Supplementary Figure 7). This is in line with  
262 observations across many taxa, and with the idea that GC-rich mRNA has increased  
263 expression via its stability and secondary structure (Kudla *et al.* 2009; Plotkin & Kudla  
264 2011). For a second approach to examining constraints, we compared codon usage  
265 between our genomes and taxa associated with this parasitoid-host-endosymbiont  
266 system (Supplementary Table 8). We found no evidence of similarity in codon usage  
267 (scaled as RSCU) nor in nitrogen content (scaled per amino acid) between parasitoids  
268 and host aphids, the primary endosymbionts *Buchnera* nor, with the secondary  
269 endosymbiont *Hamiltonella* (Supplementary Figures 8-10). Together, these tests do not  
270 support environmental constraints as the driver of low GC content in these two  
271 genomes.

272 In contrast, we did find evidence for reduced GC content in genes expressed at  
273 different parasitoid life-history stages. We found higher GC content in larvae-biased  
274 genes in *L. fabarum* (Figure 3). This was true when we compared the 10% most highly  
275 expressed genes in adults (32.6% GC) and larvae (33.2%,  $p=1.2e-116$ , Figure 3,  
276 Additional File 3), and this pattern holds even more strongly for genes that are  
277 differentially expressed between adults (upregulated in adults: 28.7% GC) and larvae  
278 (upregulated in larvae: 30.7% GC,  $p=2.2e-80$ . Note that the most highly expressed  
279 genes overlap partially with those that are differentially expressed, Additional File 3).  
280 At the same time, we found no evidence that nitrogen content differs in either of these  
281 comparisons (Figure 3). While the magnitude of these differences is not very large,  
282 subtle differences in gene content are hypothesized to be the result of selection in  
283 other systems (Acquisti *et al.* 2009). It seems plausible that GC content differences  
284 among genes expressed at different life history stages could be selected in a process  
285 analogous to the small changes in gene expression that are linked to large phenotypic  
286 differences within and between species (Romero *et al.* 2012). One explanation for  
287 lower GC content in adult-biased genes could be differences in energy demands and  
288 availability of resource across life stages. Given the extreme codon bias in these  
289 genomes (Figure 2), using codons that match this bias is expected to be more efficient  
290 and accurate, resulting in lower energy consumption and faster turnover (Chaney &  
291 Clark 2015; Galtier *et al.* 2018; Kudla *et al.* 2006; Rao *et al.* 2013). Expressing AT-rich  
292 genes is slightly more energy-efficient in itself, and this could favor otherwise neutral  
293 mutations from GC to AT (Rocha & Danchin 2002). There is good motivation for adults  
294 to have a greater demand for energy efficiency. Adult parasitoids usually feed on  
295 carbohydrate rich but protein and lipid poor resources like nectar, while performing

296 costly tasks including flying, mating, and laying eggs. Meanwhile, parasitoid larvae are  
297 feeding on their aphid host's tissue, and likely benefit further from nutrients coming  
298 from the aphids' endosymbionts, while their only task is to grow as fast as possible  
299 (Cheng *et al.* 2011; Miao *et al.* 2004; Pennacchio *et al.* 1999).

300 This supports the idea that selection at the level of gene expression is shaping  
301 the GC content of these genomes. Nonetheless, further work should more explicitly  
302 test both nutrient limitation and how selective pressures differ across life-history  
303 stages. While we do not have the power to test for GC-biased gene conversion with  
304 two taxa, the even lower third position GC content (15.5% and 10.7%, Table 1) suggests  
305 that this should be tested in relation to other parasitoids (Galtier *et al.* 2018). Further  
306 explanations to be considered include effective population size, translational efficiency,  
307 and mutational bias (Behura & Severson 2013; Bentele *et al.* 2013; Galtier *et al.* 2018).  
308 Altogether, these patterns raise important questions about how codon biases impact  
309 genome content, and whether synonymous mutations are always functionally neutral  
310 (Plotkin & Kudla 2011; Powell & Moriyama 1997).

311

### 312 ***Orphan genes in the assembly***

313 To examine genes that may underlie novel functional adaptation, we identified  
314 sequences that are unique within the predicted genes in the *A. ervi* and *L. fabarum*  
315 genomes. We defined orphan genes as predicted genes with transcriptomic support  
316 and with no identifiable homology based on searches against the NCBI *nr*, *nt*, and  
317 Swissprot databases. With this, we identified 2,568 (*A. ervi*, Additional File 4) and 968  
318 (*L. fabarum*, Additional File 5) putative orphans (Supplementary Table 9). The  
319 evolutionary origin of these orphan genes is not known (Gold *et al.* 2018; Van Oss &

320 Carvunis 2019), but their retention or evolution could be important to understanding  
321 specific functions or traits in these taxa. The higher number of orphan genes in *A. ervi*  
322 partially explains the absolute difference in the number of annotated genes between  
323 both taxa.

324

### 325 ***Gene family expansions***

326 To examine gene families that may have undergone expansions in association with  
327 functional divergence and specialization, we identified groups of orthologous genes  
328 that have increased and decreased in size in the two genomes, relative to one another.  
329 We identified these species-specific gene-family expansions using the OMA standalone  
330 package (Altenhoff *et al.* 2018). OMA predicted 8,817 OMA groups (strict 1:1 orthologs)  
331 and 8,578 HOGs (Hierarchical Ortholog Groups, Additional File 6). Putative gene-family  
332 expansions would be found in the predicted HOGs, because they are calculated to allow  
333 for >1 member per species. Among these, there were more groups in which *A. ervi*  
334 possessed more genes than *L. fabarum* (865 groups with more genes in *A. ervi*, 223  
335 with more in *L. fabarum*, Supplementary Figure 11, Additional File 6). To examine only  
336 the largest gene-family expansions, we looked further at the HOGs containing >20  
337 genes (10 HOG groups, Supplementary Figure 12). Strikingly, the four largest  
338 expansions were more abundant in *A. ervi* and were all identified as F-box proteins/  
339 Leucine-rich-repeat proteins (*LRR*, total: 232 genes in *A. ervi* and 68 in *L. fabarum*,  
340 Supplementary Figure 12, Additional File 6). This signature of expansion does not  
341 appear to be due to fragmentation in the *A. ervi* assembly: the size of scaffolds  
342 containing *LRRs* is on average larger in *A. ervi* than in *L. fabarum* (Welch two-sampled  
343 t-test,  $p=0.001$ , Supplementary Figure 13).



344           The *LRRs* are a broad class of proteins associated with protein-protein  
345 interactions, including putative venom components in these parasitoids (Colinet *et al.*  
346 2014). *LRRs* belong to a larger category of leucine rich repeat pattern recognition  
347 receptor proteins, which are an important component of innate immunity and cell-  
348 surface recognition of bacterial intruders and include toll-like receptors in insects  
349 (Soanes & Talbot 2010; Takeda & Akira 2005). While the functions of these proteins are  
350 diverse, expansion in F-box/*LRR* proteins has been shown to have specific function in  
351 immunity in parasitic insects. In the Hessian fly (*Mayetiola destructor*), fly-encoded F-  
352 box/*LRR* proteins bind with plant-encoded proteins to form a complex that blocks the  
353 plant's immune defenses against the parasitic fly (Zhao *et al.* 2015). Thus, we  
354 hypothesize that this class of proteins has expanded in these parasitoids in relation to  
355 recognizing the diverse bacterial defenses of their aphid hosts. Under this hypothesis,  
356 we argue that expansion of F-box/*LRR* proteins contributes to the broad host  
357 recognition in both species, and that their greater abundance in *A. ervi* may be  
358 associated with a recent arms race with respect to the immune defenses and protective  
359 endosymbionts of their host aphids.

360           The six largest gene families that were expanded in *L. fabarum*, relative to *A.*  
361 *ervi*, were less consistently annotated. Interestingly, they contained two different  
362 histone proteins: Histone H2B and H2A (Supplementary Figure 12). All eukaryotic  
363 genomes examined to date contain multiple histone genes for the same histone  
364 variants found in humans (e.g. 22 genes for H2B or 16 genes for H2A in humans, Singh  
365 *et al.* 2018), and it has recently been suggested that these histone variants are not  
366 functionally equivalent but rather play a role in chromatin regulation (Singh *et al.* 2018).  
367 Hence, these variants could also play a role in several *L. fabarum* specific traits,

368 including the switch from sexual to asexual reproduction (thelytoky); in mammals, sex  
369 determination has been linked to regulation via histone modification (Kuroki *et al.*  
370 2013).

371

## 372 **Venom proteins**

373 Venom injected at oviposition is crucial for successful reproduction in most parasitoid  
374 wasp species (Moreau & Asgari 2015; Poirié *et al.* 2014). The venom of *A. ervi* was  
375 previously analyzed using a combined transcriptomic and proteomic approach (Colinet  
376 *et al.* 2014), and we applied similar methods here to compare the venom composition  
377 in *L. fabarum*. The venom gland in *L. fabarum* is morphologically different from *A. ervi*  
378 (Supplementary Figure 14). A total of 35 *L. fabarum* proteins were identified as putative  
379 venom proteins using 1D gel electrophoresis and mass spectrometry, combined with  
380 transcriptomic and the genome data (Supplementary Figure 15, Additional File 7,  
381 Dennis *et al.* 2017). These putative venom proteins were identified based on predicted  
382 secretion (for complete sequences) and the absence of a match to typical cellular  
383 proteins (e.g. actin, myosin). To match the analysis between the two taxa, the previous  
384 *A. ervi* venom data (Colinet *et al.* 2014) was analyzed using the same criteria as *L.*  
385 *fabarum*. This identified 32 putative venom proteins in *A. ervi* (Additional File 7).

386         Although these two species differ in their host range (Kavallieratos *et al.* 2004),  
387 comparison of venom proteins between species revealed that more than 50% of the  
388 proteins are shared between species (Figure 4A and Additional File 7), corresponding  
389 to more than 70% of the putative function categories that were predicted (Figure 4B  
390 and Additional File 7). Among venom proteins shared between both parasitoids, a  
391 gamma glutamyl transpeptidase (GGT1) is the most abundant protein in the venom of

392 both *A. ervi* (Colinet *et al.* 2014) and *L. fabarum* (Additional File 7). This protein has  
393 been suggested to be involved in the castration of the aphid host after parasitism  
394 (Falabella *et al.* 2007). As previously reported for *A. ervi* (Colinet *et al.* 2014), a second  
395 GGT venom protein (GGT2) containing mutations in the active site was also found in  
396 the venom of *L. fabarum* (Supplementary Figure 16, 17). Phylogenetic analysis (Figure  
397 5) revealed that the *A. ervi* and *L. fabarum* GGT venom proteins occur in a single clade  
398 in which GGT1 venom proteins group separately from GGT2 venom proteins, thus  
399 suggesting that they originated from a duplication that occurred prior to the split from  
400 their most recent common ancestor. As previously shown for *A. ervi*, the GGT venom  
401 proteins of *A. ervi* and *L. fabarum* are found in one of the three clades described for  
402 the non-venomous hymenopteran GGT proteins (clade "A", Figure 5 and Colinet *et al.*  
403 2014). Within this clade, venomous and non-venomous GGT proteins had a similar exon  
404 structure, except for exon 1 that corresponds to the signal peptide only present in  
405 venomous GGT proteins (Supplementary Figure 17). *Aphidius ervi* and *L. fabarum*  
406 venomous GGT proteins thus probably result from a single imperfect duplication of the  
407 non-venomous GGT gene belonging to clade A in their common ancestor, followed by  
408 recruitment of the signal peptide coding sequence. This first imperfect duplication  
409 event would then have been followed by a second duplication of the newly recruited  
410 venomous GGT gene before the separation of both species.

411 The presence of truncated *LRR* proteins was previously reported in venom of *A.*  
412 *ervi* (Colinet *et al.* 2014) and other Braconidae (Mathé-Hubert *et al.* 2016) that likely  
413 interfere with the host immune response. Several *LRR* proteins were found in the  
414 venom of *L. fabarum* as well, however these results should be interpreted with caution  
415 since the sequences were incomplete and the presence of a signal peptide could not

416 be confirmed (Additional File 7). Moreover, these putative venom proteins were only  
417 identified from transcriptomic data of the venom apparatus and we could not find any  
418 corresponding annotated gene in the genome. This supports the idea that gene-family  
419 expansions in putative F-box/*LRR* proteins (discussed above) are not related to venom  
420 production.

421         Approximately 50% of the identified venom proteins were unique to either *A.*  
422 *ervi* or *L. fabarum*, and these could be related to their differing host ranges (Additional  
423 File 7). However, most of these proteins had no predicted function, making it difficult  
424 to hypothesize their possible role in parasitism success. Among the venom proteins  
425 with a predicted function, an apolipophorin was found in the venom of *L. fabarum* but  
426 not in *A. ervi*. Apolipophorin is an insect-specific apolipoprotein involved in lipid  
427 transport and innate immunity that is not commonly found in venoms. Among  
428 parasitoid wasps, apolipophorin has been described in the venom of the ichneumonid  
429 *Hyposoter didymator* (Dorémus *et al.* 2013) and the encyrtid *Diversinervus elegans* (Liu  
430 *et al.* 2017), but its function is yet to be deciphered. Apolipophorin is also present in  
431 low abundance in honeybee venom where it could have antibacterial activity (Kim & Jin  
432 2015; Van Vaerenbergh *et al.* 2014). Lastly, we could not find *L. fabarum* homologs for  
433 any of the three secreted cystein-rich toxin-like peptides that are highly expressed in  
434 the *A. ervi* venom apparatus (Additional File 7). However, this may not be definitive  
435 since the search for similarities in the genome is complicated by the small size of these  
436 toxin-like sequences.

437

438

439

440 *Table 3: Summary of manual curations of select gene families in the two parasitoid genomes*

Category	<i>A. ervi</i>	<i>L. fabarum</i>
Venom proteins	32	35
Desaturases*	16	15
Immune genes†	216	216
Osiris genes	21	25
Mitochondrial Oxidative Phosphorylation System (OXPHOS)**	75	74
Chemosensory group		
Chemosensory: Odorant receptors (ORs)	228	156
Chemosensory: Ionotropic chemosensory receptors (IRs)	42	40
Chemosensory: Odorant-binding proteins (OBPs)	14	14
Chemosensory: Chemosensory proteins (CSPs)	11	13
Sex determination group		
Sex determination: Core (transformer, doublesex)	4	3
Sex determination: Related genes	6	5
DNA methylation genes	2	2
<b>TOTALS</b>	<b>667</b>	<b>598</b>

441 \*Note 1: Includes genes that are partial, ambiguous, or potential pseudogenes †Note2: although the same number,  
442 the set of immune genes is not identical in the two genomes.

443

444

#### 445 **Key gene families**

446 We manually annotated more than 1,000 genes (667 for *A. ervi* and 598 for *L. fabarum*;

447 Table 3) using Apollo, hosted on the BIPAA website (Dunn *et al.* 2019;

448 <https://bipaa.genouest.org> ; Lee *et al.* 2013) to confirm and improve the results of the

449 machine annotation. This is especially important for large gene families, which are

450 usually poorly annotated by automatic prediction (Robertson *et al.* 2018); since such

451 gene families potentially underlie key adaptive differences between the two

452 parasitoids, accurate annotation is needed.

453

#### 454 **Desaturases**

455 Desaturases are an important gene family that introduce carbon-carbon double bonds

456 in fatty acyl chains in insects (Los & Murata 1998; Sperling *et al.* 2003). While these

457 function broadly across taxa, a subset of these genes (specifically acyl-CoA desaturases)  
458 have been implicated in insect chemical recognition for roles including alkene  
459 production and modification of fatty acids (Helmkampf *et al.* 2015). This gene family is  
460 particularly interesting because it has been shown that *Lysiphlebus cardui*, a close  
461 relative of *L. fabarum*, have no unsaturated cuticular hydrocarbons, just as is seen in its  
462 aphid host. This allows the parasitoid to go undetected in aphid colonies that are ant-  
463 tended and therefore better parasitize them (Liepert & Dettner 1996). We confirmed  
464 that the same is true for *L. fabarum*; its CHC profile is dominated by saturated  
465 hydrocarbons (alkanes), contains only trace alkenes, and is completely lacking dienes  
466 (Supplementary Figure 18, 20). In contrast, *A. ervi* females produce a large amount of  
467 unsaturated hydrocarbons, with a significant amount of alkenes and alkadienes in their  
468 CHC profiles (app. 70% of the CHC profile are alkenes/alkadienes, Supplementary  
469 Figure 19, 20).

470         The loss of one annotated desaturase gene in *L. fabarum* compared to *A. ervi*  
471 (Table 3) might explain these differences in the composition of their CHC profiles,  
472 especially their apparent inability to synthesize dienes. We also note there is little  
473 evidence that members of this gene family are clustered in the genome (just three and  
474 two desaturase genes in the same scaffolds of *A. ervi* and *L. fabarum*, respectively).  
475 Further investigations should verify this loss in *L. fabarum*, identify the ortholog of the  
476 missing copy in *A. ervi*, and test if this potential lost desaturase gene in *L. fabarum* is  
477 involved in the generation of unsaturated CHCs in *A. ervi*. This would determine if this  
478 loss is a key adaptation for mimicry of their aphid hosts' cuticular hydrocarbon profiles  
479 in *L. fabarum*.

480

## 481 Immune genes

482 We searched for immune genes in the two genomes based on a list of 367 immunity  
483 related genes, collected primarily from the *Drosophila* literature (Additional File 8).  
484 Using blast-based searches, 204 of these genes (59%) were found and annotated in  
485 both species. Six were present in only the *A. ervi* genome and six in only the *L. fabarum*  
486 genome. We compared these with the immune genes used to define the main  
487 *Drosophila* immune pathways (Toll, Imd, and JAK-STAT, Supplementary Table 10) and  
488 conserved in a large number of insect species (Buchon *et al.* 2014; Charroux & Royet  
489 2010; Lemaitre & Hoffman 2007). Among these genes there are several well  
490 characterized pathways. The *D. melanogaster* Toll pathway is essential for the response  
491 to fungi and Gram-positive bacteria (Valanne *et al.* 2011). It was initially identified as a  
492 developmental pathway acting via the nuclear factor kappa B (NF- $\kappa$ B). The Imd/NF-  
493 kappa-B pathway is pivotal in the humoral and epithelial immune response to Gram-  
494 negative bacteria. Signaling through *imd* (a death domain protein) ultimately activates  
495 the transcription of specific antimicrobial peptides (AMPs, Myllymäki *et al.* 2014). The  
496 JAK-STAT pathway is involved in the humoral and cellular immune response (Morin-  
497 Poulard *et al.* 2013). It is activated after a cytokine-like protein called unpaired (*upd*)  
498 binds to its receptor Domeless (Dome). Activated JAK phosphorylates STAT molecules  
499 that translocate into the nucleus, where they bind the promoters of target genes.

500 In the genome of both wasps, many genes encoding proteins of the Imd and  
501 Toll pathways were absent, such as upstream GNBPs (Gram Negative Binding Proteins)  
502 and PGRPs (Peptidoglycan Recognition Proteins) and downstream AMPs  
503 (Supplementary Table 10, Supplementary Figure 21, Additional File 8). While none of  
504 these genes were found in *L. fabarum*, one PGRP related to PGRP-SD, involved in the

505 response to Gram-positive bacteria (Bischoff *et al.* 2004), and one *defensin*-related  
506 gene were found in *A. ervi*. The *imd* gene was also absent in in both wasps; this is  
507 noteworthy because *imd* has been present in other hymenopteran genomes analyzed  
508 to date. Strikingly, all of the Imd pathway genes, including GGBP- and PGRP-encoding  
509 genes, *imd*, *FADD*, *Dredd* and *Relish* are lacking in aphid genomes (*A. pisum*, *A. gossypii*  
510 and *D. noxia*, via AphidBase (Legeai *et al.* 2010) and Gerardo *et al.* (2010)), and *imd* is  
511 absent in *A. glycines*, *M. persicae*, *M. cerisae*, *R. padi* genomes, some of which are hosts  
512 for *A. ervi* and *L. fabarum* (Kavallieratos *et al.* 2004). The lack of an Imd pathway in  
513 aphids is suggested to be an adaptation to tolerate the obligate bacterial symbiont,  
514 *Buchnera aphidicola*, as well as their facultative endosymbionts that are gram-negative  
515 gamma-proteobacteria (e.g. *Hamiltonella defensa*). These facultative symbionts exhibit  
516 defensive activities against microbial pathogens and insect parasitoids (Guo *et al.* 2017;  
517 Leclair *et al.* 2016; Oliver *et al.* 2010; Scarborough *et al.* 2005) and may at least partially  
518 compensate for the host aphids innate immune functions. Recent data also suggest  
519 that cross-talk occurs between the Imd and Toll pathways to target wider and  
520 overlapping arrays of microbes (Nishide *et al.* 2019). Whether a similar cross-talk occurs  
521 in these two Aphidiidae (*A. ervi* and *L. fabarum*) needs further study.

522 Overall, our results suggest convergent evolution of loss in immunity genes, and  
523 possibly function, between these parasitoids and their aphid hosts. One reason might  
524 be that during the early stages of development, parasitoids need host symbionts to  
525 supply their basic nutrients, and thus an immune response from the parasitoid larvae  
526 might impair this function. Alternatively, but not exclusively, mounting an immune  
527 response against bacteria by the parasitoid larvae may be energetically costly and  
528 divert resources from its development. This idea of energy conservation would be



529 especially relevant if the GC-loss discussed above is a mechanism to conserve  
530 resources. In both cases, the immune response will be costly for the parasitoid. Further  
531 work is needed to address whether other unrelated aphid parasitoids are lacking *imd*,  
532 upstream activators, and downstream effectors of the immune pathways (a  
533 preliminary blast search suggests that *imd* is present in the Aphelinidae *Aphelinus*  
534 *abdominalis*). This impaired immunity might lead to a decrease in both wasps'  
535 responses to pathogenic bacteria, or they may use other defensive components to fight  
536 bacterial infections (perhaps some in common with aphids) that await to be discovered.  
537 For example, in *L. fabarum*, recent transcriptomic work has shown that detoxifying  
538 genes may be a key component of parasitoid success (Dennis *et al.* in revision), and  
539 these could play a role in immunity.

540

#### 541 **Osiris genes**

542 The Osiris genes are an insect-specific gene family that underwent multiple tandem  
543 duplications early in insect evolution. These genes are essential for proper  
544 embryogenesis (Smoyer *et al.* 2003) and pupation (Andrade López *et al.* 2017; Schmitt-  
545 Engel *et al.* 2015), and are also tied to immune and toxin-related responses (e.g.  
546 Andrade López *et al.* 2017; Greenwood *et al.* 2017) and developmental polyphenism  
547 (Smith *et al.* 2018; Vilcinskis & Vogel 2016).

548 We found 21 and 25 putative Osiris genes in the *A. ervi* and *L. fabarum*  
549 genomes, respectively (Supplementary Tables 11, 12). In insects with well assembled  
550 genomes, there is a consistent synteny of approximately 20 Osiris genes; this cluster  
551 usually occurs in a ~150kbp stretch and gene synteny is conserved in all known  
552 Hymenoptera genomes (Supplementary Figure 22). The Osiris cluster is largely devoid

553 of non-Osiris genes in most of the Hymenoptera, but the assemblies of *A. ervi* and *L.*  
554 *fabarum* suggest that if the cluster is actually syntenic in these species, there are  
555 interspersed non-Osiris genes (those are black boxes in Supplementary Figures 23 and  
556 24).

557 In support of their role in defense (especially metabolism of xenobiotics and  
558 immunity), these genes were much more highly expressed in larvae than in adults  
559 (Supplementary Table 12). We hypothesize that their upregulation in larvae is an  
560 adaptive response to living within a host. Because of the available transcriptomic data,  
561 we could only make this comparison in *L. fabarum*. Here, 19 of the 26 annotated Osiris  
562 genes were significantly differentially expressed in larvae over adults (Supplementary  
563 Table 12, Additional File 9). In both species, transcription in adults was very low, with  
564 fewer than 10 raw reads per cDNA library sequenced, and most often less than one  
565 read per library.

566

## 567 OXPPOS

568 In most eukaryotes, mitochondria provide the majority of cellular energy (in the form  
569 of adenosine triphosphate, ATP) through the oxidative phosphorylation (OXPPOS)  
570 pathway. OXPPOS genes are an essential component of energy production, and have  
571 increased in Hymenoptera relative to other insect orders (Li *et al.* 2017). We identified  
572 69 out of 71 core OXPPOS genes in both genomes, and identified five putative  
573 duplication events that are apparently not assembly errors (Supplementary Table 13,  
574 Additional File 10). The gene sets of *A. ervi* and *L. fabarum* contained the same genes  
575 and the same genes were duplicated in each, implying duplication events occurred  
576 prior to the split from their most recent common ancestor. One of these duplicated

577 genes appears to be duplicated again in *A. ervi*, or the other copy has been lost in *L.*  
578 *fabarum*.

579

## 580 **Chemosensory genes**

581 Genes underlying chemosensory reception play important roles in parasitoid mate and  
582 host localization (Comeault *et al.* 2017; Nouhaud *et al.* 2018). Several classes of  
583 chemosensory genes were annotated separately (Table 4): odorant receptors (ORs) are  
584 known to detect volatile molecules, odorant-binding proteins (OBPs) and  
585 chemosensory proteins (CSPs) are possible carriers of chemical molecules to sensory  
586 neurons, and ionotropic receptors (IRs) are involved in both odorant and gustatory  
587 molecule reception. With these manual annotations, further studies can now be made  
588 with respect to life history characters including reproductive mode, specialization on  
589 aphid hosts, and mimicry.

590

### 591 Chemosensory: Soluble proteins (OBPs and CSPs)

592 Hymenoptera have a wide range of known OBP genes, with up to 90 in *N. vitripennis*  
593 (Vieira *et al.* 2012). However, the numbers of these genes appear to be similar across  
594 parasitic wasps, with 14 in both species studied here and 15 recently described in *D.*  
595 *alloeum* (Tvedte *et al.* 2019). Similarly, CSP numbers are in the same range within  
596 parasitic wasps (11 and 13 copies here, Table 4). Interestingly, two CSP sequences (one  
597 in *A. ervi* and one in *L. fabarum*) did not have the conserved cysteine motif,  
598 characteristic of this gene family. So although they were annotated here, further work  
599 should investigate if and how these genes function.

600

601 Chemosensory: Odorant receptors (ORs)

602 In total, we annotated 228 putative ORs in *A. ervi* and 156 in *L. fabarum* (Table 4). This  
603 is within the range of OR numbers annotated in other hymenopteran parasitoids,  
604 including: 79 in *M. cingulum* (Ahmed *et al.* 2016), 225 in *N. vitripennis* (Robertson *et*  
605 *al.* 2010), and 187 in *D. alloeum* (Tvedte *et al.* 2019). Interestingly, we annotated a  
606 larger set of ORs in *A. ervi* than in *L. fabarum*. One explanation is that *A. ervi* generally  
607 has more annotated genes than *L. fabarum*, and whatever broad pattern underlies  
608 the reduction in the gene repertoire of *L. fabarum* also affected OR genes. One  
609 functional explanations for a lower number of OR genes in *L. fabarum* is that the *A.*  
610 *ervi* strain sequenced of was derived from several field strains that parasitized  
611 different hosts on different host plants, and the ability to parasitize a broader host  
612 range could select for more OR genes (Monticelli *et al.* 2019).

613

614 Chemosensory: Ionotropic chemosensory receptors (IRs)

615 In total, we annotated 38 putative IRs in *A. ervi* and 37 in *L. fabarum* (Table 4). Three  
616 putative co-receptors (IR 8a, IR 25a and IR 76b) were annotated both species, one of  
617 which (IR 76b) was duplicated in *A. ervi*. This bring the total for the IR functional group  
618 to 42 and 40 genes for *A. ervi* and *L. fabarum*, respectively. This is within the range of  
619 IRs known from other parasitoid wasps such as *Aphidius gifuensis* (23 IRs identified in  
620 antennal transcriptome, Braconidae, Kang *et al.* 2017), *D. alloeum* (51 IRs, Braconidae,  
621 Tvedte *et al.* 2019) and *N. vitripennis* (47 IRs, Pteromalidae, Robertson *et al.* 2010). A  
622 phylogenetic analysis of these genes showed a deeply rooted expansion in the IR genes  
623 (Supplementary Figure 25). Thus, in contrast to the expansion usually observed in  
624 hymenopteran ORs compared to other insect orders, IRs have not undergone major

625 expansions in parasitic wasps, which is generally the case for a majority of insects with  
626 the exception of Blattodea (Harrison *et al.* 2018)

627

## 628 **Sex determination**

629 The core sex determination genes (*transformer*, *doublesex*) are conserved in both  
630 species (Supplementary Table 14, Additional File 11). Notably, *A. ervi* possesses a  
631 putative *transformer* duplication. This scaffold carrying the duplication (scaffold2824)  
632 is only fragmentary, but a *transformer* duplicate has also been detected in the  
633 transcriptome of a member of the *A. colemani* species complex, suggesting a conserved  
634 presence within the genus (Peters *et al.* 2017). In *A. ervi*, *transformer* appears to have  
635 an internal repeat of the CAM-domain, as is seen in the genus *Asobara* (Geuverink *et*  
636 *al.* 2018). In contrast, there is no evidence of duplication in sex determination genes in  
637 *L. fabarum*. This supports the idea that complementary sex determination (CSD) in  
638 sexually reproducing *L. fabarum* populations is based on up-stream cues that differ  
639 from those known in other CSD species (Matthey-Doret *et al.* 2019), whereas the CSD  
640 locus known from other hymenopterans locus is a paralog of *transformer* (Heimpel &  
641 de Boer 2007).

642 In addition to the core sex determination genes, we identified homologs of  
643 several genes related to sex determination (Supplementary Table 15). We identified  
644 *fruitless* in both genomes, which is associated with sex-specific behavior in taxa  
645 including *Drosophila* (Yamamoto 2008). Both genomes also have homologs of *sex-lethal*  
646 which is the main determinant of sex in *Drosophila* (Bell *et al.* 1988). *Drosophila* has  
647 two homologs of this gene, and the single version in Hymenoptera may have more in  
648 common with the non-sex-lethal copy, called *sister-of-sex-lethal*. We identified

649 homologs of the gene *CWC22*, including a duplication in *A. ervi*; this duplication is  
650 interesting because a duplicated copy of *CWC22* is the primary signal of sex  
651 determination in the house fly *Musca domestica* (Sharma *et al.* 2017). Lastly, there was  
652 a duplication of *RBP1* in both genomes. The duplication of *RBP1* is not restricted to  
653 these species, nor is the duplications of *CWC22*, which appears sporadically in  
654 Braconidae. Together, these annotations add to our growing knowledge of duplications  
655 of these genes, and provide possibilities for further examinations of the role of  
656 duplications and specialization in association with sex determination.

657

#### 658 **DNA Methylation genes**

659 DNA methyltransferase genes are thought to be responsible for the generation and  
660 maintenance of DNA methylation. In general, DNA methyltransferase 3 (*DNMT3*)  
661 introduces *de novo* DNA methylation sites and DNA methyltransferase 1 (*DNMT1*)  
662 maintains and is essential for DNA methylation (Jeltsch & Jurkowska 2014; Provataris  
663 *et al.* 2018). A third gene, *EEF1AKMT1* (formerly known as *DNMT2*), was once thought  
664 to act to methylate DNA but is now understood to methylate tRNA (Provataris *et al.*  
665 2018). In both *A. ervi* and *L. fabarum*, we successfully identified homologs *DNMT3* and  
666 *EEF1AKMT1*. In contrast, *DNMT1* was not detected in either species (Table 4,  
667 Supplementary Table 16). This adds to growing evidence that these genes are not  
668 conserved across family Braconidae, as *DNMT1* appears to be absent in several other  
669 braconid genera, including *Asobara tabida*, *A. japonica*, *Cotesia sp.*, and *F. arisanus*  
670 (Bewick *et al.* 2017; Geuverink 2017). However, *DNMT1* is present in some braconids,  
671 including *M. demolitor*, and outside of Braconidae these genes are otherwise strongly

672 conserved across insects. In contrast, DNMT3, present here, is more often lost in  
673 insects (Provataris *et al.* 2018).

674 This absence of *DNMT1* helps explain previous estimates of very low DNA  
675 methylation in *A. ervi* (0.5%, Bewick *et al.* 2017). We confirmed these low levels of  
676 methylation in *A. ervi* by mapping this previously generated bisulfite sequencing data  
677 (Bewick *et al.* 2017) to our genome assembly. We aligned >80% of their data (total  
678 94.5Mbp, 625,765 reads). The sequence coverage of this mapped data was low: only  
679 63,554 methylation-available cytosines were covered and only 1,216 were represented  
680 by two or more mapped reads. Nonetheless, of these mapped cytosines, the vast  
681 majority (63,409) were never methylated, just 143 sites were always methylated, and  
682 two were variably methylated. Methylation-available cytosine classes were roughly  
683 equally distributed among three cytosine classes (CG: 0.154%, CHG: 0.179%, and CHH:  
684 0.201%). This methylation rate is less than the 0.5% estimated by Bewick (2017) and  
685 confirms a near absence of DNA methylation in *A. ervi*. Given the parallel absence of  
686 DNMT1 in *L. fabarum*, it seems likely that both species sequenced here may have very  
687 low levels of DNA methylation, and that this is not a significant mechanism in these  
688 species.

689 This stark reduction in DNA methylation is interesting, given that epigenetic  
690 mechanisms are likely important to insect defenses, including possible responses to  
691 host endosymbionts (Huang *et al.* 2019; Vilcinskis 2016, 2017). As with the immune  
692 pathways discussed above, this could reflect a loss that is adaptive to developing within  
693 endosymbiont-protected hosts. It is also interesting that while one epigenetic  
694 mechanism seems to be absent in both *A. ervi* and *L. fabarum*, we see an increase in  
695 histone variants in *L. fabarum* (based on the OMA analysis of gene family expansion),

696 and these histones could function in gene regulation. However, whether there is a  
697 functional or causal link between these two observations is yet to be tested.

698  
699

*Table 4: Summary of annotation of putative DNA methylation genes*

Species	Gene	Scaffold	e-value ( <i>Nasonia</i> )
<i>A. ervi</i>	EEF1AKMT1 homolog	scaffold94	1.00E-66
<i>L. fabarum</i>		tig00000449	5.00E-63
<i>A. ervi</i>	DNA methyltransferase 3	scaffold45	5.00E-138
<i>L. fabarum</i>		tig00002022	9.00E-117
<i>A. ervi</i>	DNA methyltransferase 1	<i>no homolog detected</i>	
<i>L. fabarum</i>		<i>no homolog detected</i>	

700

701

## 702 Conclusions

703 These two genomes have provided insight into adaptive evolution in parasitoids that  
704 infect aphids. Both genomes are extremely GC-poor, and their extreme codon bias  
705 provides an excellent system for examining the chemical biases and selective forces  
706 that may overshadow molecular evolution in eukaryotes. We have also highlighted  
707 several groups of genes that are key to functional evolution across insects, including  
708 venom, sex determination, response to bacterial infection (F-box/*LRR* proteins), and  
709 near absence of DNA methylation. Moreover, the absence of certain immune genes  
710 (e.g. from the Imd and Toll pathways) in these two species is similar to losses in host  
711 aphids, and raises intriguing questions related to the effects of aphids' symbiosis on  
712 both aphid and parasitoid genomics.

713 Parasitoid wasps provide an excellent model for studying applied and basic  
714 biological questions, including host range (specialist vs generalist), reproductive mode  
715 (sexual vs asexual), antagonistic coevolution, genome evolution, and epigenetic  
716 regulation, to mention just a few. Our new genomic resources will open the way for a



717 broad set of future research, including work to understand host specialization, adaptive  
718 changes associated with climate, and the potential loss of diapause in *A. ervi* (Tougeron  
719 *et al.* 2019; Tougeron *et al.* 2017). Lastly, the genomes of these two non-social  
720 Hymenoptera provide a valuable comparison for understanding processes specific to  
721 social insects with complex caste structure, and are a first but essential step to better  
722 understand the genetic architecture and evolution of traits that are important for a  
723 parasitic life style and their use in biological control.

724

725

## 726 **Methods**

727 *\*More complete methods are available in the Supplementary Material*

### 728 **Insect collection and origin**

729 *Aphidius ervi*

730 *Aphidius ervi* samples used for whole-genome sequencing came from two different,  
731 sexually reproducing, isofemale lines established from parasitized aphids (recognizable  
732 as mummies) from fields of cereals and legumes in two different geographic zones in  
733 Chile: Region de Los Rios (S 39° 51', W 73° 7') and Region del Maule (S 35° 24', W 71°  
734 40'). Mummies (parasitized aphids) of *Sitobion avenae* aphids were sampled on wheat  
735 (*Triticum aestivum* L.) while mummies of *Acyrtosiphon pisum* aphids were sampled on  
736 *Pisum sativum* L. (pea aphid race). Aphid mummies were isolated in petri dishes until  
737 adult parasitoids emerged. These two parasitoid lineages were separated in two cages  
738 with hosts *ad libitum* and were propagated for approximately 75 generations under  
739 controlled conditions as described elsewhere (Ballesteros *et al.* 2017; Sepúlveda *et al.*  
740 2016). A further reduction of genetic variation was accomplished by establishing two

741 isofemale *A. ervi* lines, which were maintained as described previously and propagated  
742 for approximately 10 generations before adult parasitoids (male and female) were  
743 collected live and stored in 1.5 ml centrifuge tubes containing ethanol (95%) at -20°C.

744 *Aphidius ervi* samples used for CHC analysis (below) were purchased from Katz  
745 Biotech AG (Baruth, Germany). Species identification was confirmed with COI  
746 barcoding following Hebert *et al.* (2003). Wasps sacrificed for CHC analysis were  
747 sampled from the first generation reared in the lab on *Acyrtosiphon pisum* strain LL01  
748 (Peccoud *et al.* 2009), which were mass-reared on *Vicia faba* cv. *Dreifach Weisse*.

749

750 *Lysiphlebus fabarum*

751 *Lysiphlebus fabarum* samples used for whole-genome sequencing came from a single,  
752 asexually reproducing, isofemale line (IL-07-64). This lineage was first collected in  
753 September 2007 from Wildberg, Zürich, Switzerland as mummies of the aphid *Aphis*  
754 *fabae fabae*, collected from the host plant *Chenopodium album*. In the lab, parasitoids  
755 were reared on *A. f. fabae* raised on broad bean plants (*Vicia faba*) under controlled  
756 conditions [16 h light: 8 h dark, 20°C] until sampling in September 2013, or  
757 approximately 150 generations. Every lab generation was founded by ca. 10 individuals  
758 that were transferred to fresh host plants containing wasp-naïve aphids. Approximately  
759 700 individuals were collected for whole-genome sequencing from a single generation  
760 in December 2013 and flash frozen at -80°C. To avoid sequencing non-wasp DNA,  
761 samples were sorted over dry ice to remove any contaminating host aphid or plant  
762 material.

763 For linkage group construction, separate *L. fabarum* collections were made  
764 from a sexually reproducing lineage. Here, we collected all sons produced by a single

765 virgin female, sampled from the control lineage in a recently employed evolution  
766 experiment (H-lineage; Dennis *et al.* 2017). Wasps were stored on ethanol until RAD-  
767 seq library construction. Lastly, a third population was sampled for the proteomic  
768 analysis of the venom-apparatus (below); these females came from the genetically-  
769 diverse starting population used to found the evolution experiment of Dennis *et al.*  
770 (2017), and were sampled in December 2014.

771

## 772 DNA extraction and library preparation

### 773 *Aphidius ervi*

774 DNA was extracted from adult haploid males of *A. ervi* in seven sub-samples (ca. 120  
775 males each), reared in *S. avenae*. Total DNA was extracted using the DNEasy Plant Mini  
776 Kit (QIAGEN) following the manufacturer's instructions. DNA was quantified by  
777 spectrophotometry (Epoch Microplate Spectrophotometer, Biotek) and fluorometry  
778 (Qubit 3.0; Qubit DNA High sensitivity Assay Kit, Invitrogen), and quality was assessed  
779 using 1% agarose gel electrophoresis. DNA samples were sent on dry ice to MACROGEN  
780 (Seoul, South Korea) and were used to produce Illumina paired-end (PE) and mate-pair  
781 (MP) libraries for sequencing. A PE library was constructed from one of the seven sub-  
782 samples (120 individuals, 1 $\mu$ g DNA) sheared by ultrasonication (Covaris) company,  
783 average sheared insert size: 350bp). The remaining DNA samples were pooled (6  
784 samples, 720 individuals) and used for MP sequencing (3kb, 5kb and 8kb insert sizes),  
785 which were prepared with the Nextera mate-pair protocol (Illumina). All libraries were  
786 sequenced using an Illumina HiSeq 2000 sequencer (MACROGEN).

787 Long read PacBio (Pacific Biosciences) RS II sequencing was performed from a  
788 single DNA extraction of 270 *A. ervi* females, reared on *A. pisum*. Genomic DNA was

789 extracted using the Wizard genomic DNA purification kit (Promega) according to  
790 manufacturer instructions and quantified spectrophotometrically using a NanoDrop  
791 2000 (Thermo Scientific). Input DNA was mechanically sheared to an average size  
792 distribution of 10Kb (Covaris gTube, Kbiosciences) and the resulting library was size  
793 selected on a Blue Pippin Size Selection System (Cat #BLU0001, Sage Science) to enrich  
794 fragments > 8Kb. Quality and quantity were checked on Bioanalyzer (Agilent  
795 Technologies) and Qubit, respectively. Four SMRT RSII cells with P6 chemistry were  
796 sequenced at GenoScreen, France.

797

#### 798 *Lysiphlebus fabarum*

799 DNA was extracted from adult female *L. fabarum* in 10 sub-samples (50-100 wasps  
800 each) using the QIAmp DNA mini Kit (Qiagen) according to the manufacturer's  
801 instructions, with the inclusion of an overnight tissue digestion at 56 °C. Extracted DNA  
802 was then pooled and used to produce Illumina PE and MP, and PacBio libraries. The PE  
803 library was prepared using the Illumina Paired-End DNA protocol; the average fragment  
804 size was 180 base pair (bp). The MP library (5kb insert) was generated with the Nextera  
805 mate-pair protocol (Illumina). Both libraries were sequenced on the Illumina MiSeq in  
806 Paired-End mode at the University of Zürich.

807 Long-read libraries for PacBio RS II sequencing were produced using the DNA  
808 Template Prep Kit 2.0 (Pacific Biosciences). Input DNA was mechanically sheared to an  
809 average size distribution of 10Kb (Covaris gTube, Kbiosciences) and the resulting library  
810 was size selected on a Blue Pippin Size Selection System (Sage Science) machine to  
811 enrich fragments > 8Kb; quality and quantity were checked on the Bioanalyzer and  
812 Qubit, respectively. Ten SMRT Cells were sequenced at the University of Zürich.

813

## 814 **Genome assembly**

### 815 *Aphidius ervi*

816 Library quality was checked with FastQC ver. 0.11.3 (Andrews *et al.* 2010). Paired-end  
817 libraries were processed with Trimmomatic ver. 0.35 (Bolger *et al.*, 2014) for trimming  
818 Illumina adapters/primers, low quality bases (Q <25, 4bp window) and discarding  
819 sequences shorter than 50bp or without its mate-pair. In the case of Mate-Pair libraries,  
820 removal of improperly oriented read-pairs and removal of Nextera adapters was  
821 performed using NextClip (Leggett *et al.* 2014). Filtered PE and MP libraries were used  
822 for genome assembly with Platanus ver. 1.2.1 with default parameters (Kajitani *et al.*  
823 2014), gap closing was performed with GapCloser (Luo *et al.* 2012). Scaffolding with  
824 PacBio reads was performed using a modified version of SSPACE-LR v1.1 (Boetzer &  
825 Pirovano 2014), with the maximum link option set by `-a 250`. Finally, the gaps of this  
826 last version were filled with the Illumina reads using GapCloser.

827

### 828 *Lysiphlebus fabarum*

829 Library quality was also checked with FastQC (Andrews *et al.* 2010). Illumina reads were  
830 filtered using Trimmomatic to remove low quality sequences (Q<25, 4bp window), to  
831 trim all Illumina primers, and to discard any sequence shorter than 50bp or without its  
832 mate-pair. NextClip was used to remove all improperly oriented read pairs.

833 Raw PacBio reads were error-corrected using the quality filtered Illumina data  
834 with the program Proovread (Hackl *et al.* 2014). These error-corrected reads were then  
835 used for *de novo* assembly in the program *canu* v1.0 (Koren *et al.* 2017). Since our  
836 PacBio reads were expected to have approximately 30X coverage (based on the

837 presumed size of 128MB), *Canu* was run with the recommended settings for low  
838 coverage data (corMhapSensitivity=high corMinCoverage=2 errorRate=0.035), and  
839 with the specification that the genome is approximately 128Mbp. The resulting  
840 assembly was polished using Pilon (Walker *et al.* 2014) to correct for both single  
841 nucleotide and small indel errors, using mapping of both the MP and PE data,  
842 generated with bwa-mem (Li & Durbin 2009).

843

#### 844 Linkage map construction: *L. fabarum*

845 For linkage map construction, we followed the methodology described in Wang *et al.*  
846 (2013) and Purcell *et al.* (2014). In brief, we genotyped 124 haploid male offspring from  
847 one sexual female using ddRADseq. Whole-body DNA was high-salt extracted (Aljanabi  
848 & Martinez 1997), digested with the *EcoRI* and *MseI* restriction enzymes, and ligated  
849 with individual barcodes (Parchman *et al.* 2012; Peterson *et al.* 2012). Barcoded  
850 samples were purified and amplified with Illumina indexed primers by PCR (Peterson *et*  
851 *al.* 2012) and quality-checked on an agarose gel.

852 Pooled samples were sequenced on the Illumina HiSeq2500. Raw single-end  
853 libraries were quality filtered and de-multiplexed using the process\_radtags routine  
854 within Stacks v1.28 with default parameters (Catchen *et al.* 2011), and further filtered  
855 for possible adapter contamination using custom scripts. Genotyping was performed by  
856 mapping all samples against the *L. fabarum* draft genome assembly using bowtie2  
857 (Langmead & Salzberg 2012) with rg-id, sensitive and end-to-end options. Genotypes  
858 were extracted using samtools mpileup (Li *et al.* 2009) and bcftools (haploid option, Li  
859 2011). We filtered the resulting genotypes for a quality score >20 and removed loci  
860 with >20% missing data and/or a minor allele frequency <15% using VCFtools v0.1.12b

861 (Danecek *et al.* 2011). After filtering, 1,319 biallelic SNPs in 90 offspring remained.

862 For constructing linkage groups, we followed Gadau (2009) to account for the  
863 unknown phase of the maternal genotype. In short, we duplicated the haploid male  
864 genotypes and reversed the phase for one duplicated set and removed one of the  
865 mirror linkage group sets after mapping. We generated the map using MSTmap (Wu *et al.*  
866 *al.* 2008) on the data with following parameters: population\_type DH,  
867 distance\_function kosambi, no\_map\_dist 15.0, no\_map\_size 2, missing\_threshold 1.00,  
868 and the cut\_off\_p\_value 1e-6. The cut-off p-value was adjusted to create a linkage map  
869 of five linkage groups, however the biggest group had a gap of >70 cM, indicating a false  
870 fusion of two groups, which we split in two groups. This result corresponded to the six  
871 chromosomes previously described for *L. fabarum* (Belshaw & Quicke 2003), these  
872 were visualized with AllMaps (Tang *et al.* 2015). Initial mapping showed that 14 SNPs at  
873 one end of tig00000000 mapped to Chromosome1, while the majority of the contig  
874 (>150,000 bp) mapped to Chromosome 2. Thus, these SNPs were removed from the  
875 linkage maps, and it is advised that subsequent drafts of the *L. fabarum* genome should  
876 split this contig around position 153,900.

877

### 878 **Genome completeness and synteny**

879 Completeness of the two assemblies was assessed by identifying Benchmarking  
880 Universal Single-Copy Orthologs (BUSCOs) using the BUSCO v3.0.2 pipeline in genome  
881 mode (Simão *et al.* 2015). We identified single copy orthologs based on the  
882 Arthropoda\_db9 (1,066 genes, training species: *Nasonia vitripennis*).

883 Synteny between the two genomes was assessed using the NUCmer aligner,  
884 which is part of the MUMmer v3.23 package (Kurtz *et al.* 2004). For this, we used the

885 *L. fabarum* chromosomes as the reference, and included the scaffolds not incorporated  
886 into chromosomes (total 1,407 pieces). The *A. ervi* assembly was mapped to this using  
887 the default settings of NUCmer.

888

### 889 **Predictive gene annotation**

890 For both assembled genomes, gene predictions were generated using MAKER2 (Holt &  
891 Yandell 2011). Within MAKER2, predictive training was performed in a three step  
892 process. A first set of genes was predicted by similarity to known proteins or contigs  
893 from RNAseq in the same species (described below). This gene set was used thereafter  
894 for training both Augustus (Keller *et al.* 2011) and SNAP (Korf 2004), in two steps, with  
895 the results of the first training re-used to train the software in the second round.  
896 Transcriptomic evidence was provided separately for each species. For *A. ervi*, six  
897 separate *de novo* transcriptome assemblies from Trinity (Grabherr *et al.* 2011) were  
898 constructed, one each for the adults reared on different hosts (NCBI PRJNA377544,  
899 Ballesteros *et al.* 2017). For each transcript, we only included variants based on filtering  
900 with RSEM v 1.2.21 using the option `-fpm_cutoff 1.0, --isopct_cutoff=15.00`. This  
901 resulted in 452,783 transcripts. For *L. fabarum*, we utilized a joint transcriptome, built  
902 using RNAseq data (NCBI PRJNA290156) collected from adults (Dennis *et al.* 2017) and  
903 4-5 day old larvae (Dennis *et al.* in review). Peptide evidence came from the  
904 Hymenoptera genomes database (<http://hymenopteragenome.org>, *Acromyrmex*  
905 *echinator* v3.8, *Apis mellifera* v3.2, *Nasonia vitripennis* v1.2), from the Bioinformatics  
906 Platform of Agroecosystems Arthropod database (<https://bipaa.genouest.org>,  
907 *Hyposoter didymator* v1.0), and *Drosophila melanogaster* (<http://flybase.org>, v6.13), and  
908 SwissProt (October 2016) databases. Summary statistics were generated with GAG



909 (Hall *et al.* 2014). Transcriptomic support for the predicted genes was estimated by  
910 mapping available transcriptomic data (same as above) to the respective genomes  
911 using STAR (Dobin *et al.* 2013) in the “quantMode”.

912

### 913 **Functional annotation**

914 The putative functions of the proteins predicted by the above pipeline were identified  
915 based on blastp (v2.5.0) matches against Genbank *nr* (non-redundant GenBank CDS  
916 translations+PDB+SwissProt+PIR+PRF) release 12/2016 and interproscan v5 against  
917 Interpro (1.21.2017). GO terms associations were collected from blast *nr* and  
918 interproscan results with blast2GO (v2.2). Finally, transmembrane domains were  
919 identified with Hidden Markov Models (HMM) in tmhmm v2.0c, and peptide signals  
920 with signalP (euk v4.1, Emanuelsson *et al.* 2007; Nielsen 2017).

921

### 922 **Transposable elements**

923 Transposable elements (TE) were predicted using the REPET pipeline (Flutre *et al.*  
924 2011), combining *de novo* and homology-based annotations. *De novo* prediction of TEs  
925 was restricted to scaffolds larger than the scaffold N50 for each species. Within these,  
926 repetitive elements were identified using a blast-based alignment of each genome to  
927 itself followed by clustering with Recon (Bao & Eddy 2002), Grouper (Quesneville *et al.*  
928 2005) and Piler (Edgar & Myers 2005). For each cluster, a consensus sequence was  
929 generated by multiple alignment of all clustered elements with MAP (Huang 1994). The  
930 resulting consensus was then scanned for conserved structural features or homology  
931 to nucleotide and amino acid sequences from known TEs (RepBase 20.05, Bao *et al.*  
932 2015; Jurka 1998) using BLASTER (tblastx, blastx, Flutre *et al.* 2011) or HMM profiles of

933 repetitive elements (Pfam database 27.0) using hmmer3 (Mistry *et al.* 2013). Based on  
934 identified features, repeats were classified using Wicker's TE classification as  
935 implemented in the PASTEclassifier (Hoede *et al.* 2014). The resulting *de novo* TE library  
936 for the genome was then filtered to retain only the elements with at least one perfect  
937 match in the genome. Subsequently, all TEs in the genomes were annotated with  
938 REPET's TE annotation pipeline. Reference TE sequences were aligned to the genome  
939 using BLASTER, Repeat Masker (Smit *et al.* 2013-2015) and CENSOR (Kohany *et al.*  
940 2006). The resulting HSPs were filtered using an empirical statistical filter implemented  
941 in REPET (Flutre *et al.* 2011) and combined using MATCHER (Quesneville *et al.* 2005).  
942 Short repeats were identified using TRF (Benson 1999) and Mreps (Kolpakov *et al.*  
943 2003). Elements in genomic sequences with homology with known rebase elements  
944 (RepBase 20.05) were identified with BLASTER (blastx, tblastx) and curated by  
945 MATCHER. Finally, redundant TEs and spurious SSR annotations were filtered and  
946 separate annotations for the same TE locus were combined using REPET's "long join  
947 procedure".

948

#### 949 **GC content and codon usage**

950 We examined several measures of nucleotide composition, at both the nucleotide and  
951 protein level. Whole genome GC content was calculated by totaling the numbers of A,  
952 C, T, and G in the entire assembly. In the predicted coding sequences, this was also  
953 calculated separately for each predicted gene and third position GC composition was  
954 calculated separately in the predicted coding sequences. In all cases, this was done with  
955 the sscu package in R (Sun 2016). Relative Synonymous Codon Usage (RSCU) was  
956 extracted from the entire CDS using the seqinR package in R (Charif & Lobry 2007), and

957 visualized with a PCA (R packages factoextra, reshape, and ggplot2, Kassambara &  
958 Mundt 2016; Wickham 2007, 2009). To examine GC content in coding genes of other  
959 insects, we downloaded the 118 available CDS in the RefSeq database of NCBI (date:  
960 October 2018) and again calculated per-gene GC content.

961 To examine the GC content of life-stage biased transcripts, we compared GC  
962 content in the genes that are significantly (FDR < 0.05) differentially expressed between  
963 previously generated transcriptomes from adult (Dennis *et al.* 2017) and larval (Dennis  
964 *et al.* in revision) *L. fabarum*, as well in the 10% most highly expressed genes in adults  
965 and larvae.

966

#### 967 **Orphan genes**

968 We identified orphan genes as those for which we could not find orthologs in any other  
969 sequenced genomes. To do this, we first used OrthoFinder (Emms & Kelly 2015) to  
970 generate clusters of orthologous and paralogous genes among the predicted genes  
971 (CDS) from the genomes of *A. ervi* and *L. fabarum*, as well as five other sequenced  
972 parasitoids (*Diachasma alloeum*, *Fopius arisanus*, *Macrocentrus cingulum*, *Microplitis*  
973 *demolitor* and *Nasonia vitripennis*). OrthoFinder produces a set of genes that were not  
974 assigned to any orthogroup. We identified species specific genes, which we are calling  
975 orphan genes, by removing all genes that had hits to any other genes in the *nt*, *nr*, and  
976 *swissprot* NCBI database (June 2019). Within these putative orphans, we only retained  
977 those with transcriptomic support.

978

979

980

## 981 **Gene family expansions**

982 We examined gene families that have expanded and contracted in *A. ervi* and *L.*  
983 *fabarum* relative to one another using the OMA standalone package (v2.2.0, default  
984 values, Altenhoff *et al.* 2018). OMA was used to compute orthologs (OMA groups) and  
985 Hierarchical Orthologous Groups (HOGs) for the predicted proteins of *L. fabarum* and  
986 *A. ervi*: 15,203 and 20,344, respectively. While OMA groups consist of strict 1:1  
987 orthologs between OGS1 and OGS3, HOGs contain all orthologs and paralogs of a given  
988 predicted gene family. HOGs were parsed with a custom Perl script to identify all gene  
989 families in which one of the wasp species contained more members than the other. We  
990 focused on only the groups that contained more than 20 genes (ten groups,  
991 Supplementary Figure 12). These were identified by blastx against the *nr* database in  
992 NCBI.

993

## 994 **Venom proteins**

995 The *L. fabarum* venom proteomic analysis was performed from 10 extracted venom  
996 glands (Supplementary Figure 14). The 16 most visible bands in 1D gel electrophoresis  
997 were cut, digested with trypsin and analyzed by mass spectrometry. All raw data files  
998 generated by mass spectrometry were processed to generate mgf files and searched  
999 against: (i) the *L. fabarum* proteome predicted from the genome (*L. fabarum*  
1000 annotation v1.0 proteins) and (ii) the *L. fabarum de novo* transcriptome (Dennis *et al.*  
1001 2017) using the MASCOT software v2.3 (Perkins *et al.* 1999). The mass spectrometry  
1002 proteomics data have been deposited to the ProteomeXchange Consortium  
1003 (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository  
1004 (Hanrahan & Johnston 2011), with the ID PXD015758.

1005           Sequence annotation was performed based on blast similarity searches. Signal  
1006 peptide prediction was performed with SignalP (Emanuelsson *et al.* 2007; Nielsen  
1007 2017). Searches for protein domains was performed with PfamScan (Finn *et al.* 2013)  
1008 and venom protein genes were identified using the blast tools in Apollo (Dunn *et al.*  
1009 2019; Lee *et al.* 2013). Multiple amino acid sequence alignments were made with  
1010 MUSCLE (Edgar 2004a, b). Phylogenetic analysis was performed using maximum  
1011 likelihood (ML) with PhyML 3.0 (Guindon *et al.* 2010). SMS was used to select the best-  
1012 fit model of amino acid substitution for ML phylogeny (Lefort *et al.* 2017).

1013

#### 1014 **Manual gene curation**

1015 The two genome assemblies were manually curated for a number of gene families of  
1016 interest. This improved their structural and functional annotation for more in-depth  
1017 analysis. Manual curation, performed in Apollo included the inspection of stop/start  
1018 codons, duplications (both true and erroneous), transcriptomic support, and  
1019 concordance with the predicted gene models.

1020

#### 1021 ***Desaturases***

1022 Desaturase genes in both genomes were automatically identified and annotated with  
1023 GeMoMa (Keilwagen *et al.* 2016) using desaturase gene annotations from *Diachasma*  
1024 *alloeum*, *Fopius arisanus*, and *Microplitis demolitor*, retrieved from NCBI's protein  
1025 database as queries (retrieved May 2017). Additionally, all desaturase genes were  
1026 manually inspected.

1027           To measure the production of desaturases in *A. ervi*, wasps were freeze-killed  
1028 and stored separately by sex at - 20 °C. For CHC extraction, single individuals were

1029 covered with 50  $\mu$ l of MS pure hexane (UniSolv) in 2 ml GC vials (Agilent Technologies,) 1030 and swirled for 10 minutes on a Thermo-shaker (IKA KS 130 Basic, Staufen). The hexane 1031 extracts were then transferred to a fresh conical 250  $\mu$ l GC insert (Agilent 1032 Technologies), where the hexane was completely evaporated under a constant flow of 1033 CO<sub>2</sub>. The dried extract was then resuspended in 5  $\mu$ l of a hexane solution containing 1034 7.5 ng/ $\mu$ l of n-dodecane (EMD Millipore Corp.) as an internal standard. 3  $\mu$ l of the 1035 extract were then injected into a GC-QQQ Triple Quad (GC: 7890B, Triple Quad: 7010B, 1036 Agilent) with a PAL Autosampler system operating in electron impact ionization mode. 1037 The split/splitless injector was operated at 300 °C in Pulsed splitless mode at 20 psi until 1038 0.75 min with the Purge Flow to Split Vent set at 50 mL/min at 0.9 min. Separation of 1039 compounds was performed on a 30 m x 0.25 mm ID x 0.25  $\mu$ m HP-1 1040 Dimethylpolysiloxane column (Agilent) with a temperature program starting from 60 1041 °C, held for 2 min, and increasing by 50 °C per min to 200 °C, held for 1 min, followed 1042 by an increase of 8 °C per min to 250 °C, held again for 1 min, and finally 4 °C per min 1043 to 320 °C, held for 10 min. Post Run was set to 325 °C for 5 min. Helium served as carrier 1044 gas with a constant flow of 1.2 ml per min and a pressure of 10.42 psi. Initially CHC 1045 peaks were identified and the chromatogram was generated using the Qualitative 1046 Analysis Navigator of the MassHunter Workstation Software (vB.08.00 / Build 1047 8.0.8208.0, Agilent). CHC quantification was performed using the Quantitative Analysis 1048 MassHunter Workstation Software (vB.09.00 / Build 9.0.647.0, Agilent). Peaks were 1049 quantified using their diagnostic (or the neighboring most abundant) ion as quantifier 1050 and several characteristic ions in their mass spectra as qualifiers to allow for 1051 unambiguous detection by the quantification software. The pre-defined integrator 1052 Agile 2 was used for the peak integration algorithm to allow for maximum flexibility. All

1053 peaks were then additionally checked for correct integration and quantification, and,  
1054 where necessary, re-integrated manually. Percentages were based on the respective  
1055 averages of four individual female CHC extracts.

1056

### 1057 **Immune genes**

1058 The list of immune genes to be searched against the *A. ervi* and *L. fabarum* genomes  
1059 was established based on *Drosophila melanogaster* lists from the Lemaitre laboratory  
1060 ([lemaitrelab.epfl.ch/fr/ressources](http://lemaitrelab.epfl.ch/fr/ressources), adapted from De Gregorio *et al.* 2001; De Gregorio  
1061 *et al.* 2002) and from the interactive fly web site  
1062 ([www.sdbonline.org/sites/fly/aigfam/immune.htm](http://www.sdbonline.org/sites/fly/aigfam/immune.htm) and Buchon *et al.* 2014). Each *D.*  
1063 *melanogaster* protein sequence was used in blast similarity searches against the two  
1064 predicted wasp proteomes. The best match was retained, and its protein sequence was  
1065 used to perform a new blast search using the NCBI non-redundant protein sequence  
1066 database to confirm the similarity with the *D. melanogaster* sequence. When both  
1067 results were concordant, the retained sequence was then searched for in *Nasonia*  
1068 *vitripennis* and *Apis mellifera* proteomes to identify homologous genes in these species.

1069

### 1070 **Osiris genes**

1071 Osiris gene orthologs were determined with a two-part approach: candidate gene  
1072 categorization followed by phylogenetic clustering. Candidate Osiris genes were  
1073 generated using HMM (with hmmer v3.1b2, Wheeler & Eddy 2013) and local alignment  
1074 searching (blast, Altschul *et al.* 1990). A custom HMM was derived using all 24 well  
1075 annotated and curated Osiris genes of *Drosophila melanogaster*. Next, an HMM search  
1076 was performed on the *A. ervi* and *L. fabarum* proteomes, extracting all protein models

1077 with  $P < 0.05$ . Similarly, all *D. melanogaster* Osiris orthologs were searched in the  
1078 annotated proteomes of *A. ervi* and *L. fabarum* using protein BLAST ( $e < 0.05$ ). The top  
1079 BLAST hit for each ortholog was then searched within each parasitoid genome for  
1080 additional paralogs ( $e < 0.001$ ). All unique candidates from the above approaches were  
1081 then aligned using MAFFT (Kato & Standley 2013), and an approximate maximum-  
1082 likelihood phylogeny was constructed using FastTree (Price *et al.* 2009) via the CIPRES  
1083 science gateway of Xsede (Miller *et al.* 2015). The species used were: the fruit fly (*D.*  
1084 *melanogaster*), the tobacco hornworm moth (*Manduca sexta*), the silkworm moth  
1085 (*Bombyx mori*), the flour beetle (*Tribolium castaneum*), the jewel wasp (*Nasonia*  
1086 *vitripennis*), the honeybee (*Apis mellifera*), the buff tail bumble bee (*Bombus terrestris*),  
1087 the red harvester ant (*Pogonomyrmex barbatus*), the Florida carpenter ant  
1088 (*Camponotus floridanus*), and Jerdon's jumping ant (*Harpegnathos saltator*).

1089

## 1090 OXPPOS

1091 Genes involved in the oxidative phosphorylation pathway (OXPHOS) were identified in  
1092 several steps. Initial matches were obtained using the nuclear-encoded OXPPOS  
1093 proteins from *Nasonia vitripennis* (Gibson *et al.* 2010; J. D. Gibson unpublished) and  
1094 *Drosophila melanogaster* (downloaded from [www.mitocomp.uniba.it](http://www.mitocomp.uniba.it): Porcelli *et al.*  
1095 2007). These two protein sets were used as queries to search the protein models  
1096 predicted for *A. ervi* and *L. fabarum* (blastp, Altschul *et al.* 1997). Here, preference was  
1097 given to matches to *N. vitripennis*. Next, genes from the *N. vitripennis* and *D.*  
1098 *melanogaster* reference set that did not have a match in the predicted proteins were  
1099 used as queries to search the genome-assembly (blastn), in case they were not in the  
1100 predicted gene models. Gene models for all matches were then built up manually,



1101 based on concurrent evidence from the matches in both *A. ervi* and *L. fabarum* and  
1102 their available expression evidence. The resulting protein models were aligned to one  
1103 another and to *N. vitripennis* using MAFFT (Katoh & Standley 2013) to identify missing  
1104 or extraneous sections. These results were used as queries to search the *N. vitripennis*  
1105 proteins to ensure that all matches are reciprocal-best-blast-hits. Gene naming was  
1106 assigned based on the existing *N. vitripennis* nomenclature. Potential duplicates were  
1107 flagged based on blast-matches back to *N. vitripennis* (Additional Data 10).

1108

### 1109 **Olfactory genes**

#### 1110 Odorant-binding proteins (OBPs) and chemosensory Proteins (CSPs)

1111 To identify OBPs based on homology to known sequences, we retrieved 60 OBP amino  
1112 acid sequences from other Braconidae (namely *Fopius arisanus* and *Microplitis*  
1113 *demolitor*) from GenBank. To this, we added seven OBPs found in a previous  
1114 transcriptome of *A. ervi* (Patrizia Falabella, unpublished, EBI SRI Accessions:  
1115 ERS3933807- ERS3933809). To identify CSPs, we used CSP amino acid sequences from  
1116 more Hymenoptera species (*Apis mellifera*, *Nasonia vitripennis*, *Fopius arisanus* and  
1117 *Microplitis demolitor*). These sets were used as query against *A. ervi* and *L. fabarum*  
1118 genomes using tblastn (e-value cutoff 10e-3 for OBPs and 10e-2 for CSPs). Genomic  
1119 scaffolds that presented a hit with at least one of the query sequences were selected.  
1120 To identify precise intron/exon boundaries, the Braconidae OBP and CSP amino acid  
1121 sequences were then aligned on these scaffolds with Scipio (Keller *et al.* 2008) and  
1122 Exonerate (Slater & Birney 2005). These alignments were used to generate gene  
1123 models in Apollo. Gene models were manually curated based on homology with other  
1124 Hymenoptera OBP and CSP genes and on RNAseq data, when available. Lastly, the

1125 deduced amino acid sequences of *A. ervi* and *L. fabarum* OBP and CSP candidates were  
1126 then used as query for another tblastn search against the genomes in an iterative  
1127 process to identify any additional OBPs. Since both OBPs and CSPs are secreted  
1128 proteins, the occurrence of a signal peptide was verified using SignalP (Emanuelsson *et*  
1129 *al.* 2007; Nielsen 2017).

1130

### 1131 Odorant receptors (ORs)

1132 ORs were annotated using available OR gene models from *Diachasma alloeum*, *Fopius*  
1133 *arisanus*, and *Microplitis demolitor* retrieved from NCBI's protein database (retrieved  
1134 May 2017). Preliminary OR genes models for *A. ervi* and *L. fabarum* were predicted with  
1135 exonerate (v2.4.0), GeMoMa (v1.4, Keilwagen 2016), and combined with Evidence  
1136 Modeler (v.1.1.1, Haas *et al.* 2008). These preliminary models were subsequently  
1137 screened for the 7tm\_6 protein domain (with PfamScan v1.5) and manually curated in  
1138 WebApollo2.

1139 In an iterative approach, we annotated the IRs using known IR sequences from  
1140 *Apis mellifera*, *Drosophila melanogaster*, *Microplitis demolitor* and *Nasonia vitripennis*  
1141 as queries to identify IRs in the genomes of *A. ervi* and *L. fabarum*. The hymenopteran  
1142 IR sequences served as input for the prediction of initial gene model with Exonerate  
1143 (Slater & Birney 2005) and GeMoMa (Keilwagen *et al.* 2016). Then, we inspected and  
1144 edited homologous gene models from each tool in the Apollo genome browser to  
1145 adjust for proper splice sites, start and stop codons in agreement with spliced RNA-Seq  
1146 reads. After a first round of prediction, we repeated the whole process and provided  
1147 the amino acid sequences of curated IR genes as queries for another round of  
1148 predictions to identify any remaining paralogous IRs.

1149 Multiple sequence alignments of the IRs were computed with hmalign (Eddy  
1150 1998) using a custom IR HMM to guide the alignments (Harrison *et al.* 2018). Gene  
1151 trees were generated with FastTree v2 (Price *et al.* 2010) using the pseudocount option  
1152 and further parameters for the reconstruction of an exhaustive, accurate tree (options:  
1153 -pseudo -spr 4 -mlacc 2 -slowni). Resulting trees were visualized with iTOL v4 (Letunic  
1154 & Bork 2019), well supported IR clusters and expansions were highlighted by color  
1155 (branch support > 0.9).

1156

### 1157 ***Sex Determination***

1158 Ortholog searches were performed with tblastn (Altschul *et al.* 1997) against the  
1159 genomic scaffolds. Hits with an e-value smaller than 1e-20 were assessed, apart from  
1160 *transformer* and *doublesex* where any hit was surveyed. Doublesex, Transformer-2 and  
1161 Transformer peptide sequences of *Asobara tabida* (NCBI accessions MF074326-  
1162 MF074334) were used as queries for the core sex determination genes. This braconid  
1163 species is the closest relative whose sex determination mechanism has been examined  
1164 (Geuverink *et al.*, 2018). The putative *transformerB* sequence of *A. ervi* was blasted for  
1165 verification against the transcriptome of *Aphidius colemani* (Peters *et al.* 2017) and a  
1166 highly conserved fragment was detected (GBVE01021531). Peptide sequences of sex  
1167 determination related genes to use as queries were taken from *Nasonia vitripennis*:  
1168 Fruitless (NP\_001157594), Sex-Lethal homolog (XP\_016836645), pre-mRNA-splicing  
1169 factor *CWC22* homolog (XP\_001601117) and RNA-binding protein 1-like  
1170 (XP\_008202465). Hidden Markov models were not used as gene models because the  
1171 ensuing peptide predictions did not contain all putative homologs (e.g. *transformerB* in  
1172 *A. ervi*) due to fragmentation of the scaffolds containing the candidate genes.

1173

1174 ***DNA methylation genes***

1175 The genomes were searched with tblastn (Altschul *et al.* 1997) for the presence of  
1176 potential DNA methyltransferase genes using peptide sequences from *Apis mellifera*  
1177 and *N. vitripennis* as queries. These species differ in their copy number of *DNMT1*, with  
1178 two copies (NP\_001164522, XP\_006562865) in the honeybee *A. mellifera* (Wang *et al.*  
1179 2006) and three copies (NP\_001164521, XP\_008217946, XP\_001607336) in the wasp  
1180 *N. vitripennis* (Werren *et al.* 2010). *DNMT2*, currently characterized as EEF1AKMT1  
1181 (EEF1A Lysine Methyltransferase 1), has become redundant in the list of DNA  
1182 methyltransferase genes as it methylates tRNA instead, but was surveyed here as a  
1183 positive control (*N. vitripennis* NP\_001123319, *A. mellifera* XP\_003251471). *DNMT3*  
1184 peptide sequences from *N. vitripennis* (XP\_001599223) and from *A. mellifera*  
1185 (NP\_001177350) were used as queries for this gene. Low levels of methylation were  
1186 confirmed by mapping the whole genome bisulfite sequencing data generated by  
1187 Bewick *et al.* (2017) back to the *A. ervi* genome assembly.

1188

1189

1190

1191

1192

1193

1194

1195

1196

1197 **List of abbreviations**

1198 **A, T, C, G, and U:** Adenine, Thymine , Cytosine, Guanine, and Uracile, nucleotides

1199 **bp:** Base Pair

1200 **BIPAA:** Bioinformatics Platform for Agroecosystem Arthropods ([bipaa.genouest.org](http://bipaa.genouest.org))

1201 **BUSCO:** Benchmarking Universal Single-Copy Orthologs

1202 **CDS:** Predicted Coding Sequence

1203 **CSD:** Complementary Sex Determination

1204 **CHC:** Cuticular Hydrocarbons

1205 **DNMT:** DNA Methyltransferase genes

1206 **CSP:** Chemosensory Protein

1207 **GO:** Gene Ontology

1208 **HMM:** Hidden Markov Model

1209 **HOG:** Hierarchical Ortholog Group

1210 **IR:** Ionotropic Receptor

1211 **LRR:** Leucine Rich Repeat Proteins

1212 **Mbp:** Mega Base Pairs, or 1,000,000bp

1213 **MP:** Mate-pair sequence data

1214 **NCBI:** National Center for Biotechnology Information

1215 **N50:** A measure of genome completeness. The length of the scaffold containing the

1216 middle nucleotide

1217 **OXPHOS:** Oxidative Phosphorylation

1218 **OBP:** Odorant-binding Protein

1219 **OR:** Odorant Receptor

1220 **PE:** Paired-end sequence data

1221 **RSCU:** Relative Synonymous Codon Usage

1222 **TE:** Transposable Element

1223

1224

1225

1226

1227

1228

1229

1230

1231

1232

1233

1234

1235

1236

1237

1238

1239

1240

1241

1242

1243 **Availability of data and materials**

1244 Both genomes are available from the NCBI Genome database (PRJNA587428, *A. ervi*:  
1245 SAMN13190903, *L. fabarum*: SAMN13190904). The assemblies, predicted genes, and  
1246 annotations are also available at <https://bipaa.genouest.org>. Raw Illumina and PacBio  
1247 sequence data used to construct genomes is available in NCBI SRA for both *A. ervi*  
1248 (SAMN12878248) and *L. fabarum* (accessions SAMN10617865, SAMN10617866,  
1249 SAMN10617867), and is further detailed in Supplementary Tables 1 and 2. Venom  
1250 protein data are available via ProteomeXchange with identifier PXD015758.

1251

1252

1253

1254

1255

1256

## 1257 Acknowledgements

1258 Thanks to the many people who helped with collections and insect rearing, especially  
1259 Paula Rodriguez (EAWAG), Laury Arthaud and Christian Rebuf (ESIM, INRA-ISA),  
1260 Francisca Zepeda-Paulo (UACH), Sebastian Ortiz-Martinez, Cinthya Villegas, and  
1261 Daniela Sepúlveda (UTALCA). Lucia Briones (UTALCA) and Dominique Cazes (ESIM,  
1262 INRA-ISA) provided help in with *A. ervi* DNA extractions. Paul Saffert (Uni Potsdam)  
1263 provided valuable discussion leading to the codon usage analysis. David Pratella (ESIM,  
1264 INRA-ISA) provided help on the annotation of immune genes.

1265 *Aphidius ervi* sequencing was funded by FONDECYT grant 1130483 and  
1266 Iniciativa Científica Milenio (ICM) NC120027 (both to Christian Figueroa and Blas  
1267 Lavandero, Universidad de Talca, Chile), INRA (AIP “séquençage” INRA Rennes, France),  
1268 and funding from ESIM team (Marylène Poirié, INRA-ISA Sophia Antipolis, France) and  
1269 BGI (funding to Denis Tagu, INRA Rennes, France). The ESIM team is supported by the  
1270 French Government (National Research Agency, ANR) through the "Investments for the  
1271 Future" LABEX SIGNALIFE : program reference # ANR-11-LABX-0028-01. Mark Lammers  
1272 would like to thank Panagiotis Provataris (ZFMK, Bonn, Germany) for bringing to our  
1273 attention the existent small whole genome bisulfite sequencing data set for *Aphidius*  
1274 *ervi*.

1275 *Lysiphlebus fabarum* data were generated in collaboration with the Genetic  
1276 Diversity Centre (GDC, with particular thanks to Stefan Zoller and Jean-Claude Walser),  
1277 ETH Zurich, and utilized the ETH Scientific Computing Cluster (Euler). Orthologs were  
1278 computed on the University of Potsdam's High Performance Computing Cluster  
1279 Orson2, managed by the ZIM. *Lysiphlebus fabarum* sequencing was funded by an SNSF  
1280 professorship to Christoph Vorburger (grant nrs. PP00P3\_123376 and

1281 PP00P3\_146341). TS acknowledges SNSF grant nr PP00P3\_170627. JG and LS  
1282 acknowledge DFG grant SPP 1819 Rapid evolutionary adaptation (GA 661/4-1, SCHR  
1283 1554/3-1).

1284 DNA sequences from the *Myzus* genomes used in comparative analysis of  
1285 codon usage were downloaded from AphidBase. Funding for *Myzus persicae* clone  
1286 G006 genomic sequencing was provided by USDA-NIFA award 2010-65105-20558.  
1287 Funding for *M. persicae* clone O genomic sequencing was provided by The Genome  
1288 Analyses Centre (TGAC) Capacity and Capability Challenge program (project CCC-15 and  
1289 BB/J004553/1), from the Biotechnology and Biological Sciences Research Council  
1290 (BBSRC), and the John Innes Foundation.

1291

#### 1292 **Statement on competing interests**

1293 The authors declare no competing interests

1294

1295

1296

1297

1298

1299

1300

1301

1302

1303

1304



1305

1306 REFERENCES

- 1307 Acquisti C, Elser JJ, Kumar S (2009) Ecological nitrogen limitation shapes the DNA  
1308 composition of plant genomes. *Molecular Biology and Evolution* **26**, 953-956.
- 1309 Ahmed T, Zhang T, Wang Z, He K, Bai S (2016) Gene set of chemosensory receptors  
1310 in the polyembryonic endoparasitoid *Macrocentrus cingulum*. *Scientific*  
1311 *Reports* **6**, 24078-24078.
- 1312 Aljanabi S, Martinez I (1997) Universal and rapid salt-extraction of high quality  
1313 genomic DNA for PCR-based techniques. *Nucleic Acids Research* **25**, 4692 -  
1314 4693.
- 1315 Almpanis A, Swain M, Gatherer D, McEwan N (2018) Correlation between bacterial  
1316 G+C content, genome size and the G+C content of associated plasmids and  
1317 bacteriophages. *Microbial Genomics* **4**, -.
- 1318 Altenhoff AM, Levy J, Zarowiecki M, *et al.* (2018) OMA standalone: orthology  
1319 inference among public and custom genomes and transcriptomes. *bioRxiv*.
- 1320 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment  
1321 search tool. *Journal of Molecular Biology* **215**, 403-410.
- 1322 Altschul SF, Madden TL, Schäffer AA, *et al.* (1997) Gapped BLAST and PSI-  
1323 BLAST: a new generation of protein database search programs. *Nucleic Acids*  
1324 *Research* **25**, 3389-3402.
- 1325 Andrade López JM, Lanno SM, Auerbach JM, *et al.* (2017) Genetic basis of octanoic  
1326 acid resistance in *Drosophila sechellia*: functional analysis of a fine-mapped  
1327 region. *Molecular Ecology* **26**, 1148-1160.
- 1328 Andrews S, Krueger F, Segonds-Pichon A, *et al.* (2010) FastQC: a quality control tool  
1329 for high throughput sequence data. Available online at:  
1330 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- 1331 Ardila-Garcia AM, Umphrey GJ, Gregory TR (2010) An expansion of the genome  
1332 size dataset for the insect order Hymenoptera, with a first test of parasitism  
1333 and eusociality as possible constraints. *Insect Molecular Biology* **19**, 337-346.
- 1334 Ballesteros GI, Gadau J, Legeai F, *et al.* (2017) Expression differences in *Aphidius*  
1335 *ervi* (Hymenoptera: Braconidae) females reared on different aphid host  
1336 species. *PeerJ* **5**, e3640.
- 1337 Bao W, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive  
1338 elements in eukaryotic genomes. *Mobile DNA* **6**, 11.
- 1339 Bao Z, Eddy SR (2002) Automated *de novo* identification of repeat sequence families  
1340 in sequenced genomes. *Genome Research* **12**, 1269-1276.
- 1341 Barahimipour R, Strenkert D, Neupert J, *et al.* (2015) Dissecting the contributions of  
1342 GC content and codon usage to gene expression in the model alga  
1343 *Chlamydomonas reinhardtii*. *The Plant Journal* **84**, 704-717.
- 1344 Behura SK, Severson DW (2013) Codon usage bias: causative factors, quantification  
1345 methods and genome-wide patterns: with emphasis on insect genomes.  
1346 *Biological Reviews* **88**, 49-61.
- 1347 Bell LR, Maine EM, Schedl P, Cline TW (1988) Sex-lethal, a *Drosophila* sex  
1348 determination switch gene, exhibits sex-specific RNA splicing and sequence  
1349 similarity to RNA binding proteins. *Cell* **55**, 1037-1046.
- 1350 Belshaw R, Quicke DL (2003) The cytogenetics of thelytoky in a predominantly  
1351 asexual parasitoid wasp with covert sex. *Genome* **46**, 170-173.

- 1352 Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences.  
1353 *Nucleic Acids Research* **27**, 573-580.
- 1354 Bentele K, Saffert P, Rauscher R, Ignatova Z, Blüthgen N (2013) Efficient translation  
1355 initiation dictates codon usage at gene start. *Molecular systems biology* **9**, 675-  
1356 675.
- 1357 Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ (2017) Evolution of DNA methylation  
1358 across insects. *Molecular Biology and Evolution* **34**, 654-665.
- 1359 Bischoff V, Vignal C, Boneca IG, *et al.* (2004) Function of the *Drosophila* pattern-  
1360 recognition receptor PGRP-SD in the detection of Gram-positive bacteria.  
1361 *Nature Immunology* **5**, 1175-1180.
- 1362 Boetzer M, Pirovano W (2014) SSPACE-LongRead: scaffolding bacterial draft  
1363 genomes using long read sequence information. *BMC Bioinformatics* **15**, 211.
- 1364 Boivin G, Hance T, Brodeur J (2012) Aphid parasitoids in biological control. *Can J*  
1365 *Plant Sci* **92**.
- 1366 Buchon N, Silverman N, Cherry S (2014) Immunity in *Drosophila melanogaster* —  
1367 from microbial recognition to whole-organism physiology. *Nature Reviews*  
1368 *Immunology* **14**, 796.
- 1369 Burke GR, Strand MR (2014) Systematic analysis of a wasp parasitism arsenal.  
1370 *Molecular Ecology* **23**, 890-901.
- 1371 Burke GR, Walden KKO, Whitfield JB, Robertson HM, Strand MR (2018) Whole  
1372 genome sequence of the parasitoid wasp *Microplitis demolitor* that harbors an  
1373 endogenous virus mutualist. *G3: Genes/Genomes/Genetics*.
- 1374 Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks:  
1375 Building and genotyping loci *de novo* from short-read sequences. *G3: Genes,*  
1376 *Genomes, Genetics* **1**, 171-182.
- 1377 Chaney JL, Clark PL (2015) Roles for synonymous codon usage in protein  
1378 biogenesis. *Annual Review of Biophysics* **44**, 143-166.
- 1379 Charif D, Lobry JR (2007) SeqinR 1.0-2: A contributed package to the R project for  
1380 statistical computing devoted to biological sequences retrieval and analysis.  
1381 In: *Structural Approaches to Sequence Evolution: Molecules, Networks,*  
1382 *Populations* (eds. Bastolla U, Porto M, Roman HE, Vendruscolo M), pp. 207-  
1383 232. Springer Berlin Heidelberg, Berlin, Heidelberg.
- 1384 Charroux B, Royet J (2010) *Drosophila* immune response: From systemic  
1385 antimicrobial peptide production in fat body cells to local defense in the  
1386 intestinal tract. *Fly* **4**, 40-47.
- 1387 Chau A, Mackauer M (2000) Host-instar selection in the aphid parasitoid *Monoctonus*  
1388 *paulensis* (Hymenoptera: Braconidae, Aphidiinae): a preference for small pea  
1389 aphids. *EJE* **97**, 347-353.
- 1390 Chen X-x, van Achterberg C (2018) Systematics, phylogeny, and evolution of  
1391 braconid wasps: 30 years of progress. *Annual Review of Entomology*.
- 1392 Cheng R-X, Meng L, Mills NJ, Li B (2011) Host preference between symbiotic and  
1393 aposymbiotic *Aphis fabae*, by the aphid parasitoid, *Lysiphlebus ambiguus*.  
1394 *Journal of Insect Science* **11**, 81-81.
- 1395 Colinet D, Anselme C, Deleury E, *et al.* (2014) Identification of the main venom  
1396 protein components of *Aphidius ervi*, a parasitoid wasp of the aphid model  
1397 *Acyrtosiphon pisum*. *BMC Genomics* **15**, 342.
- 1398 Comeault AA, Serrato-Capuchina A, Turissini DA, *et al.* (2017) A nonrandom subset  
1399 of olfactory genes is associated with host preference in the fruit fly *Drosophila*  
1400 *orena*. *Evolution Letters* **1**, 73-85.

- 1401 Danecek P, Auton A, Abecasis G, *et al.* (2011) The variant call format and VCFtools.  
1402 *Bioinformatics* **27**, 2156-2158.
- 1403 De Gregorio E, Spellman PT, Rubin GM, Lemaitre B (2001) Genome-wide analysis  
1404 of the *Drosophila* immune response by using oligonucleotide microarrays.  
1405 *Proceedings of the National Academy of Sciences* **98**, 12590.
- 1406 De Gregorio E, Spellman PT, Tzou P, Rubin GM, Lemaitre B (2002) The Toll and  
1407 Imd pathways are the major regulators of the immune response in *Drosophila*.  
1408 *The EMBO Journal* **21**, 2568-2579.
- 1409 Dennis AB, Käch H, Vorburger C (in revision) Dual RNA-seq in an aphid parasitoid  
1410 reveals plastic and evolved adaptation.
- 1411 Dennis AB, Patel V, Oliver KM, Vorburger C (2017) Parasitoid gene expression  
1412 changes after adaptation to symbiont-protected hosts. *Evolution* **71**, 2599-  
1413 2617.
- 1414 Dion E, Zélé F, Simon JC, Outreman Y (2011) Rapid evolution of parasitoids when  
1415 faced with the symbiont-mediated resistance of their hosts. *Journal of*  
1416 *Evolutionary Biology* **24**, 741-750.
- 1417 Dobin A, Davis CA, Schlesinger F, *et al.* (2013) STAR: ultrafast universal RNA-seq  
1418 aligner. *Bioinformatics* **29**.
- 1419 Dorémus T, Urbach S, Jouan V, *et al.* (2013) Venom gland extract is not required for  
1420 successful parasitism in the polydnavirus-associated endoparasitoid *Hyposoter*  
1421 *didymator* (Hym. Ichneumonidae) despite the presence of numerous novel and  
1422 conserved venom proteins. *Insect Biochemistry and Molecular Biology* **43**,  
1423 292-307.
- 1424 *Drosophila* 12 Genomes C, Clark AG, Eisen MB, *et al.* (2007) Evolution of genes and  
1425 genomes on the *Drosophila* phylogeny. *Nature* **450**, 203.
- 1426 Dunn NA, Unni DR, Diesh C, *et al.* (2019) Apollo: Democratizing genome  
1427 annotation. *PLoS Comput Biol* **15**, e1006790.
- 1428 Dupas S, Carton Y, Poiriè M (2003) Genetic dimension of the coevolution of  
1429 virulence–resistance in *Drosophila* – parasitoid wasp relationships. *Heredity*  
1430 **90**, 84.
- 1431 Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* **14**, 755-763.
- 1432 Edgar RC (2004a) MUSCLE: a multiple sequence alignment method with reduced  
1433 time and space complexity. *BMC Bioinformatics* **5**, 113.
- 1434 Edgar RC (2004b) MUSCLE: multiple sequence alignment with high accuracy and  
1435 high throughput. *Nucleic Acids Research* **32**, 1792-1797.
- 1436 Edgar RC, Myers EW (2005) PILER: identification and classification of genomic  
1437 repeats. *Bioinformatics* **21**, i152-i158.
- 1438 Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the  
1439 cell using TargetP, SignalP and related tools. *Nature Protocols* **2**, 953-971.
- 1440 Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole  
1441 genome comparisons dramatically improves orthogroup inference accuracy.  
1442 *Genome Biology* **16**, 157.
- 1443 Falabella P, Riviello L, Caccialupi P, *et al.* (2007) A  $\gamma$ -glutamyl transpeptidase of  
1444 *Aphidius ervi* venom induces apoptosis in the ovaries of host aphids. *Insect*  
1445 *Biochemistry and Molecular Biology* **37**, 453-465.
- 1446 Falabella P, Tremblay E, Pennacchio F (2003) Host regulation by the aphid parasitoid  
1447 *Aphidius ervi*: the role of teratocytes. *Entomologia Experimentalis Et*  
1448 *Applicata* **97**, 1-9.
- 1449 Finn RD, Bateman A, Clements J, *et al.* (2013) Pfam: the protein families database.  
1450 *Nucleic Acids Research* **42**, D222-D230.

- 1451 Flutre T, Duprat E, Feuillet C, Quesneville H (2011) Considering transposable  
1452 element diversification in *de novo* annotation approaches. *Plos One* **6**, e16526.  
1453 Foerstner KU, von Mering C, Hooper SD, Bork P (2005) Environments shape the  
1454 nucleotide composition of genomes. *EMBO reports* **6**, 1208.  
1455 Forbes AA, Bagley RK, Beer MA, Hippee AC, Widmayer HA (2018) Quantifying the  
1456 unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal  
1457 order. *BMC Ecology* **18**, 21.  
1458 Gadau J (2009) Phase-unknown linkage mapping in ants. *Cold Spring Harb Protoc*  
1459 **2009**, pdb prot5251.  
1460 Galtier N, Roux C, Rousselle M, *et al.* (2018) Codon usage bias in animals:  
1461 Disentangling the effects of natural selection, effective population size, and  
1462 GC-biased gene conversion. *Molecular Biology and Evolution* **35**, 1092-1103.  
1463 Geib SM, Liang GH, Murphy TD, Sim SB (2017) Whole genome sequencing of the  
1464 Braconid parasitoid wasp *Fopius arisanus*, an important biocontrol agent of  
1465 pest teplitid fruit flies. *G3: Genes/Genomes/Genetics* **7**, 2407-2411.  
1466 Gerardo NM, Altincicek B, Anselme C, *et al.* (2010) Immunity and other defenses in  
1467 pea aphids, *Acyrtosiphon pisum*. *Genome Biology* **11**, R21.  
1468 Geuverink E (2017) *Parental and endosymbiont effects on sex determination in*  
1469 *haplodiploid wasps : Who is in control?*, University of Groningen.  
1470 Geuverink E, Verhulst EC, van Leussen M, van de Zande L, Beukeboom LW (2018)  
1471 Maternal provision of non-sex-specific transformer messenger RNA in sex  
1472 determination of the wasp *Asobara tabida*. *Insect Molecular Biology* **27**, 99-  
1473 109.  
1474 Gibson JD, Niehuis O, Verrelli BC, Gadau J (2010) Contrasting patterns of selective  
1475 constraints in nuclear-encoded genes of the oxidative phosphorylation  
1476 pathway in holometabolous insects and their possible role in hybrid  
1477 breakdown in *Nasonia*. *Heredity* **104**, 310.  
1478 Godfray HCJ (1994) *Parasitoids: behavioral and evolutionary ecology* Princeton  
1479 University Press, Princeton, N.J.  
1480 Gold DA, Katsuki T, Li Y, *et al.* (2018) The genome of the jellyfish *Aurelia* and the  
1481 evolution of animal complexity. *Nat Ecol Evol*.  
1482 Grabherr MG, Haas BJ, Yassour M, *et al.* (2011) Full-length transcriptome assembly  
1483 from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**,  
1484 644-652.  
1485 Greenwood JM, Milutinovic B, Peuss R, *et al.* (2017) Oral immune priming with  
1486 *Bacillus thuringiensis* induces a shift in the gene expression of *Tribolium*  
1487 *castaneum* larvae. *BMC Genomics* **18**, 329.  
1488 Gross P (1993) Insect behavioral and morphological defenses against parasitoids.  
1489 *Annual Review of Entomology* **38**, 251-273.  
1490 Guindon S, Dufayard J-F, Lefort V, *et al.* (2010) New algorithms and methods to  
1491 estimate maximum-likelihood phylogenies: Assessing the performance of  
1492 PhyML 3.0. *Systematic Biology* **59**, 307-321.  
1493 Guo J, Hatt S, He K, *et al.* (2017) Nine facultative endosymbionts in aphids. A  
1494 review. *Journal of Asia-Pacific Entomology* **20**, 794-801.  
1495 Haas BJ, Salzberg SL, Zhu W, *et al.* (2008) Automated eukaryotic gene structure  
1496 annotation using EVIDENCEModeler and the Program to Assemble Spliced  
1497 Alignments. *Genome Biology* **9**, R7.  
1498 Hackl T, Hedrich R, Schultz J, Förster F (2014) proofread: large-scale high-accuracy  
1499 PacBio correction through iterative short read consensus. *Bioinformatics* **30**,  
1500 3004-3011.

- 1501 Hanrahan SJ, Johnston JS (2011) New genome size estimates of 134 species of  
1502 arthropods. *Chromosome Res* **19**, 809-823.
- 1503 Harrison MC, Jongepier E, Robertson HM, *et al.* (2018) Hemimetabolous genomes  
1504 reveal molecular basis of termite eusociality. *Nature Ecology & Evolution* **2**,  
1505 557-566.
- 1506 Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications  
1507 through DNA barcodes. *Proceedings. Biological sciences* **270**, 313-321.
- 1508 Heimpel GE, de Boer JG (2007) Sex determination in the Hymenoptera. *Annual*  
1509 *Review of Entomology* **53**, 209-230.
- 1510 Heimpel GE, Mills NJ (2017) *Biological control : ecology and applications*.  
1511 <http://dx.doi.org/10.1017/9781139029117>
- 1512 Helmkampf M, Cash E, Gadau J (2015) Evolution of the insect desaturase gene  
1513 family with an emphasis on social Hymenoptera. *Molecular Biology and*  
1514 *Evolution* **32**, 456-471.
- 1515 Henry LM, Roitberg BD, Gillespie DR (2008) Host-range evolution in *Aphidius*  
1516 parasitoids: Fidelity, virulence and fitness trade-offs on an ancestral host.  
1517 *Evolution* **62**, 689-699.
- 1518 Henter HJ, Via S (1995) The potential for coevolution in a host-parasitoid system. I.  
1519 Genetic variation within an aphid population in susceptibility to a parasitic  
1520 wasp. *Evolution* **49**, 427-438.
- 1521 Heraty J (2009) Parasitoid biodiversity and insect pest management. *Insect*  
1522 *Biodiversity*.
- 1523 Herzog J, Muller CB, Vorburger C (2007) Strong parasitoid-mediated selection in  
1524 experimental populations of aphids. *Biol Lett* **3**, 667-669.
- 1525 Hoede C, Arnoux S, Moisset M, *et al.* (2014) PASTEC: An automatic transposable  
1526 element classification tool. *Plos One* **9**, e91929.  
1527 <https://bipaa.genouest.org> *Bioinformatics Platform for Agroecosystem Arthropods*  
1528 *(BIPAA)*. <https://bipaa.genouest.org>
- 1529 Huang H, Wu P, Zhang S, *et al.* (2019) DNA methylomes and transcriptomes analysis  
1530 reveal implication of host DNA methylation machinery in BmNPV  
1531 proliferation in *Bombyx mori*. *BMC Genomics* **20**, 736.
- 1532 Huang X (1994) On global sequence alignment. *Comput Appl Biosci* **10**, 227-235.
- 1533 Hufbauer RA, Bogdanowicz SM, Harrison RG (2004) The population genetics of a  
1534 biological control introduction: mitochondrial DNA and microsatellite  
1535 variation in native and introduced populations of *Aphidus ervi*, a parasitoid  
1536 wasp. *Molecular Ecology* **13**, 337-348.
- 1537 Jeltsch A, Jurkowska RZ (2014) New concepts in DNA methylation. *Trends in*  
1538 *Biochemical Sciences* **39**, 310-318.
- 1539 Jurka J (1998) Repeats in genomic DNA: mining and meaning. *Current Opinion in*  
1540 *Structural Biology* **8**, 333-337.
- 1541 Kajitani R, Toshimoto K, Noguchi H, *et al.* (2014) Efficient de novo assembly of  
1542 highly heterozygous genomes from whole-genome shotgun short reads.  
1543 *Genome Research* **24**, 1384-1395.
- 1544 Kang Z-W, Tian H-G, Liu F-H, *et al.* (2017) Identification and expression analysis of  
1545 chemosensory receptor genes in an aphid endoparasitoid *Aphidius gifuensis*.  
1546 *Scientific Reports* **7**, 3939.
- 1547 Kassambara A, Mundt F (2016) Factoextra: extract and visualize the results of  
1548 multivariate data analyses (ed. package r).

- 1549 Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software  
1550 Version 7: Improvements in performance and usability. *Molecular Biology*  
1551 *and Evolution* **30**, 772-780.
- 1552 Kavallieratos NG, Tomanovi, x, *et al.* (2004) A survey of aphid parasitoids  
1553 (Hymenoptera: Braconidae: Aphidiinae) of Southeastern Europe and their  
1554 aphid-plant associations. *Applied Entomology and Zoology* **39**, 527-563.
- 1555 Keilwagen J, Wenk M, Erickson JL, *et al.* (2016) Using intron position conservation  
1556 for homology-based gene prediction. *Nucleic Acids Research* **44**, e89-e89.
- 1557 Keller O, Kollmar M, Stanke M, Waack S (2011) A novel hybrid gene prediction  
1558 method employing protein multiple sequence alignments. *Bioinformatics* **27**,  
1559 757-763.
- 1560 Keller O, Odronitz F, Stanke M, Kollmar M, Waack S (2008) Scipio: using protein  
1561 sequences to determine the precise exon/intron structures of genes and their  
1562 orthologs in closely related species. *BMC Bioinformatics* **9**, 278-278.
- 1563 Kim BY, Jin BR (2015) Apolipoprotein III from honeybees (*Apis cerana*) exhibits  
1564 antibacterial activity. *Comparative Biochemistry and Physiology Part B:*  
1565 *Biochemistry and Molecular Biology* **182**, 6-13.
- 1566 Kohany O, Gentles AJ, Hankus L, Jurka J (2006) Annotation, submission and  
1567 screening of repetitive elements in Repbase: RepbaseSubmitter and Censor.  
1568 *BMC Bioinformatics* **7**, 474.
- 1569 Kolpakov R, Bana G, Kucherov G (2003) mreps: Efficient and flexible detection of  
1570 tandem repeats in DNA. *Nucleic Acids Research* **31**, 3672-3678.
- 1571 Koren S, Walenz BP, Berlin K, *et al.* (2017) Canu: scalable and accurate long-read  
1572 assembly via adaptive k-mer weighting and repeat separation. *Genome*  
1573 *Research* **27**, 722-736.
- 1574 Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59.
- 1575 Kraaijeveld AR, Van Alphen JJM, Godfray HCJ (1998) The coevolution of host  
1576 resistance and parasitoid virulence. *Parasitology* **116**, S29-S45.
- 1577 Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M (2006) High guanine and  
1578 cytosine content increases mRNA levels in mammalian cells. *Plos Biology* **4**,  
1579 e180.
- 1580 Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants  
1581 of gene expression in *Escherichia coli*. *Science* **324**, 255.
- 1582 Kuroki S, Matoba S, Akiyoshi M, *et al.* (2013) Epigenetic regulation of mouse sex  
1583 determination by the histone demethylase *Jmjd1a*. *Science* **341**, 1106.
- 1584 Kurtz S, Phillippy A, Delcher AL, *et al.* (2004) Versatile and open software for  
1585 comparing large genomes. *Genome Biology* **5**, R12.
- 1586 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature*  
1587 *Methods* **9**, 357-359.
- 1588 Leclair M, Pons I, Mahéo F, *et al.* (2016) Diversity in symbiont consortia in the pea  
1589 aphid complex is associated with large phenotypic variation in the insect host.  
1590 *Evolutionary Ecology* **30**, 925-941.
- 1591 Lee E, Helt GA, Reese JT, *et al.* (2013) Web Apollo: a web-based genomic  
1592 annotation editing platform. *Genome Biology* **14**, R93.
- 1593 Lefort V, Longueville J-E, Gascuel O (2017) SMS: Smart Model Selection in PhyML.  
1594 *Molecular Biology and Evolution* **34**, 2422-2424.
- 1595 Legeai F, Shigenobu S, Gauthier JP, *et al.* (2010) AphidBase: a centralized  
1596 bioinformatic resource for annotation of the pea aphid genome. *Insect*  
1597 *Molecular Biology* **19 Suppl 2**, 5-12.

- 1598 Leggett RM, Clavijo BJ, Clissold L, Clark MD, Caccamo M (2014) NextClip: an  
1599 analysis and read preparation tool for Nextera Long Mate Pair libraries.  
1600 *Bioinformatics* **30**, 566-568.
- 1601 Lemaitre B, Hoffman J (2007) The host defence of *Drosophila melanogaster*. *Ann*  
1602 *Rev Immunol* **25**.
- 1603 Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new  
1604 developments. *Nucleic Acids Research* **47**, W256-W259.
- 1605 Li H (2011) A statistical framework for SNP calling, mutation discovery, association  
1606 mapping and population genetical parameter estimation from sequencing data.  
1607 *Bioinformatics (Oxford, England)* **27**, 2987-2993.
- 1608 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler  
1609 transform. *Bioinformatics* **25**.
- 1610 Li H, Handsaker B, Wysoker A, *et al.* (2009) The Sequence Alignment/Map format  
1611 and SAMtools. *Bioinformatics* **25**, 2078-2079.
- 1612 Li Y, Park H, Smith TE, Moran NA (2019) Gene family evolution in the pea aphid  
1613 based on chromosome-level genome assembly. *Molecular Biology and*  
1614 *Evolution* **36**, 2143-2156.
- 1615 Li Y, Zhang R, Liu S, *et al.* (2017) The molecular evolutionary dynamics of oxidative  
1616 phosphorylation (OXPHOS) genes in Hymenoptera. *BMC Evolutionary*  
1617 *Biology* **17**, 269.
- 1618 Liepert C, Dettner K (1993) Recognition of aphid parasitoids by honeydew-collecting  
1619 ants: The role of cuticular lipids in a chemical mimicry system. *Journal of*  
1620 *Chemical Ecology* **19**, 2143-2153.
- 1621 Liepert C, Dettner K (1996) Role of cuticular hydrocarbons of aphid parasitoids in  
1622 their relationship to aphid-attending ants. *Journal of Chemical Ecology* **22**,  
1623 695-707.
- 1624 Liu N-Y, Wang J-Q, Zhang Z-B, Huang J-M, Zhu J-Y (2017) Unraveling the venom  
1625 components of an encyrtid endoparasitoid wasp *Diversinervus elegans*.  
1626 *Toxicon* **136**, 15-26.
- 1627 Los DA, Murata N (1998) Structure and expression of fatty acid desaturases.  
1628 *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **1394**, 3-  
1629 15.
- 1630 Łukasik P, Dawid MA, Ferrari J, Godfray HCJ (2013) The diversity and fitness  
1631 effects of infection with facultative endosymbionts in the grain aphid, *Sitobion*  
1632 *avenae*. *Oecologia* **173**, 985-996.
- 1633 Luo R, Liu B, Xie Y, *et al.* (2012) SOAPdenovo2: an empirically improved memory-  
1634 efficient short-read *de novo* assembler. *GigaScience* **1**, 18-18.
- 1635 Lüthi MN, Vorburger C, Dennis AB (submitted) A novel RNA virus in the parasitoid  
1636 wasp *Lysiphlebus fabarum*: genomic structure, prevalence and transmission.
- 1637 Martinez AJ, Kim KL, Harmon JP, Oliver KM (2016) Specificity of multi-modal  
1638 aphid defenses against two rival parasitoids. *Plos One* **11**, e0154670.
- 1639 Mateo Leach I, Pannebakker BA, Schneider MV, *et al.* (2009) Thelytoky in  
1640 Hymenoptera with *Venturia canescens* and *Leptopilina clavipes* as Case  
1641 Studies. In: *Lost Sex: The Evolutionary Biology of Parthenogenesis* (eds.  
1642 Schön I, Martens K, Dijk P), pp. 347-375. Springer Netherlands, Dordrecht.
- 1643 Matthey-Doret C, van der Kooij CJ, Jeffries DL, *et al.* (2019) Mapping of multiple  
1644 complementary sex determination loci in a parasitoid wasp. *Genome Biology*  
1645 *and Evolution*.
- 1646 McCutcheon JP, McDonald BR, Moran N (2009) *Origin of an alternative genetic*  
1647 *code in the extremely small and GC-rich genome of a bacterial symbiont.*

- 1648 Miao X, Huang Y, Zhu X, Ding D (2004) A comparative study on development and  
1649 reproduction of the parasitoid *Lysiphlebus japonicus* (Hymenoptera:  
1650 Aphidiidae) in symbiotic and aposymbiotic host aphids. *Applied Entomology  
1651 and Zoology* **39**, 243-248.
- 1652 Miller MA, Schwartz T, Pickett BE, *et al.* (2015) A RESTful API for access to  
1653 phylogenetic tools via the CIPRES science gateway. *Evolutionary  
1654 bioinformatics online* **11**, 43-48.
- 1655 Mistry J, Finn RD, Eddy SR, Bateman A, Punta M (2013) Challenges in homology  
1656 search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic  
1657 Acids Research* **41**, e121-e121.
- 1658 Monticelli LS, Nguyen LTH, Amiens-Desneux E, *et al.* (2019) The preference-  
1659 performance relationship as a means of classifying parasitoids according to  
1660 their specialization degree. *Evolutionary Applications* **0**.
- 1661 Moreau S, Asgari S (2015) Venom proteins from parasitoid wasps and their biological  
1662 functions. *Toxins* **7**, 2385.
- 1663 Morin-Poulard I, Vincent A, Crozatier M (2013) The *Drosophila* JAK-STAT  
1664 pathway in blood cell formation and immunity. *JAK-STAT* **2**, e25700.
- 1665 Mugal CF, Arndt PF, Holm L, Ellegren H (2015) Evolutionary consequences of DNA  
1666 methylation on the GC content in vertebrate genomes. *G3:  
1667 Genes/Genomes/Genetics* **5**, 441.
- 1668 Müller CB, Adriaanse ICT, Belshaw R, Godfray HCJ (2004) The structure of an  
1669 aphid–parasitoid community. *Journal of Animal Ecology* **68**, 346-370.
- 1670 Myllymäki H, Valanne S, Rämet M (2014) The *Drosophila* imd signaling pathway.  
1671 *The Journal of Immunology* **192**, 3455.
- 1672 NCBI NCfBI, Bethesda (MD) *nr database, available from  
1673 ftp.ncbi.nlm.nih.gov/blast/db/*.
- 1674 Nielsen H (2017) Predicting Secretory Proteins with SignalP, available at:  
1675 <http://www.cbs.dtu.dk/services/SignalP/>. In: *Protein Function Prediction:  
1676 Methods and Protocols* (ed. Kihara D), pp. 59-73. Springer New York, New  
1677 York, NY.
- 1678 Nishide Y, Kageyama D, Yokoi K, *et al.* (2019) Functional crosstalk across IMD and  
1679 Toll pathways: insight into the evolution of incomplete immune cascades.  
1680 *Proceedings of the Royal Society B: Biological Sciences* **286**, 20182207.
- 1681 Nouhaud P, Gautier M, Gouin A, *et al.* (2018) Identifying genomic hotspots of  
1682 differentiation and candidate genes involved in the adaptive divergence of pea  
1683 aphid host races. *Molecular Ecology* **27**, 3287-3300.
- 1684 Obbard DJ, Shi M, Longdon B, Dennis AB (in revision) A new family of segmented  
1685 RNA viruses infecting animals.
- 1686 Oliver KM, Degan PH, Burke GR, Moran NA (2010) Facultative symbionts in  
1687 aphids and the horizontal transfer of ecologically important traits. *Annual  
1688 Review of Entomology* **55**, 247-266.
- 1689 Oliver KM, Higashi CHV (2018) Variations on a protective theme: *Hamiltonella  
1690 defensa* infections in aphids variably impact parasitoid success. *Current  
1691 Opinion in Insect Science*.
- 1692 Oliver KM, Russell JA, Moran NA, Hunter MS (2003) Facultative bacterial  
1693 symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the  
1694 National Academy of Sciences* **100**, 1803-1807.
- 1695 Parchman TL, Gompert Z, Mudge J, *et al.* (2012) Genome-wide association genetics  
1696 of an adaptive trait in lodgepole pine. *Molecular Ecology* **21**, 2991-3005.



- 1697 Peccoud J, Simon J-C, McLaughlin HJ, Moran NA (2009) Post-Pleistocene radiation  
1698 of the pea aphid complex revealed by rapidly evolving endosymbionts.  
1699 *Proceedings of the National Academy of Sciences* **106**, 16315.
- 1700 Pennacchio F, Fanti P, Falabella P, *et al.* (1999) Development and nutrition of the  
1701 braconid wasp, *Aphidius ervi* in aposymbiotic host aphids. *Archives of Insect*  
1702 *Biochemistry and Physiology* **40**, 53-63.
- 1703 Pennacchio F, Strand MR (2006) Evolution of developmental strategies in parasitic  
1704 hymenoptera. *Annual Review of Entomology* **51**, 233-258.
- 1705 Perkins DN, Pappin DJC, Creasy DM, Cottrell JS (1999) Probability-based protein  
1706 identification by searching sequence databases using mass spectrometry data.  
1707 *Electrophoresis* **20**, 3551-3567.
- 1708 Peters RS, Krogmann L, Mayer C, *et al.* (2017) Evolutionary history of the  
1709 Hymenoptera. *Current Biology* **27**, 1013-1018.
- 1710 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest  
1711 RADseq: An inexpensive method for *de novo* SNP discovery and genotyping  
1712 in model and non-model species. *Plos One* **7**, e37135.
- 1713 Plotkin JB, Kudla G (2011) Synonymous but not the same: the causes and  
1714 consequences of codon bias. *Nature reviews. Genetics* **12**, 32-42.
- 1715 Poelman EH, Bruinsma M, Zhu F, *et al.* (2012) Hyperparasitoids use herbivore-  
1716 induced plant volatiles to locate their parasitoid host. *Plos Biology* **10**,  
1717 e1001435.
- 1718 Poirié M, Colinet D, Gatti J-L (2014) Insights into function and evolution of  
1719 parasitoid wasp venoms. *Current Opinion in Insect Science* **6**, 52-60.
- 1720 Porcelli D, Barsanti P, Pesole G, Caggese C (2007) The nuclear OXPPOS genes in  
1721 insecta: a common evolutionary origin, a common cis-regulatory motif, a  
1722 common destiny for gene duplicates. *BMC Evol Biol* **7**, 215.
- 1723 Poulin R (2007) *Evolutionary Ecology of Parasites (Second Edition)* Princeton  
1724 University Press.
- 1725 Powell JR, Moriyama EN (1997) Evolution of codon usage bias in *Drosophila*.  
1726 *Proceedings of the National Academy of Sciences* **94**, 7784.
- 1727 Price MN, Dehal PS, Arkin AP (2009) FastTree: Computing large minimum  
1728 evolution trees with profiles instead of a distance matrix. *Molecular Biology*  
1729 *and Evolution* **26**, 1641-1650.
- 1730 Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – Approximately maximum-  
1731 likelihood trees for large alignments. *Plos One* **5**, e9490.
- 1732 Provataris P, Meusemann K, Niehuis O, Grath S, Misof B (2018) Signatures of DNA  
1733 methylation across insects suggest reduced DNA methylation levels in  
1734 holometabola. *Genome Biology and Evolution* **10**, 1185-1197.
- 1735 Purcell J, Brelsford A, Wurm Y, Perrin N, Chapuisat M (2014) Convergent genetic  
1736 architecture underlies social organization in ants. *Current Biology* **24**, 2728-  
1737 2732.
- 1738 Quesneville H, Bergman CM, Andrieu O, *et al.* (2005) Combined evidence annotation  
1739 of transposable elements in genome sequences. *PLoS Comput Biol* **1**, e22.
- 1740 Quicke DLJ (2014) *The Braconid and Ichneumonid parasitoid wasps : biology,*  
1741 *systematics, evolution and ecology.*  
1742 <http://public.eblib.com/choice/publicfullrecord.aspx?p=1882154>
- 1743 Ran W, Higgs PG (2010) The influence of anticodon–codon interactions and  
1744 modified bases on codon usage bias in bacteria. *Molecular Biology and*  
1745 *Evolution* **27**, 2129-2140.

- 1746 Rao YS, Chai XW, Wang ZF, Nie QH, Zhang XQ (2013) Impact of GC content on  
1747 gene expression pattern in chicken. *Genetics, selection, evolution : GSE* **45**, 9-  
1748 9.
- 1749 Rasekh A, Michaud JP, Kharazi-Pakdel A, Allahyari H (2010) Ant mimicry by an  
1750 aphid parasitoid, *Lysiphlebus fabarum*. *J Insect Sci* **10**, 126.
- 1751 Robertson HM, Gadau J, Wanner KW (2010) The insect chemoreceptor superfamily  
1752 of the parasitoid jewel wasp *Nasonia vitripennis*. *Insect Molecular Biology* **19**,  
1753 121-136.
- 1754 Robertson HM, Waterhouse RM, Walden KKO, *et al.* (2018) Genome sequence of  
1755 the wheat stem sawfly, *Cephus cinctus*, representing an early-branching  
1756 lineage of the Hymenoptera, illuminates evolution of hymenopteran  
1757 chemoreceptors. *Genome Biology and Evolution* **10**, 2997-3011.
- 1758 Rocha EPC, Danchin A (2002) Base composition bias might result from competition  
1759 for metabolic resources. *Trends in Genetics* **18**, 291-294.
- 1760 Romero IG, Ruvinsky I, Gilad Y (2012) Comparative studies of gene expression and  
1761 the evolution of gene regulation. *Nat Rev Genet* **13**, 505-516.
- 1762 Rouchet R, Vorburger C (2014) Experimental evolution of parasitoid infectivity on  
1763 symbiont-protected hosts leads to the emergence of genotype specificity.  
1764 *Evolution* **68**, 1607-1616.
- 1765 Sandrock C, Schirrmeister B, Vorburger C (2011) Evolution of reproductive mode  
1766 variation and host associations in a sexual-asexual complex of aphid  
1767 parasitoids. *BMC Evolutionary Biology* **11**, 348.
- 1768 Sandrock C, Vorburger C (2011) Single-locus recessive inheritance of asexual  
1769 reproduction in a parasitoid wasp. *Curr Biol* **21**, 433-437.
- 1770 Scarborough CL, Ferrari J, Godfray HCJ (2005) Aphid protected from pathogen by  
1771 endosymbiont. *Science* **310**, 1781.
- 1772 Schmitt-Engel C, Schultheis D, Schwirz J, *et al.* (2015) The iBeetle large-scale RNAi  
1773 screen reveals gene functions for insect development and physiology. *Nature*  
1774 *Communications* **6**, 7822.
- 1775 Schmitz A, Anselme C, Ravallec M, *et al.* (2012) The cellular immune response of  
1776 the pea aphid to foreign intrusion and symbiotic challenge. *Plos One* **7**,  
1777 e42114.
- 1778 Sepúlveda D, Zepeda-Paulo F, Ramírez C, Lavandero B, Figueroa C (2016) *Loss of*  
1779 *host fidelity in highly inbred populations of the parasitoid wasp Aphidius ervi*  
1780 *(Hymenoptera: Braconidae)*.
- 1781 Seward EA, Kelly S (2016) Dietary nitrogen alters codon bias and genome  
1782 composition in parasitic microorganisms. *Genome Biology* **17**, 226.
- 1783 Sharma A, Heinze SD, Wu Y, *et al.* (2017) Male sex in houseflies is determined by  
1784 *Mdmd*, a paralog of the generic splice factor gene *CWC22*. *Science* **356**, 642.
- 1785 Shi M, Chen X-X (2005) Molecular phylogeny of the Aphidiinae (Hymenoptera:  
1786 Braconidae) based on DNA sequences of 16S rRNA, 18S rDNA and ATPase  
1787 6 genes. *EJE* **102**, 133-138.
- 1788 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015)  
1789 BUSCO: assessing genome assembly and annotation completeness with  
1790 single-copy orthologs. *Bioinformatics* **31**, 3210-3212.
- 1791 Singh R, Bassett E, Chakravarti A, Parthun MR (2018) Replication-dependent histone  
1792 isoforms: a new source of complexity in chromatin structure and function.  
1793 *Nucleic Acids Research* **46**, 8665-8678.
- 1794 Slater GSC, Birney E (2005) Automated generation of heuristics for biological  
1795 sequence comparison. *BMC Bioinformatics* **6**, 31.

- 1796 Šmarda P, Bureš P, Horová L, *et al.* (2014) Ecological and evolutionary significance  
1797 of genomic GC content diversity in monocots. *Proceedings of the National*  
1798 *Academy of Sciences* **111**, E4096.
- 1799 Smit A, Hubley R, Green P (2013-2015) RepeatMasker Open-4.0.
- 1800 Smith CR, Morandin C, Noureddine M, Pant S (2018) Conserved roles of Osiris  
1801 genes in insect development, polymorphism and protection. *Journal of*  
1802 *Evolutionary Biology* **31**, 516-529.
- 1803 Smoyer LK, Dorer DR, Nickerson KW, Christensen AC (2003) Phenotype of the  
1804 Triplo-lethal locus of *Drosophila melanogaster* and its suppression by  
1805 hyperoxia. *Genet Res* **82**, 163-170.
- 1806 Soanes DM, Talbot NJ (2010) Comparative genome analysis reveals an absence of  
1807 leucine-rich repeat pattern-recognition receptor proteins in the kingdom Fungi.  
1808 *Plos One* **5**, e12725.
- 1809 Sperling P, Ternes P, Zank TK, Heinz E (2003) The evolution of desaturases.  
1810 *Prostaglandins, Leukotrienes and Essential Fatty Acids* **68**, 73-95.
- 1811 Strand MR (2014) Teratocytes and their functions in parasitoids. *Current Opinion in*  
1812 *Insect Science* **6**, 68-73.
- 1813 Sun Y (2016) sscu: Strength of Selected Codon Usage (ed. 2.6.0 Rpv).
- 1814 Takeda K, Akira S (2005) Toll-like receptors in innate immunity. *International*  
1815 *Immunology* **17**, 1-14.
- 1816 Tang H, Zhang X, Miao C, *et al.* (2015) ALLMAPS: robust scaffold ordering based  
1817 on multiple maps. *Genome Biology* **16**, 3.
- 1818 Tougeron K, Brodeur J, Le Lann C, van Baaren J (2019) How climate change affects  
1819 the seasonal ecology of insect parasitoids. *Ecological Entomology* **0**.
- 1820 Tougeron K, Le Lann C, Brodeur J, van Baaren J (2017) Are aphid parasitoids from  
1821 mild winter climates losing their winter diapause? *Oecologia* **183**, 619-629.
- 1822 Tvedte ES, Walden KKO, McElroy KE, *et al.* (2019) Genome of the Parasitoid Wasp  
1823 *Diachasma alloeum*, an Emerging Model for Ecological Speciation and  
1824 Transitions to Asexual Reproduction. *Genome Biology and Evolution* **11**,  
1825 2767-2773.
- 1826 Valanne S, Wang J-H, Rämetsä M (2011) The *Drosophila* toll signaling pathway. *The*  
1827 *Journal of Immunology* **186**, 649.
- 1828 Van Oss SB, Carvunis A-R (2019) De novo gene birth. *PLoS Genetics* **15**, e1008160.
- 1829 Van Vaerenbergh M, Debyser G, Devreese B, de Graaf DC (2014) Exploring the  
1830 hidden honeybee (*Apis mellifera*) venom proteome by integrating a  
1831 combinatorial peptide ligand library approach with FTMS. *Journal of*  
1832 *Proteomics* **99**, 169-178.
- 1833 Veleba A, Zedek F, Šmerda J, *et al.* (2016) Evolution of genome size and genomic  
1834 GC content in carnivorous holokinetics (Droseraceae). *Annals of Botany* **119**,  
1835 409-416.
- 1836 Vieira FG, Forêt S, He X, *et al.* (2012) Unique features of odorant-binding proteins of  
1837 the parasitoid wasp *Nasonia vitripennis* revealed by genome annotation and  
1838 comparative analyses. *Plos One* **7**, e43034.
- 1839 Vilcinskis A (2016) The role of epigenetics in host-parasite coevolution: lessons  
1840 from the model host insects *Galleria mellonella* and *Tribolium castaneum*.  
1841 *Zoology* **119**, 273-280.
- 1842 Vilcinskis A (2017) The impact of parasites on host insect epigenetics. *Advances in*  
1843 *Insect Physiology*.

- 1844 Vilcinskis A, Vogel H (2016) Seasonal phenotype-specific transcriptional  
1845 reprogramming during metamorphosis in the European map butterfly  
1846 *Araschnia levana*. *Ecol Evol* **6**, 3476-3485.
- 1847 Vinogradov AE (1998) Genome size and GC-percent in vertebrates as determined by  
1848 flow cytometry: The triangular relationship. *Cytometry* **31**, 100-109.
- 1849 Völkl W (1992) Aphids or their parasitoids: Who actually benefits from ant-  
1850 attendance? *Journal of Animal Ecology* **61**, 273-281.
- 1851 Völkl W (1997) Interactions between ants and aphid parasitoids: Patterns and  
1852 consequences for resource utilization. In: *Vertical Food Web Interactions:  
1853 Evolutionary Patterns and Driving Forces* (eds. Dettner K, Bauer G, Völkl  
1854 W), pp. 225-240. Springer Berlin Heidelberg, Berlin, Heidelberg.
- 1855 Vorburger C, Gehrler L, Rodriguez P (2010) A strain of the bacterial symbiont  
1856 *Regiella insecticola* protects aphids against parasitoids. *Biology Letters* **6**, 109-  
1857 111.
- 1858 Vorburger C, Perlman SJ (2018) The role of defensive symbionts in host-parasite  
1859 coevolution. *Biological Reviews* **93**, 1747-1764.
- 1860 Walker BJ, Abeel T, Shea T, *et al.* (2014) Pilon: An integrated tool for  
1861 comprehensive microbial variant detection and genome assembly  
1862 improvement. *Plos One* **9**, e112963.
- 1863 Wang J, Wurm Y, Nipitwattanaphon M, *et al.* (2013) A Y-like social chromosome  
1864 causes alternative colony organization in fire ants. *Nature* **493**, 664-668.
- 1865 Wang Y, Jorda M, Jones PL, *et al.* (2006) Functional CpG methylation system in a  
1866 social insect. *Science* **314**, 645.
- 1867 Werren JH, Richards S, Desjardins CA, *et al.* (2010) Functional and evolutionary  
1868 insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**,  
1869 343-348.
- 1870 Wheeler TJ, Eddy SR (2013) nhmmer: DNA homology search with profile HMMs.  
1871 *Bioinformatics* **29**, 2487-2489.
- 1872 Wickham H (2007) Reshaping data with the reshape package. *Journal of Statistical  
1873 Software; Vol 1, Issue 12 (2007)*.
- 1874 Wickham H (2009) *Ggplot2 elegant graphics for data analysis* Springer, New York.
- 1875 Windsor DA (1998) Controversies in parasitology: Most of the species on Earth are  
1876 parasites. *International Journal for Parasitology* **28**, 1939-1941.
- 1877 Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of  
1878 genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet*  
1879 **4**, e1000212.
- 1880 Yamamoto D (2008) Brain sex differences and function of the *fruitless* gene in  
1881 *Drosophila*. *Journal of Neurogenetics* **22**, 309-332.
- 1882 Yin C, Li M, Hu J, *et al.* (2018) The genomic features of parasitism, polyembryony  
1883 and immune evasion in the endoparasitic wasp *Macrocentrus cingulum*. *BMC  
1884 Genomics* **19**, 420.
- 1885 Zepeda-Paulo F, Lavandero B, Mahéo F, *et al.* (2015) Does sex-biased dispersal  
1886 account for the lack of geographic and host-associated differentiation in  
1887 introduced populations of an aphid parasitoid? *Ecology and Evolution* **5**, 2149-  
1888 2161.
- 1889 Zepeda-Paulo FA, Ortiz-Martínez SA, Figueroa CC, Lavandero B (2013) Adaptive  
1890 evolution of a generalist parasitoid: implications for the effectiveness of  
1891 biological control agents. *Evolutionary Applications* **6**, 983-999.

1892 Zhao C, Escalante Lucio N, Chen H, *et al.* (2015) A massive expansion of effector  
1893 genes underlies gall-formation in the wheat pest *Mayetiola destructor*. *Current*  
1894 *Biology* **25**, 613-620.

1895  
1896

1897

1898

1899

1900

1901

1902

1903

1904

1905

1906

1907

1908

1909

1910

1911

1912

1913

1914

1915

1916

1917

1918 **FIGURES IN MAIN TEXT (title)**

1919

1920 Figure 1. **Aphid parasitoid life cycle:** Generalized life cycle of *Aphidius ervi* and  
1921 *Lysiphlebus fabarum*, two different parasitoid wasps that target aphid hosts.

1922

1923 Figure 2. **Codon usage in predicted genes:** Proportions of all possible codons, as used in  
1924 the predicted genes in *A. ervi* (top) and *L. fabarum* (bottom). Codon usage was  
1925 measured as relative synonymous codon usage (RSCU), which scales usage to the  
1926 number of possible codons for each amino acid (RSCU). Codons are listed at the bottom  
1927 and are grouped by the amino acid that they encode. The green line depicts GC content  
1928 (%) of the codon.

1929

1930 Figure 3. **GC and nitrogen content of expressed genes:** We observe significant  
1931 differences (p-values from two-sided t-test) in the GC content between adult and larval  
1932 *L. fabarum* in: (A) the most highly expressed 10% of the genes and (B) genes that are  
1933 differentially expressed between adults and larvae. In contrast, there is no difference  
1934 in the nitrogen content of the same set of genes (C, D).

1935

1936 Figure 4. **Overlap in Venom proteins between *A. ervi* and *L. fabarum*:** Overlap in venom  
1937 proteins (A) and venom protein putative function (B) between *A. ervi* and *L. fabarum*

1938

1939 Figure 5: **Phylogeny of hymenopteran GGT sequences.** *A. ervi*/*L. fabarum* and *N.*  
1940 *vitripennis*/*P. puparum* venom GGT sequences are marked with blue and orange  
1941 rectangles respectively. Letters A, B and C indicate the major clades observed for  
1942 hymenopteran GGT sequences. Numbers at corresponding nodes are aLRT values.  
1943 Only aLRT support values greater than 0.8 are shown. The outgroup is human GGT6  
1944 sequence.

1945

1946

1947

1948

1949

1950

1951

1952

1953

1954

1955

1956 List of additional data files

1957

1958 **Additional Data 1:** details of genetic positions used to construct linkage groups for *L.*

1959 *fabarum*.

1960 **Additional Data 2:** Genbank numbers and taxa information for genome (CDS) graphed

1961 in Supplemental Figure 6.

1962 **Additional Data 3:** file detailing (a) the most highly expressed genes in both taxa and

1963 (b) differential expression between adult and larval *L. fabarum*.

1964 **Additional Data 4:** fasta file of orphan genes for *A. ervi*

1965 **Additional Data 5:** fasta file of orphan genes for *L. fabarum*

1966 **Additional Data 6:** Summary of OMA output, including details of *LRR* genes

1967 **Additional Data 7:** Annotation of venom genes in *L. fabarum* and *A. ervi*

1968 **Additional Data 8:** Details of immune gene annotation

1969 **Additional Data 9:** Expression details of Osiris genes in *L. fabarum* and *A. ervi*

1970 **Additional Data 10:** Details of annotated OXPHOS genes, including duplications in the

1971 assembly

1972 **Additional Data 11:** Details of sex determination gene annotations

1973

1974

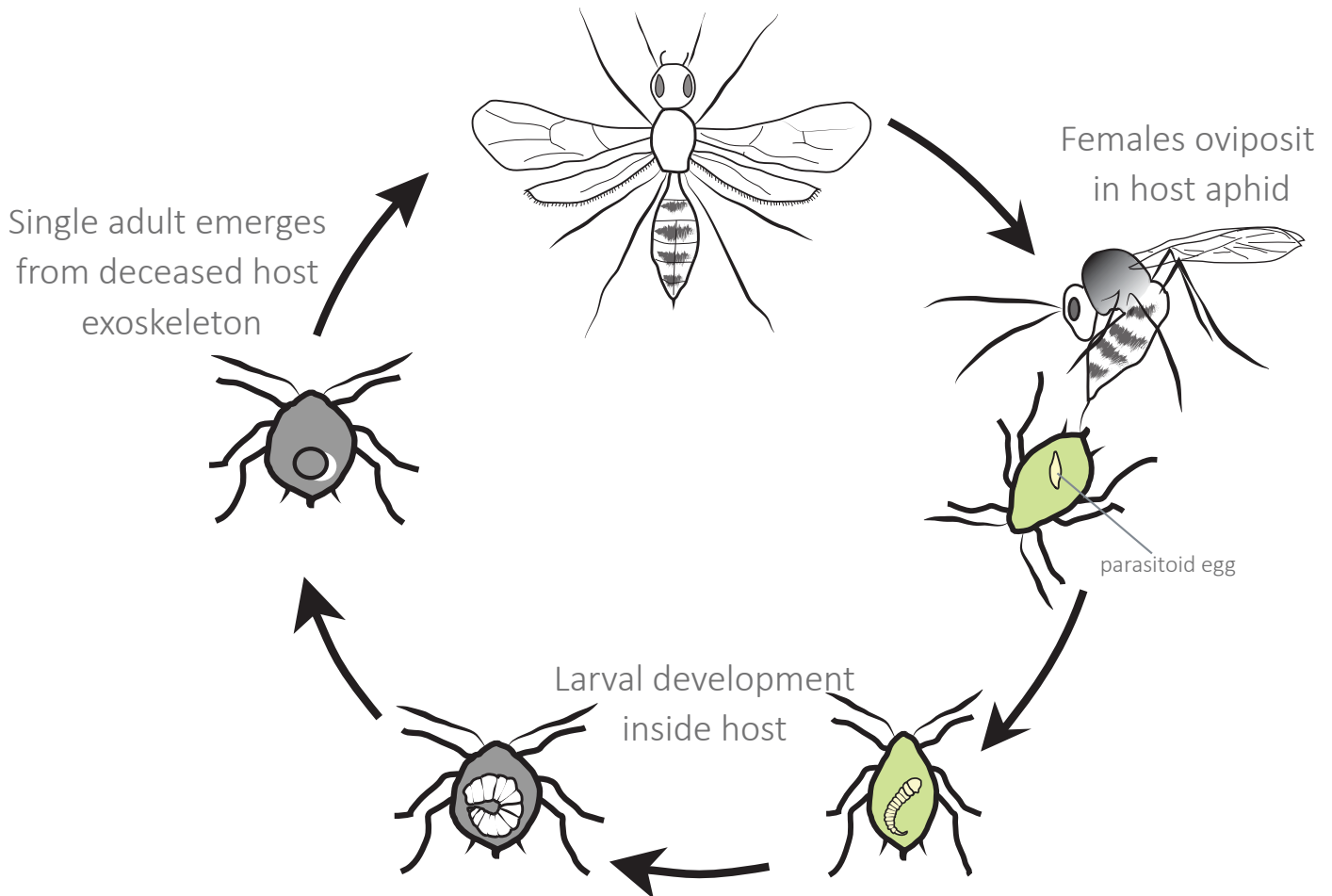
1975

1976

1977

## Figure 1

### Life cycle of aphid parasitoids



### Life history characteristics

	<i>Aphidius ervi</i>	<i>Lysiphlebus fabarum</i>
Host insects	Aphididae	Aphididae
Reproductive mode	Sexual	Asexual or sexual
Host is ant tended	No	Yes, usually
Native range	Europe	Europe
Primary host aphid tribe	Macrosiphini	Aphidini



Figure 2

Codon usage and GC content

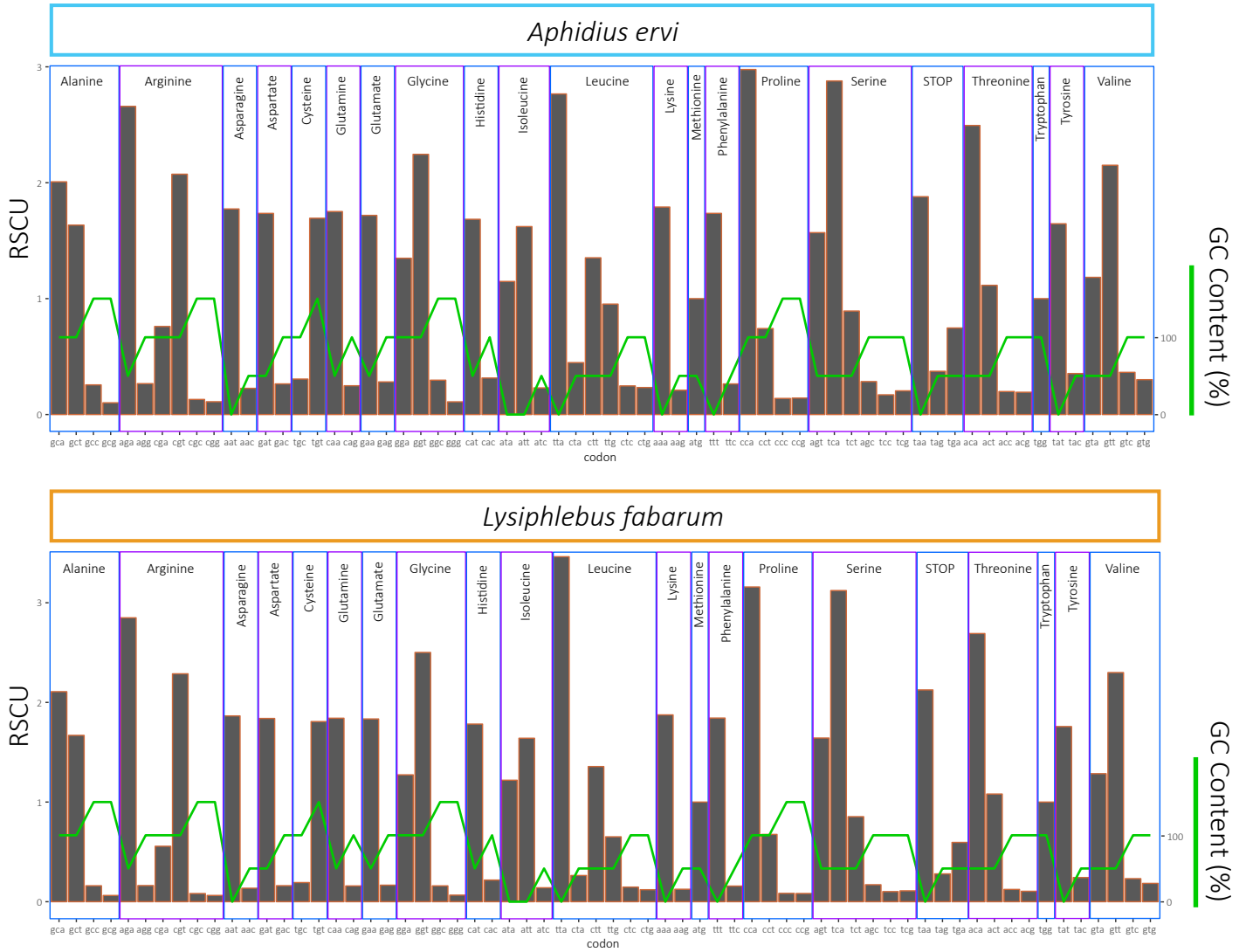


Figure 3

# Genes expressed in larval and adult *L. fabarum*

Highly expressed genes

Differentially expressed genes

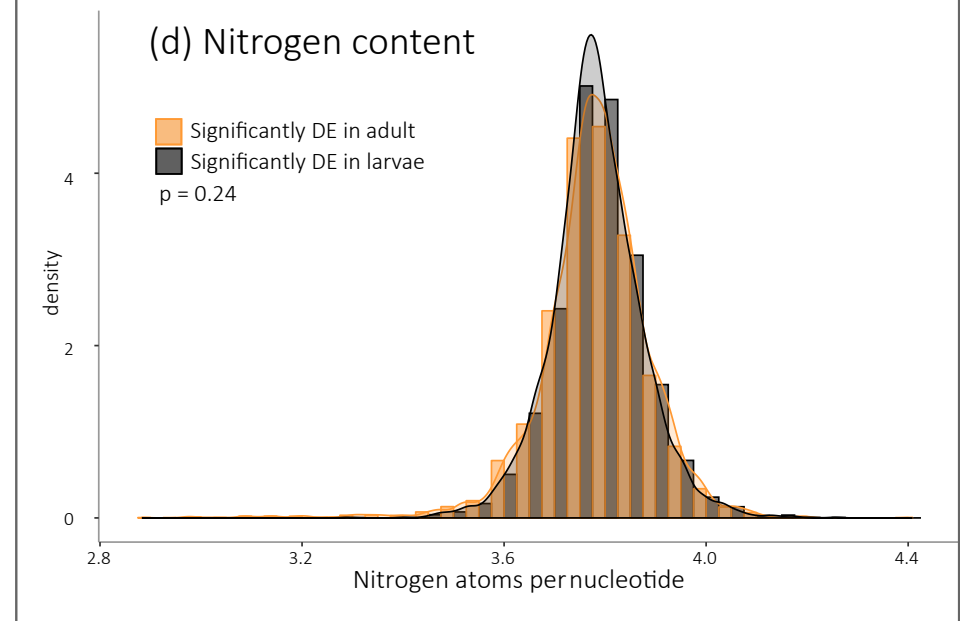
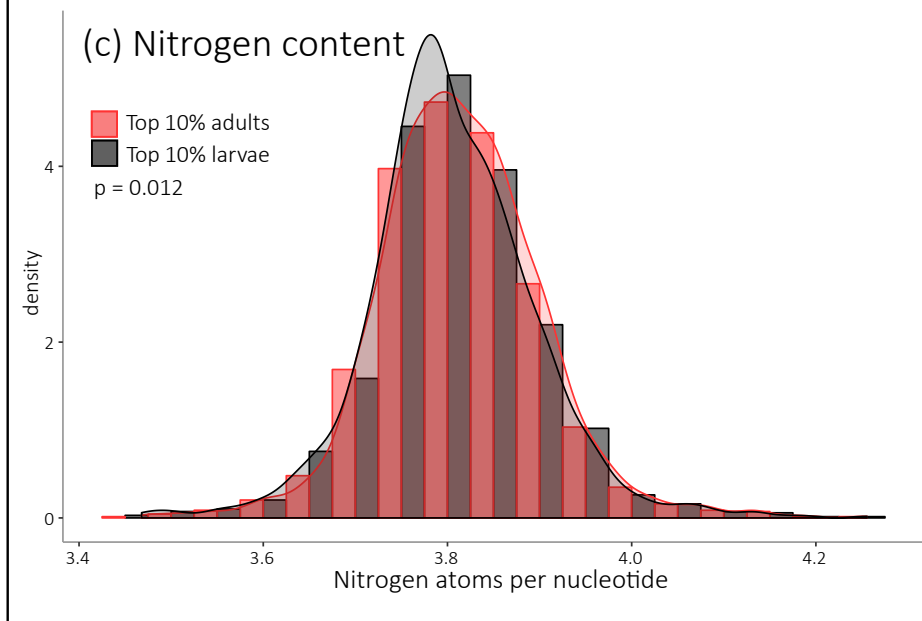
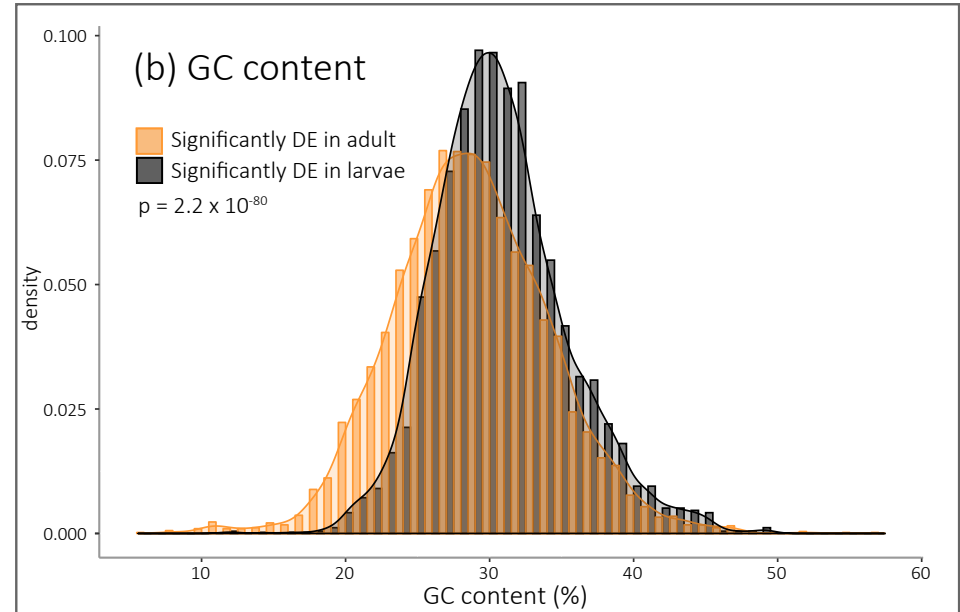
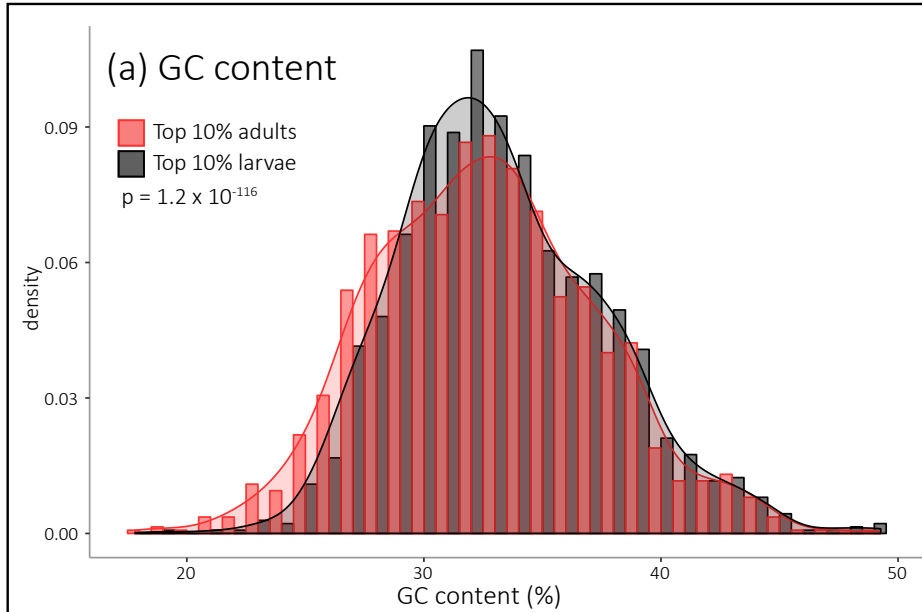
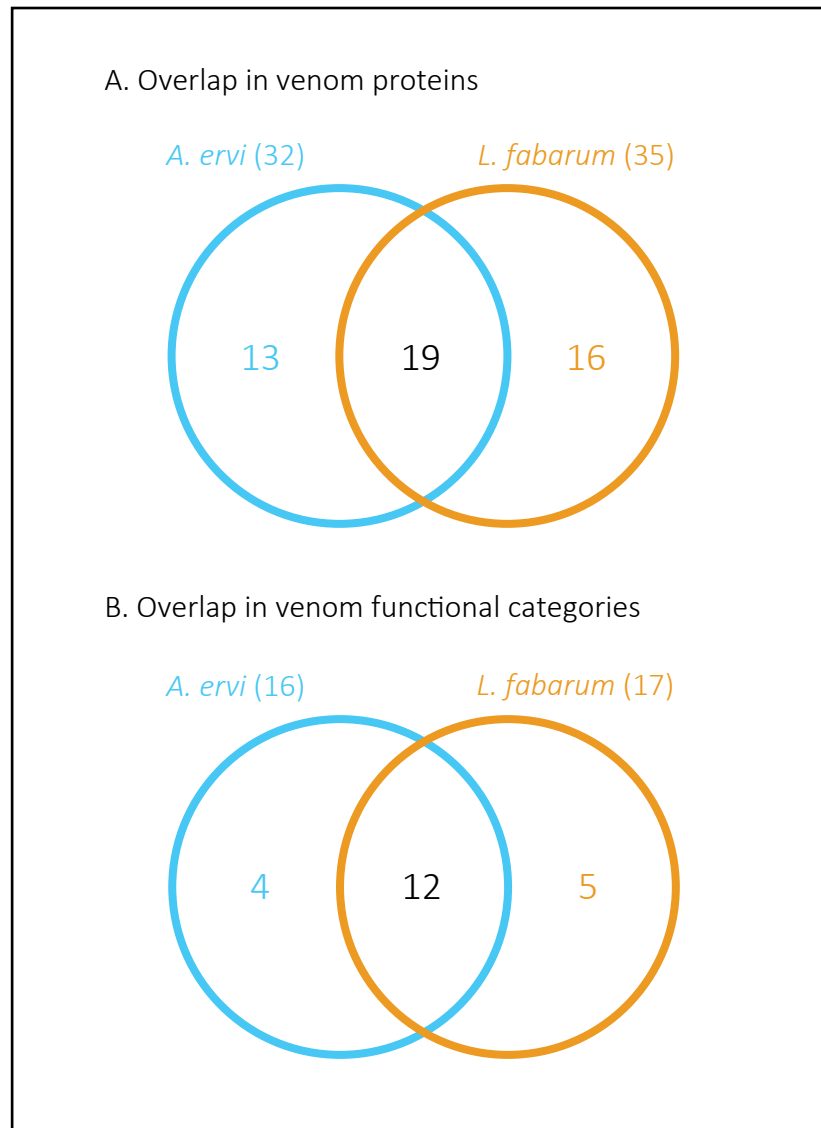


Figure 4



## Figure 5

