



HAL
open science

Detection and decay rates of prey and prey symbionts in the gut of a predator through metagenomics

Débora P. Paula, Benjamin Linard, David A Andow, Edison R Sujii, Carmen S S Pires, Alfried P. Vogler

► **To cite this version:**

Débora P. Paula, Benjamin Linard, David A Andow, Edison R Sujii, Carmen S S Pires, et al..
Detection and decay rates of prey and prey symbionts in the gut of a predator through metagenomics.
Molecular Ecology Resources, 2015, 15 (4), pp.880-892. 10.1111/1755-0998.12364 . hal-01636897

HAL Id: hal-01636897

<https://hal.science/hal-01636897v1>

Submitted on 17 Jan 2019

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.

1 MOLECULAR ADVANCES

2 **Detection and decay rates of prey and prey symbionts in the gut of a predator through**
3 **metagenomics**

4
5
6
7 **Running title:** Prey and symbiont detection by metagenomics

8
9 DEBORA P. PAULA,^{*†} BENJAMIN LINARD,[†] DAVID A. ANDOW,[‡] EDISON R. SUJII,^{*}
10 CARMEN S. S. PIRES,^{*} ALFRIED P. VOGLER^{†\$}

11
12 **Embrapa Genetic Resources and Biotechnology, Parque Estação Biológica, W5 Norte, P.O.*
13 *Box 02372, Brasília, DF, 70770-917, Brazil, [†]Department of Life Sciences, Natural History*
14 *Museum, Cromwell Rd, London, SW7 5BD, UK,[‡] Department of Entomology, University of*
15 *Minnesota, 219 Hodson Hall, 1980 Folwell Ave., St. Paul, MN 55108, USA; ^{\$} Department of*
16 *Life Sciences, Imperial College London, Silwood Park Campus, Ascot, SL7 5PY, UK*

17
18 Correspondence: D. P. Paula, Fax: +55 (61) 34484929;

19 E-mail: debora.pires@embrapa.br

20
21 Keywords: analyte detectability half-life, analyte detectability period, gut content, mtDNA,
22 nuclear genome, aphid, coccinellid

24 **Abstract**

25 DNA methods are useful to identify ingested prey items from the gut of predators, but reliable
26 detection is hampered by low amounts of degraded DNA. PCR-based methods can retrieve
27 minute amounts of starting material but suffer from amplification biases and cross-reactions
28 with the predator and related species genomes. Here, we use PCR-free direct shotgun
29 sequencing of total DNA isolated from the gut of the harlequin ladybird *Harmonia axyridis* at
30 five time points after feeding on a single pea aphid *Acyrtosiphon pisum*. Sequence reads
31 were matched to three reference databases: Insecta mitogenomes of 587 species, including *H.*
32 *axyridis* sequenced here; *A. pisum* nuclear genome scaffolds; and scaffolds and complete
33 genomes of 13 potential bacterial symbionts. Immediately after feeding, multicopy mtDNA of
34 *A. pisum* was detected in tens of reads, while hundreds of matches to nuclear scaffolds were
35 detected. Aphid nuclear DNA and mtDNA decayed at similar rates (0.281 and 0.11 h⁻¹
36 respectively), and the detectability periods were 32.7 and 23.1 h. Metagenomic sequencing
37 also revealed thousands of reads of the obligate *Buchnera aphidicola* and facultative *Regiella*
38 *insecticola* aphid symbionts, which showed exponential decay rates significantly faster than
39 aphid DNA (0.694 and 0.80 h⁻¹ respectively). However, the facultative aphid symbionts
40 *Hamiltonella defensa*, *Arsenophonus* spp. and *Serratia symbiotica* showed an unexpected
41 temporary increase in population size by 1-2 orders of magnitude in the predator guts before
42 declining. Metagenomics is a powerful tool that can reveal complex relationships and the
43 dynamics of interactions among predators, prey and their symbionts.

44 **Introduction**

45 Molecular gut content analysis has been used to identify the prey consumed by invertebrate
46 predators, allowing the study of specific trophic interactions that naturally occur in the field
47 (Pompanon *et al.* 2012, Greenstone *et al.* 2014). Various approaches have been developed to
48 assess the presence of target prey remaining in predator guts via protein-based analyses (e.g.
49 by isoenzymes electrophoresis, ELISA, Western-Blot) or DNA-based analyses (e.g. by PCR,
50 and qPCR) (Symondson 2002; Hardwood & Obrycki 2005; Greenstone *et al.* 2007; Weber &
51 Lundgren 2009; Zeale *et al.* 2010). These molecular tools require the development of species-
52 specific antibodies or DNA primers for amplification of target genes, or time-consuming
53 cloning of PCR products and subsequent Sanger sequencing. Despite their great contribution
54 to contemporary studies of trophic interactions due to their high specificity and sensitivity,
55 they are limited to detect a few target prey molecules.

56 Since the advent of high-throughput DNA sequencing, diet analyses based on feces
57 have been assessed in several mammals, birds and insects through barcode region sequencing,
58 known as metabarcoding (Valentini *et al.* 2009a; 2009b; Deagle *et al.* 2010; Hereward &
59 Walter 2012; Pompanon *et al.* 2012; Vesterinen *et al.* 2013). In these studies, DNA barcodes
60 allow the detection of a spectrum of species against a set of DNA reference sequences,
61 without need of cloning PCR products. However, despite being less time-consuming and very
62 sensitive, there are still limitations, such as the need to design taxon-specific or group-specific
63 primers that avoid predator DNA amplification (Jarman *et al.* 2004; Deagle *et al.* 2005), or to
64 digest or block predator template DNA (Green & Minz 2005; Vestheim & Jarman 2008;
65 Deagle *et al.* 2009, 2010; Shehzad *et al.* 2012; but see Piñol *et al.* 2014). In addition,
66 problems with non-target template amplification (Zeale *et al.* 2010) or cross-amplification
67 when predator and prey are phylogenetically close (Thomas *et al.* 2012) need to be addressed.

68 Further, the amplification of target DNA limits the study of the decay dynamics of DNA
69 inside the predators because of the difficulties of quantifying the amount of starting material
70 with the PCR procedure, and because of the focus on a single gene region.

71 Shotgun sequencing of total DNA extracted from the gut or even feces is an
72 alternative approach that, compared to PCR-based (meta)barcoding, provides a broader
73 taxonomic range of target organisms (Srivathsan *et al.* 2014). It could also be used to study
74 the symbiont communities closely associated with a prey (Oliver *et al.* 2010) and the
75 dynamics of their interactions. Conceivably, total DNA extraction from the gut content of a
76 predator, followed by direct sequencing of any identifiable DNA fragment from the prey and
77 from its associated symbionts could enlarge even further the spectrum of species detection.
78 The number of genomes (nuclear or mitochondrial) elucidated and available in public
79 databanks is increasing rapidly or can be readily generated, and these could be used as a
80 reference to match the sequenced DNA fragments to identify prey. In that sense, three sources
81 of DNA could possibly be used to identify the prey spectrum without need of genetic
82 amplification: the prey nuclear and mitochondrial genomes and the genomes of its associated
83 symbionts. As some symbionts are prey specific (Oliver *et al.* 2010), their detection could
84 indicate or support the identification of the prey.

85 It is widely agreed that prey DNA susceptibility to predator digestion (Harwood &
86 Obrycki 2005; Greenstone *et al.* 2007; Weber & Lundgren 2009) and the molecular technique
87 (Greenstone *et al.* 2014) used for prey detection are important factors influencing the
88 sensitivity of prey detection. Consequently, the prey detection system proposed here based on
89 the detection of any part of the prey genomes (and on associated symbionts) and on shotgun
90 sequencing of the DNA in the predator gut should be investigated more deeply by, for
91 example, estimating the DNA decay rate and detectability period. These decay parameters

92 indicate how long prey can be detected according to the speed and DNA susceptibility to
93 digestion, providing a basis for comparison with other molecular techniques.

94 This study aimed to test the detection of prey nuclear and mitochondrial genomes and
95 bacterial symbiont genomes through a direct metagenomic approach without any
96 amplification of genetic material, based on a feeding experiment with pea aphid *Acyrtosiphon*
97 *pisum* (Hemiptera: Aphididae) in the gut of the widely invasive, aphidophagous harlequin
98 ladybird beetle *Harmonia axyridis* (Coleoptera: Coccinellidae). We used pea aphid as a prey
99 because its genomes have been elucidated and its associated symbionts are well characterized,
100 which enabled the study of the fate of various bacterial symbionts after prey ingestion. These
101 symbionts include the obligate *Buchnera aphidicola* and several facultative symbionts, such
102 as *Regiella insecticola*, *Hamiltoniella defensa* and *Serratia symbiotica*, in addition to other
103 known aphid symbionts, e.g., *Arsenophonus* (Oliver *et al.* 2010).

104 In addition, this study aimed to characterize the specificity and sensitivity of prey
105 detection using the proposed methodology, and estimate the DNA decay rate, half-life, and
106 detectability period. As prey items are ingested with their entire associated microbial and
107 parasite community, the analysis of these components potentially can provide additional
108 information on the fate of the prey and the impact of the feeding event on the predator.

109 **Material and methods**

110 *Insects and description of the study system*

111 *Harmonia axyridis* (Coleoptera: Coccinellidae) was used as a model for prey detection in a
112 common worldwide aphidophagous predator. Pupae (over 600) were collected in August
113 2013 (summer) in soybean fields in St. Paul, Minnesota-USA. Upon emergence adults were
114 transferred to individual petri dishes (35x10 mm) with moistened filter paper and held under
115 controlled conditions (25°C and 16:8 h L:D cycle) without food. After 24 h post-emergence,
116 the individuals were used in the feeding bioassay.

117 Pea aphid, *Acyrtosiphon pisum* (Hemiptera: Aphididae), were used as a prey model
118 because it has both mitogenome (GenBank gi|213948225|ref|NC_011594.1|) and nuclear
119 genome (GenBank Assembly ID: GCA_000142985.2) elucidated (Richards *et al.* 2010).
120 Additionally, it is the best studied aphid regarding symbionts (Oliver *et al.* 2010). Adults were
121 obtained from a laboratory colony collected from North Dakota, USA, containing unidentified
122 symbionts. Soybean aphids, *Aphis glycines* (Hemiptera: Aphididae), were obtained daily from
123 the same soybean field where the *H. axyridis* pupae were collected.

124 The presence of symbionts was tested against the genomes of the genera
125 *Arsenophonus*, *Buchnera*, *Hamiltonella*, *Regiella*, *Rickettsia*, *Rickettsiella*, *Serratia*,
126 *Spiroplasma* and *Wolbachia*. These genera were chosen because either they are known to
127 confer fitness advantages and costs to aphids (Wille & Hartman 2009, Oliver *et al.* 2010,
128 Jones *et al.* 2011, Jouselin *et al.* 2012), or to coccinellids (Majerus 2006; Weinert *et al.*
129 2007). *Buchnera* is an obligate symbiont occurring in high numbers in specialized host
130 organs. Except for *Arsenophonus*, all of these symbionts have been reported in *A. pisum*
131 (Simon *et al.* 2011; Russell *et al.* 2013). In addition, *Nosema* was included, as it could be
132 associated to *H. axyridis* (Vilcinskas *et al.* 2013), and three insect non-aphid and non-

133 coccinellid symbionts, *Blattabacterium*, *Cardinium*, and *Midichloria*, were included as false
134 positive controls (Fein-Zchori & Bourtzis 2012).

135

136 *Feeding bioassay*

137 To estimate the decay of the prey using metagenomics in the predator gut after consumption,
138 a feeding bioassay was conducted. Freshly emerged unfed adults were used because the gut
139 would be totally empty, they would have the same age and physiological state, and it would
140 avoid potential complications from secondary predation and scavenging. In addition,
141 preliminary observations indicated that adults did not readily consume prey during the first 24
142 h post-eclosion. The 24-h-old beetles were individually supplied with a single *A. pisum* adult.

143 At six time points, immediately before feeding (negative control, denoted “Pre”), 0 h
144 (immediately after feeding), 3, 24, 48 and 96 h after the target-species consumption, batches
145 of 10 beetles were harvested and stored at -80°C in 100% ethanol. These time points were
146 chosen because they contain the minimum and maximum interval time of detection currently
147 reported in the literature for detection of a prey target molecule (protein and DNA)
148 (Greenstone *et al.* 2014). Four hours after pea aphid consumption, *Aph. glycines* were offered
149 once a day as a sustaining food to *H. axyridis* adults, until the last time point of the bioassay.

150

151 *DNA sample preparation*

152 The guts of the preserved predators were dissected out using clean forceps under a
153 stereomicroscope in order to increase the chances of detecting prey DNA in the sample. Guts
154 from the same time point were pooled into one sample. The total DNA of each sample was
155 extracted with a DNeasy Blood and Tissue kit (Qiagen, Hilden-Germany) and quantified by
156 fluorescence using the Qubit system (Invitrogen™) after quality checking

157 spectrophotometrically (ratio $A_{260/280}$ nm). The total DNA concentration of each sample was
158 normalized to 20 ng/ μ L and sonicated to construct TruSeq libraries of insert size of 450 bp
159 (250 bp paired-end, 500 cycle kit). Each library was sequenced on MiSeq-Illumina using 17%
160 of the flowcell.

161

162 *Sequence quality controls*

163 The quality assessment of raw sequence data was made using FastQC
164 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and PrinSEQ (Schmieder &
165 Edwards 2011) with a minimum quality score of 20, maximum ambiguous base N of 0 and
166 trim quality from the right (3') to minimum of 20. Overrepresented sequences (e.g. library
167 index adapters) were trimmed with Trimmomatic (Lohse *et al.* 2012). The scripts used for the
168 main analyses are presented in the Supporting Information (SI) section.

169

170 *Predator mitogenome assembly*

171 For the elucidation of the *H. axyridis* mitogenome, first the reads were filtered for similarity
172 of E-value $< 10^{-5}$ with NCBI Insecta mitogenomes that included partial and complete
173 sequences of 587 species (taxonomic ID: 50557) using the BLASTn algorithm (Altschul *et al.*
174 1990). Filtering simplifies the assembly by reducing the size of the dataset and enriching it
175 with putative mitochondrial reads. The retained reads were assembled using Celera (Myers *et*
176 *al.* 2000) and IDBA-UD (Yu *et al.* 2011), and for the latter after quality control by PrinSEQ
177 (Schmieder & Edwards 2011) with a minimum quality score and mean of 20, maximum
178 ambiguous base N of 0 and trim quality from the right (3') to a minimum of 20. The scaffolds
179 generated by both assemblers were concatenated in Geneious 7.0.5 (Biomatters,
180 <http://www.geneious.com/>) using the parameters: no gaps allowed, minimum overlap 150,

181 maximum mismatches per read 0, minimum overlap identity 99%, maximum ambiguity 1.
182 The mitogenome annotation was made by first annotating the tRNA genes using models
183 based on the NCBI Insecta mitogenomes and the COVE software package (Eddy & Durbin
184 1994). The annotation process was finished manually in Geneious 7.0.5. The nearly complete
185 mitogenome sequence of 15,322 bp includes the expected two rRNA, 22 tRNA and 13 protein
186 coding genes arranged in the canonical gene order of Coleoptera (Timmermans & Vogler
187 2012). The control region was not completely sequenced. The mitogenome was deposited at
188 GenBank under the accession code KJ778886.

189

190 *Identification of aphid mtDNA*

191 Good quality reads from all time points were matched to the NCBI Insecta mtDNA reference
192 database of 587 species, including pea aphid and five other aphid species (November 2013),
193 added to the sequenced mitogenome of *H. axyridis*. The matches were made by BLASTn with
194 an E-value $< 10^{-5}$. Custom scripts (Supporting Information) were used to associate the
195 GenBank general identification (gi) number and its taxonomic identification with reads
196 obtained by Illumina sequencing requiring sequence identity $>98\%$ over a minimum hit
197 length of 225 bp (90% of read length). A species match was retained when it equaled or
198 exceeded the thresholds for minimum length and identity. Preliminary analyses using lower
199 identity thresholds indicated that all false positives and ambiguous identifications were
200 eliminated at 98% identity. Many of these false positives were repetitive DNA with high AT
201 content. The matched pea aphid mtDNA reads were mapped onto the prey mitogenome using
202 Geneious 7.0.5 to evaluate the overall coverage of the mtDNA. The map position of reads on
203 the mitochondrial *A. pisum* genome was tested for a random distribution using the Poisson
204 Dispersion test.

205

206 *Identification of aphid nuclear DNA*

207 For each time point, nuclear reads from the guts were identified by MegaBLAST alignment to
208 the *A. pisum* complete genome (assembly Acyr_2.0; placed and unplaced scaffolds; GenBank
209 Assembly ID: GCA_000142985.2) (Richards *et al.* 2010). All of the reads that matched with
210 the *A. pisum* nuclear genome with more than 245 bp of >99% sequence identity and E-value <
211 10^{-9} were examined. Nuclear repeat regions gave ambiguous species identifications, so the
212 ones in *A. pisum* (including short sequence repeats-SSR), were identified and excluded with
213 the following filters. First, we identified all rRNA reads by BLAST to the rRNA SILVA
214 database (Quast *et al.* 2013) and discarded them. To complement the filtering of possible non-
215 specific reads, we submitted the remaining aphid reads to the RepeatMasker pipeline (Tarailo-
216 Graovac & Chen 2009). It first uses the TandemRepeatFinder program to detect simple
217 tandem repeats (Benson 1999). Next, all sequences were compared to two databases of
218 currently known structured repeats; the RepBase database specialized on repeat definitions
219 (Jurka *et al.* 2005) and the Dfam database of repeat HMMs (Hidden Markov Models)
220 (Wheeler *et al.* 2012). All reads containing potential non-specific SSR or microsatellites were
221 also discarded. Finally, the filtered reads were aligned to the whole content of the NCBI
222 Refseq Protein database with BLASTx. All translated reads matching a protein sequences
223 associated to the pea aphid (taxonomy id:7029), with >90% sequence identity over more than
224 30 amino acids are considered as potential coding gene hits.

225

226 *Identification of prey-associated symbionts*

227 Thirteen bacterial genera with known insect symbiotic interactions were used to create a
228 database of symbiont sequences. For each genus, we retrieved all available NCBI Genbank

229 sequences and complete genomes to build the database (Table S1, Supporting Information).
230 DNA reads from each sample tested for the presence of these symbionts were aligned to this
231 database with MegaBLAST and all reads aligned over > 225 bp with >95% sequence identity
232 and E-value <10⁻⁹ were retained. Reads similar to the conservative rRNA sequences were
233 removed to avoid misidentification due to insufficient sequence variability between related
234 genera. The thresholds used discarded several reads that could be associated with one of the
235 studied genera, but the need to discriminate several genera in a bacterial family (e.g.
236 Enterobacteriaceae) required such a measure. The number of available reference genomes
237 used to identify symbiont reads differed for each genus, which may affect the capacity for
238 species detection (Table S1).

239

240 *Statistical analyses to estimate decay parameters*

241 An advantage of the metagenomics method is that the number of prey reads detected in the
242 predator guts can be used to estimate the dynamics of analyte detectability. Although
243 considerable work has been done with detectability half-lives, little use has been made of
244 analyte detectability parameters (Greenstone *et al.* 2014). Here we provide methods for
245 estimating three analyte degradation parameters: analyte decay rate, analyte detectability half-
246 life, and analyte detectability period (D_{max}). Two critical points must be kept in mind. First the
247 detection of a prey or symbiont read is a stochastic process that combines random events
248 associated with a) the subsample of the total DNA in the gut sample, and b) the subsample of
249 reads sequenced from the resulting DNA library. This means that the number of reads
250 observed at any time point is a random variable, and there is some probability that the actual
251 number of reads was greater (or less) than the number observed. Specifically, a time point
252 with zero observed reads must be treated as a random zero (i.e., there could have been one or

253 more reads in the original sample, but the sampling and sequencing processes did not retain
254 any of these reads), and not a true biological zero (i.e., there were no reads in the original
255 sample), and is an important and meaningful datum. [Although similar random processes
256 occur in PCR-based methods, in these methods, the sources of randomness simply add
257 variance to the estimated probability of detecting a positive individual.] Second, because the
258 bioassay used different individuals to evaluate digestion at each time point, the time points are
259 statistically independent samples of the digestion process.

260 Treating digestion as a stochastic process makes explicit the uncertainty associated
261 with the observed data. Assuming that all reads have the same probability of detection, the
262 number of observed reads at each time point will follow a Poisson process. Further, the
263 observed numbers can be used to estimate the underlying Poisson process and calculate the
264 probability distribution for possible observed read numbers. This is done using Bayesian
265 methods with a Jeffries prior. Bayesian methods were used because the observed read
266 numbers are single realizations of the underlying random process. The number of reads and
267 the Bayesian posterior distribution were normalized by the library size at each time point, and
268 multiplied by 10^6 for presentation purposes. Estimation of the detectability half-life usually
269 assumes a first-order or exponential decay in the quantity of target DNA or protein degraded
270 (Lovei *et al.* 1985; Sopp & Sunderland 1989; Weber & Lundgren 2009). Therefore, expected
271 values from the posterior distributions (not the observed numbers) were used to fit an
272 exponential decay model by non-linear regression. This initial analysis allowed the
273 identification of species and genera that did not decay exponentially in the predator guts.

274 For those that did decay exponentially, Monte Carlo simulation was used to estimate
275 the analyte decay parameters. Three parameters were estimated: a) instantaneous decay rate of
276 the DNA (analyte decay rate), b) analyte detectability half-life, and c) the maximum period

277 during which DNA could be detected (analyte detectability period, which is analogous to
278 D_{\max} , Sutherland *et al.* 1987). Random read numbers were drawn from the normalized
279 posterior distributions for each time point, an exponential decay model was fit to these values
280 by non-linear regression, and the estimated parameter values (analyte decay rate and initial
281 number of reads) were saved. This was repeated 200,000 times to generate a joint probability
282 distribution function (jpdf) of the two parameter values. The analyte decay rate and its 95%
283 CI were estimated from the marginal distributions of the jpdf. The analyte detectability half-
284 life and its 95% CI were estimated from the inverse of the decay rate. The jpdf was also used
285 to estimate the 95% confidence region of the model parameters, and the border of this region
286 was used to estimate the 95% confidence envelope of the non-linear regression. Analyte D_{\max}
287 and its 95% CI were estimated using the original read numbers, the analyte decay rate, and the
288 95% confidence envelope of the regression to calculate the time when only one read would be
289 left. A similar method was used to estimate D_{\max} from the original data published in
290 McMillan *et al.* (2007), Kuusk *et al.* (2008) and Kerzicnik *et al.* (2012), who studied the
291 detectability of single aphid prey using PCR. In these cases, we calculated the time when
292 only one individual would test positive. All calculations were done in Mathematica 7.0.

293 **Results**

294 *Library basic statistics and recovery of predator DNA*

295 Each of the six Illumina libraries was made from the guts of 10 individuals of *H. axyridis* and
296 corresponded to different time points after feeding on *A. pisum*. These had similar DNA
297 concentrations and produced similar total number of reads (Table 1). Many thousands of reads
298 in each library showed exact matches to *H. axyridis* mtDNA, and their number broadly
299 covaried with the total number of reads in each library. Reads matching mtDNA could be
300 assembled to recover the mitogenome of *H. axyridis*, although read coverage was not uniform
301 and was low in some intergenic regions (Fig. S1, Supporting Information). As non-predator
302 reads, we detected *A. pisum* and some bacterial aphid symbionts after predator feeding,
303 detailed below, and no other species were detected.

304

305 *Prey detection and decay parameters*

306 *a) mtDNA*

307 Twenty-three reads were identified as *A. pisum* mtDNA (Table 2). As expected, there was no
308 *A. pisum* mtDNA in the negative control, i.e. before the predator has fed. Aphid mtDNA
309 detection occurred immediately (0 h) and 3 h after feeding, and more prey sequences were
310 detected earlier than later. The *A. pisum* reads covered different regions of the mitogenome
311 (Fig. 1). The majority of the genes had matches to a single read only, but some genes were
312 repeatedly hit. The *cox1* gene was detected only once, in the sample obtained immediately
313 after feeding.

314 The decay of the mtDNA for a single *A. pisum* in *H. axyridis* fit the first order
315 exponential decay model extremely well ($p = 1.94 \times 10^{-3}$) with an adjusted $r^2 = 0.974$ (Fig.
316 2A). On average, the instantaneous analyte decay rate was 0.11 reads per hour with 95% CI of

317 0.05 to 0.30 h⁻¹. The analyte detectability half-life was 8.9 h with 95% CI of 3.3 to 18.3 h.
318 The analyte D_{\max} to detect a single *A. pisum* read based on mtDNA, was 23.1 h with 95% CI
319 of 9.5 to 81.4 h.

320

321 *b) Nuclear genome*

322 The number of reads with matches to the *A. pisum* nuclear genome exceeded the mtDNA
323 reads by a factor of about 30, reaching over 500 reads at the moment of feeding (Table 2). No
324 aphid sequences were detected in the pre-feeding negative control. Aphid nuclear DNA
325 detection continued for all time points, including the last one at 96 h after feeding. The latter
326 was due to the recovery of three reads, which was unexpected given the already very low
327 counts at the two earlier time points. Over the hundreds of reads showing a nearly perfect
328 match in the pea aphid genome at 0 h, 48 matched 29 different pea aphid protein sequences
329 retrieved from the NCBI RefseqP database (Table S2). Similarly, at 3 h 13 reads matched 9
330 different aphid proteins. In many cases, both reads of the same pair matched the same aphid
331 protein. Many aphid proteins are computational predictions based on the pea aphid genomes
332 (“uncharacterized” and “predicted” annotations) but they were nevertheless the closest hit in
333 the database (which includes proteins from all domains of life). Some matches seem to be
334 linked to integrated viral genomes (XP_008184955.1, an HIV Tat-specific factor-like
335 element), but we also uncovered genes linked to specific functions. For instance, one of the
336 reads matched an O-linked-mannose beta-1,2-N-acetylglucosaminyltransferase
337 (XP_001948219.2, Table S2), a protein with a domain signature (NCBI domain cd13937)
338 conserved in animals.

339 The decay of the nuclear DNA for a single *A. pisum* in *H. axyridis* fit the first order
340 exponential decay model extremely well ($p = 1.07 \times 10^{-5}$) with an adjusted $r^2 = 0.999$ (Fig.

341 2B). On average, the instantaneous analyte decay rate was 0.281 reads per hour with a 95%
342 CI of 0.225 to 0.338 h⁻¹. The analyte detectability half-life was 3.6 h with 95% CI of 3.0 to
343 4.4 h. The analyte D_{\max} was 32.7 h with 95% CI of 29.8 to 96 h. None of these values were
344 significantly different from the corresponding parameters for *A. pisum* mtDNA, although the
345 D_{\max} was somewhat greater because many more nuclear reads were detected and reads were
346 found at the final sampling time.

347

348 *Detection characterization of prey symbionts*

349 In addition to the detection of aphid nuclear and mitochondrial DNA, we identified reads
350 homologous to known aphid bacterial symbionts, some of them in high numbers (Table 2).
351 The symbionts *Buchnera aphidicola*, *Arsenophonus* spp., *Hamiltonella defensa*, *Regiella*
352 *insecticola*, and *Serratia symbiotica* were detected only after *H. axyridis* feeding, indicating
353 that they were exclusively associated with the ingested pea aphids. The obligate symbiont *B.*
354 *aphidicola* was present in the highest numbers, with an even read sampling over its whole
355 genome, with 1,651 reads at 0 h and 171 reads at 3 h (Fig. S2). Symbionts from the genera
356 *Blattabacterium*, *Cardinium*, *Midichloria*, *Rickettsia*, *Rickettsiella*, *Spiroplasma* and
357 *Wolbachia* were not detected at any time point, and neither was the microsporidian fungus
358 *Nosema*. Reads matching the genomes of the genus *Serratia* (mainly *S. marcescens*) were
359 detected in high numbers at all time points even before feeding, which indicates its
360 association with the predator. In contrast, *S. symbiotica*, which is known to be an aphid
361 symbiont (Lamelas *et al.* 2011), was observed only after feeding on pea aphids at all time
362 points.

363 Two of the detected symbionts decayed according to the exponential decay model, and
364 three of them did not. The first order exponential decay model fit the data for the obligatory

365 aphid symbiont *B. aphidicola* ($p = 2.39 \times 10^{-12}$; adjusted $r^2 = 1.000$) and the facultative *R.*
366 *insecticola* ($p = 6.04 \times 10^{-7}$; adjusted $r^2 = 0.998$) very well. *Buchnera aphidicola* was detected
367 in large numbers immediately after feeding and continued to be detected 24 h later (Table 2).
368 The decay of *B. aphidicola* in *A. pisum* ingested by *H. axyridis* is presented in Fig. 2C. On
369 average, the instantaneous analyte decay rate was 0.694 reads per hour with a 95% CI of
370 0.642 to 0.747 h⁻¹. The mean analyte detectability half-life was 3.4 h with 95% CI of 2.5 to
371 4.1 h. The analyte D_{\max} was 15.4 h with 95% CI of 5.7 to 25 h. Decay of *B. aphidicola* DNA
372 therefore was significantly faster than for either *A. pisum* mtDNA or nuclear DNA as there
373 was no overlap in the 95% CIs.

374 Decay rate was similarly fast for *R. insecticola*, which however was detected in much
375 smaller numbers than *B. aphidicola* (Table 2). On average, the instantaneous analyte decay
376 rate was 0.80 reads per hour with a 95% CI of 0.39 to 1.34 h⁻¹. The analyte detectability half-
377 life was 1.3 h with 95% CI of 0.7 to 2.5 h. The analyte D_{\max} was only 5.1 h with 95% CI of
378 2.4 to 7.6 h (Fig. 2D).

379 The facultative symbionts *H. defensa*, *Arsenophonus* spp., and *S. symbiotica* did not fit
380 the exponential decay model (*H. defensa*: $p = 0.245$, adjusted $r^2 = 0.007$; *Arsenophonus* spp.:
381 $p = 0.185$, adjusted $r^2 = 0.115$; *S. symbiotica*: $p = 0.072$, adjusted $r^2 = 0.407$). All three
382 exhibited a similar pattern, with no or almost no reads detected immediately after feeding, and
383 a large, statistically significant increase in the number detected at 3 h after feeding, followed
384 by a statistically significant decline in detection thereafter (Fig. 3). The rate of analyte decay
385 with 95% CIs from 3 h onwards was 0.25 [0.13, 0.36] h⁻¹ for *H. defensa*, 0.09 [0.02, 0.15] h⁻¹
386 for *Arsenophonus* sp., and 0.04 [0.01, 0.08] h⁻¹ for *S. symbiotica*. These values were
387 significantly slower than for *B. aphidicola*, and *R. insecticola*, and equal to or slower than for
388 *A. pisum* nuclear and mitochondrial DNA.

389 **Discussion**

390 *Metagenomic approaches in gut analyses*

391 This work showed that metagenomic approaches are sensitive enough to detect a single aphid
392 prey and its associated bacterial symbionts without prior DNA amplification, based on dozens
393 of mtDNA reads or hundreds of matches to the nuclear genome of the pea aphid. A key aspect
394 for prey DNA recovery was the use of stringent thresholds, which not only ensured the use of
395 high-quality reads but also limited false positives and established species identity of prey and
396 symbionts with great precision. These parameters were clearly sufficient to discriminate the
397 *A. pisum* mitogenome reads from *Aph. glycines*, which were provided as sustaining food later
398 in the feeding trial. Available genome sequences serving as reference are an important
399 resource for this approach. In the case of the pea aphid both mitochondrial and nuclear
400 genomes had been assembled (Richards *et al.* 2010). In addition, the NCBI pea aphid scaffold
401 archive, containing many genome segments that remained unplaced in the final genome
402 assembly, was an important source for aphid read identifications. Thirty-four percent of
403 complex-repeat families are in the unplaced scaffold archive, and produced a greater number
404 of hits than the placed scaffolds. The complex-repeat families need characterization, because
405 they can be very powerful nuclear markers for species recognition (Dodsworth *et al.* 2014).
406 Conceivably, similar databases can be created readily for other aphid species that lack these
407 genomic resources, e.g. by low-coverage genomic sequencing ('genome skimming', Straub *et*
408 *al.* 2012) from which scaffolds of repetitive regions are readily assembled as a potentially
409 large source of taxon-specific markers.

410 A fraction of the selected aphid reads corresponded to potential non-species-specific
411 reads, i.e. highly conserved regions such as rRNAs or Simple Sequence Repeats (SSRs). A
412 small proportion (3.5%) of SSRs was present in the NCBI pea aphid scaffold archive, but they

413 generally did not produce matches to our read-to-genome BLAST-based mapping.
414 Nevertheless, any detected rRNAs and SSRs matches were excluded, and therefore species
415 misidentifications based on these sequences are unlikely.

416 While the number of mtDNA reads detected for the predator was the overwhelming
417 fraction of the reads and was always at least 400 times higher than for the prey, by dissecting
418 the guts we recovered sufficient genetic signal for the detection of prey DNA and for
419 analysing decay rates. The metagenomic approach provided a refined estimate of abundance
420 and ultimately the decay rate because detection is less limited by amplification efficiency of
421 one or a few target genes, but is related to the degree of preservation of a broader portion of
422 the prey genome. By avoiding the amplification step of prey DNA, the number of detected
423 reads is more directly correlated to the amount of prey material, which was neatly confirmed
424 by the decay of read numbers over time after feeding.

425 In addition to improved DNA abundance measures, the metagenomic approach is
426 powerful due to its holistic analysis of the gut content. This includes the recovery of the
427 obligate *B. aphidicola* genome that produced a roughly uniform distribution of matching
428 reads over its genome of 643.5 kb, nearly all of which were exact (100%) matches (Fig. S2).
429 With the read mapping approach used here, the recovery relies on the completeness of the
430 reference databases used to match the sequenced DNA community. Additional reference
431 databases can be constructed to search for other associated organisms, such as pathogens,
432 parasitoids, and possible food plants. The metagenomics approach holds the advantage that
433 the number of reads can be interpreted quantitatively for the entire system simultaneously
434 without the vagaries of PCR reactions on multiple targets.

435

436 *Prey decay in the predator gut*

437 The use of time-points separated by 24 h intervals, which bracketed the known D_{\max}
438 periods for PCR-based methods, seemed to be too long for mtDNA detection of only a single
439 aphid prey item without amplification of a target prey mtDNA gene. More prey mtDNA
440 might have been detected using a shorter evaluation interval of perhaps up to 12 h after prey
441 ingestion. This might improve the precision of the decay parameters for mtDNA and reduce
442 the large confidence region around the regression (Fig. 2A), but the values for the analyte
443 decay rate and analyte D_{\max} would not change much. On the other hand, the use of a library
444 with an insert of 450 bp might have precluded the detection of prey mtDNA reads for periods
445 longer than 3 h after prey ingestion, as most of the prey mtDNA in the predators' guts content
446 could have already been digested to smaller lengths (Chen *et al.* 2000). By increasing the
447 number of reads detected after 3 h, the analyte decay rate would be reduced, and analyte D_{\max}
448 would be longer.

449 The analyte detectability half-life of *A. pisum* genetic materials was 3.6 to 8.9 h, which
450 is similar to the 2.0-4.9 h detectability half-life for PCR-based detection of a single aphid
451 consumed by different predators (Greenstone *et al.* 2014). However, analyte D_{\max} , determined
452 here from the metagenomic data, was 2-11 times longer than PCR-based D_{\max} . We estimated
453 the D_{\max} for a single aphid prey using PCR to be 4.0 h for *Pardosa sternalis* (Aranae:
454 Lycosidae) and 9.8 h for *Tetragnatha laboriosa* (Aranae: Tetragnathidae) (data from
455 Kerzicnik *et al.* 2012), 12.9 h for *Pardosa* spp. (data from Kuusk *et al.* 2008), and 16.1 h at
456 14°C and 14.5 h at 21°C for *Adalia bipunctata* (Coleoptera: Coccinellidae) (data from
457 McMillan *et al.* 2007). When considering the decay of similar prey items, metagenomic
458 sequencing appears to enable prey detection for a longer period of time than PCR-based
459 methods.

460

461 *Symbiont detection and population dynamics in the predator gut*

462 The secondary detection of several genera and species of prey symbionts in this work was
463 possible because we could construct a symbiont reference database from GenBank. Secondary
464 detection is defined here as the detection of exogenous DNA that was inside the first source of
465 exogenous DNA (prey, in this case). Usually the detection of insect symbionts has been done
466 with PCR based-methods, including metabarcoding through 16S rRNA (Jones *et al.* 2011,
467 Hirsch *et al.* 2012). As found here, metagenomics can be used to monitor symbiont
468 population fluctuations after prey ingestion.

469 The detected *B. aphidicola*, *R. insecticola*, *H. defensa*, *Arsenophonus* spp., and *S.*
470 *symbiotica* are all known to be aphid symbionts, and none have been reported from
471 coccinellids, and they were not found in the negative control. In contrast, the genus *Serratia*,
472 which includes the widespread, non-symbiotic *S. marcescens* and other free-living species,
473 was detected in large numbers in all of the bioassay treatments, including the never-fed,
474 negative control.

475 Two kinds of decay patterns of prey symbionts were detected. One kind was for *B.*
476 *aphidicola* and *R. insecticola*, which decayed according to the first order exponential decay
477 model similar to prey DNA. This result suggests that the population dynamics of *B.*
478 *aphidicola* and *R. insecticola* in the gut of *H. axyridis* could be characterized as a pure death
479 process, where they are introduced into the predator gut via their aphid host and then die and
480 are digested at a fixed rate. *Buchnera aphidicola* was detected in large numbers immediately
481 after feeding, and up to 24 h after feeding, but not thereafter. A similar dynamic was found for
482 *R. insecticola*, which is only known from aphids (Oliver *et al.* 2010). Interestingly, both
483 decayed at a faster rate than *A. pisum* nuclear or mitochondrial DNA. Although *Aph. glycines*
484 aphids were provided once a day, starting four hours after *A. pisum* aphid feeding, no *B.*

485 *aphidicola* or *R. insecticola* were found at 48 and 96 h after feeding on *A. pisum*. Their decay
486 rates may have been so fast that any *B. aphidicola* or *R. insecticola* DNA introduced via *Aph.*
487 *glycines* aphids was already degraded by the time the predators were collected at 48 and 96 h
488 in the bioassay.

489 The second kind of decay pattern was observed for *H. defensa*, *S. symbiotica* and
490 *Arsenophonus* spp. *Hamiltonella defensa* and *S. symbiotica* are associated with *A. pisum*
491 where they coexist with *Buchnera* in bacteriocytes and also occur in sheath cells around
492 bacteriocytes and in the hemolymph (Oliver *et al.* 2010). *Arsenophonus* is widespread in
493 related Aphidinae, but not in pea aphid (Jousselin *et al.* 2012), and never has been reported
494 from any beetle species. Because the only food consumed by the 3 h post-feeding *H. axyridis*
495 was *A. pisum*, *Arsenophonus* was most likely present in the North Dakota *A. pisum* population
496 used in this study.

497 One possible explanation for the unusual decay pattern is that it was generated due to
498 a random association of infected aphid hosts with beetles at the different time points, because
499 the facultative symbionts do not infect all of their aphid hosts (Russell *et al.* 2013). We
500 rejected this possible explanation, by calculating the probability that this could have happened
501 just by chance. An upper bound on this probability is 0.33% (see Supporting Information), so
502 the observed patterns probably reflect changes in the relative population size of these three
503 symbionts in the predator gut. In addition, the large number of reads at 3 h could not have
504 come from *Aph. glycines* aphids, as none of the predators had access to this food until 4 h
505 after consumption of *A. pisum*.

506 All three symbionts (*H. defensa*, *Arsenophonus* spp., and *S. symbiotica*) started with
507 small or undetectable numbers immediately after *H. axyridis* fed on *A. pisum*, and by 3 h
508 later, their populations grew in the predator guts by 1-2 orders of magnitude. Subsequently

509 they declined at different rates, with *H. defensa* declining fastest and *S. symbiotica* declining
510 slowest. The predator gut appears to be suitable for initial high rates of reproduction of these
511 symbionts, suggestive of an infection attempt during the 3 h after *A. pisum* ingestion. Indeed,
512 Degnan *et al.* (2009) found that *H. defensa* had abundant putative pathogenicity loci and
513 regulatory genes that may be important for infecting new hosts. In addition, Costopoulos *et al.*
514 (2014) fed the coccinellid *Hippodamia convergens* with aphids containing either *H. defensa*
515 or *S. symbiotica* which, compared to a control diet, reduced coccinellid survival and increased
516 adult size. The transient increase in symbiont populations reported here could account for how
517 a prey symbiont could affect the predator. The observed decline in symbiont populations later
518 in the bioassay indicates deterioration of the predator gut environment, possibly caused by the
519 predator immunity defense and increased competition from other gut bacteria.

520 Although infective horizontal transmission of prey symbionts to predators has not
521 been reported, it eventually could happen through repeated transient infections by prey
522 symbionts after prey ingestion, especially if the symbiont conferred advantageous ecological
523 effects. From our results, we can hypothesize that only the less specialized symbionts can
524 survive such transmission. *Hamiltonella* and *Regiella* species are generally distinguished from
525 their “free-living” Enterobacteriaceae relatives by their reduced genomes and the loss of some
526 essential pathways (Moran *et al.* 2005; Degnan & Moran 2008; Rao *et al.* 2012). On the other
527 hand, *Arsenophonus* species possess larger genomes and are morphologically and functionally
528 very diverse in different aphid lineages, while *Serratia* species are widespread in many insects
529 (Nováková *et al.* 2009; Russell *et al.* 2013). The fastest decay rates observed for *Hamiltonella*
530 and *Regiella* species and the slowest decay rates observed for *Arsenophonus* and *S.*
531 *symbiotica* seem to be directly correlated to these different levels of symbiosis.

532

533 *Broader implications*

534 The use of metagenomics in predator gut content analysis is a powerful tool that can reveal
535 complex relationships among predators, prey, and their symbionts. Because the copy number
536 of the genetic materials does not change during sample processing, the dynamics of these
537 relationships can be studied quantitatively. Although it does not require development of
538 specific PCR primers or antibodies, it requires reference DNA databases to make possible
539 species identification. These databases could focus on either prey nuclear or mitochondrial
540 DNA or symbiont genomes, and can be acquired from GenBank, or provided by the
541 investigator. The prey DNA databases allow definitive identification of prey species, while
542 the symbiont database may reinforce the prey identifications and reveal prey symbiont
543 population dynamics in the predator gut. Finally, because of its high analyte D_{\max} and
544 specificity, metagenomics can be especially useful for trophic interaction studies with a high
545 number of prey species to be detected at the same time, identifying unknown prey and
546 revealing species not previously known to be preyed upon by a predator.

547 **Acknowledgments**

548 We would like to thank Jason Harmon at North Dakota State University for providing the *A.*
549 *pisum* aphids; Kristina Prescott for helping with the feeding bioassays; and Alex Crampton-
550 Platt, Martijn Timmermans and Amrita Srivathsan for help with bioinformatics.

551 **References**

- 552 Altschul SF, Gish W, Miller W, *et al.* (1990) Basic local alignment search tool. *Journal of*
553 *Molecular Biology*, **215**, 403-410.
- 554 Chen Y, Giles KL, Payton ME, Greenstone MH (2000) Identifying key cereal aphid predators
555 by molecular gut analysis. *Molecular Ecology*, **9**, 1887-1898.
- 556 Costopoulos K, Kovacs JL, Kamins A, Gerardo NM (2014) Aphid facultative symbionts
557 reduce survival of the predator lady beetle *Hippodamia convergens*. *BMC Ecology*, **14**, 5.
- 558 Chen Y, Giles KL, Payton ME, Greenstone MH (2000) Identifying key cereal aphid predators
559 by molecular gut analysis. *Molecular Ecology*, **9**, 1887–98.
- 560 Deagle BE, Tollit DJ, Jarman SN, Hindell MA, Trites AW, Gales NJ (2005) Molecular
561 scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea
562 lions. *Molecular Ecology*, **14**, 1831–1842.
- 563 Deagle BE, Chiaradia A, McInnes J, Jarman SN (2010) Pyrosequencing faecal DNA to
564 determine diet of little penguins: is what goes in what comes out? *Conservation Genetics*,
565 **11**, 2039–2048
- 566 Degnan PH, Moran NA (2008) Evolutionary genetics of a defensive facultative symbiont of
567 insects: exchange of toxin-encoding bacteriophage. *Molecular Ecology*, **17** (3), 916-929.
- 568 Degnan PH, Yub Y, Sisneros N, Wing RA, Moran NA (2009) *Hamiltonella defensa*, genome
569 evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceeding of*
570 *the National Academy of Sciences (USA)*, **106** (22), 9063-9068.
- 571 Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, Novák P, Piednoël M, Weiss-
572 Schneeweiss H, Leitch AR (2014) Genomic repeat abundances contain phylogenetic
573 signal. *Systematic Biology*, doi: 10.1093/sysbio/syu080.

574 Eddy S, Durbin R (1994) RNA sequence analysis using co-variance models. *Nucleic Acids*
575 *Research*, **22**, 2079-2088.

576 Fein-Zchori E, Bourtzis K (2012) Manipulative tenants, bacteria associated with Arthropods.
577 CRC Press, Frontiers in Microbiology Series. 306 p.

578 Green SJ, Minz D (2005) Suicide polymerase endonuclease restriction, a novel technique for
579 enhancing PCR amplification of minor DNA templates. *Applied and Environmental*
580 *Microbiology*, **71**, 4721–4727.

581 Greenstone MH, Rowley DL, Weber DC, Payton ME, Hawthorne DJ (2007) Feeding mode
582 and prey detectability half-lives in molecular gut-content analysis: an example with two
583 predators of the Colorado potato beetle. *Bulletin of Entomological Research*, **97**, 201-209.

584 Greenstone MH, Payton ME, Weber DC, Simmons AM (2014) The detectability half-life in
585 arthropod predator–prey research: what it is, why we need it, how to measure it, and how
586 to use it. *Molecular Ecology* **23**, 3799–3813. doi: 10.1111/mec.12552.

587 Harwood JD, Obrycki JJ (2005) Quantifying aphid predation rates of generalist predators in
588 the field. *European Journal of Entomology*, **102**, 335-350.

589 Hereward JP, Walter GH (2012) Molecular interrogation of the feeding behaviour of field
590 captured individual insects for interpretation of multiple host plant use. *PLoS ONE*, **7** (9),
591 e44435.

592 Hirsch J, Strohmeier S, Pfannkuchen M, Reineke A (2012) Assessment of bacterial
593 endosymbiont diversity in *Otiorhynchus* spp. (Coleoptera: Curculionidae) larvae using a
594 multitag 454 pyrosequencing approach. *BMC Microbiology*, **12** (Suppl 1), S6.

595 Jarman SN, Wilson SG (2004) DNA-based species identification of krill consumed by whale
596 sharks. *Journal of Fish Biology*, **65**, 586–591.

597 Jones RT, Bressan A, Greenwell AM, Fierer N (2011) Bacterial communities of two
598 parthenogenetic aphid species colonizing two host plants across the Hawaiian Islands.
599 *Applied and Environmental Microbiology*, **77**, 8345–8349.

600 Jousselein E, D’Acier AC, Vanlerberghe-Masutti F, Duron O (2012) Evolution and diversity of
601 *Arsenophonus* endosymbionts in aphids. *Molecular Ecology*, **22**, 260-270.

602 Kerzicnik LM, Chapman EG, Harwood JD, Peairs FB, Cushing PE (2012) Molecular
603 characterization of Russian wheat aphid consumption by spiders in winter wheat. *Journal*
604 *of Arachnology*, **40**, 71-77.

605 Kuusk A-K, Cassel-Lundhagen A, Kvarheden A, Ekbom B (2008) Tracking aphid predation
606 by lycosid spiders in spring-sown cereals using PCR-based gut-content analysis. *Basic and*
607 *Applied Ecology*, **9**, 718-725.

608 Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B (2012) RobiNA: a
609 user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic*
610 *Acids Research*, **40**, W622-7.

611 Lövei GL, Monostori É, István A (1985) Digestion rate in relation to starvation in the larva of
612 a carabid predator, *Poecilus cupreus*. *Entomologia Experimentalis et Applicata*, **37**, 123-
613 127.

614 Majerus MEN (2006) The impact of male-killing bacteria on the evolution of aphidophagous
615 coccinellids. *European Journal of Entomology*, **103**, 1–7.

616 McMillan S, Kuusk A-K, Cassel-Lunhagen A, Ekbom B (2007) The influence of time and
617 temperature on molecular gut-content analysis: *Adalia bipunctata* fed *Rhopalosiphum*
618 *padi*. *Insect Science*, **14**, 353-358.

619 Moran NA, Russell JA, Koga R, Fukatsu T (2005) Evolutionary relationships of three new
620 species of *Enterobacteriaceae* living as symbionts of aphids and other insects. *Applied and*
621 *Environmental Microbiology*, **71**, 3302–3310.

622 Myers EW, Sutton GG, Delcher AL *et al.* (2000) A whole-genome assembly of *Drosophila*.
623 *Science*, **287**, 2196-2204.

624 Nováková E, Hypsa V, Moran NA (2009) *Arsenophonus*, an emerging clade of intracellular
625 symbionts with a broad host distribution. *BMC Microbiology*, **9**, 143.

626 Oliver KM, Degnan PH, Burke GR, Moran NA (2010) Facultative symbionts in aphids and
627 the horizontal transfer of ecologically important traits. *Annual Review of Entomology*, **55**,
628 247–266.

629 Piñol J, San Andres V, Clare EL, Mir G, Symondson WOC (2014) A pragmatic approach to
630 the analysis of diets of generalist predators: the use of next-generation sequencing with no
631 blocking probes. *Molecular Ecology Resources*, **14**, 18-26.

632 Pompanon F, Deagle BE, Symondson WOC *et al.* (2012) Who is eating what: Diet
633 assessment using next generation sequencing. *Molecular Ecology*, **21** (8), 1931-1950.

634 Quast C, Pruesse E, Yilmaz P *et al.* (2013) The SILVA ribosomal RNA gene database
635 project: improved data processing and web-based tools. *Nucleic Acids Research*, 41 (D1),
636 D590-D596.

637 Rao Q, Wang S, Su Y-L *et al.* (2012) Draft genome sequence of "*Candidatus Hamiltonella*
638 *defensa*," an endosymbiont of the whitefly *Bemisia tabaci*. *Journal of Bacteriology*, **194**
639 (13), 3558.

640 Richards S, Gibbs RA, Gerard NM *et al.* (2010) Genome sequence of the pea aphid
641 *Acyrtosiphon pisum*. *PLoS Biology*, **8** (2), e1000313.

642 Russell JA, Weldon S, Smith AH, Kim KL, Hu1 Y, Łukasik P, Doll S, Anastopoulos J, Novin
643 M, Oliver KM (2013) Uncovering symbiont-driven genetic diversity across North
644 American pea aphids. *Molecular Ecology*, **22** (7), 2045–2059.

645 Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets.
646 *Bioinformatics*, **27**, 863-864.

647 Shehzad WM, Riaz T, Nawaz MA, Miquel C, Poillot C, Shah SA, Pompanon F, Coissac E,
648 Taberlet P (2012) Carnivore diet analysis based on next-generation sequencing: application
649 to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*, **21**, 1951–
650 1965.

651 Simon J-C, Boutin S, Tsuchida T, Koga R, Le Gallic J-F, Frantz A, Outreman Y, Fukatsu T
652 (2011) Facultative symbiont infections affect aphid reproduction. *PLoS ONE*, **6** (7),
653 e21831.

654 Srivathsan, A., J. C. M. Sha, A. P. Vogler, and R. Meier. Comparing the effectiveness of
655 metagenomics and metabarcoding for diet analysis of a leaf feeding monkey (*Pygathrix*
656 *nemaeus*). *Molecular Ecology Resources*, Accepted manuscript online: 7 JUL 2014
657 11:12AM EST | DOI: 10.1111/1755-0998.12302.

658 Sopp PI, Sunderland KD (1989) Some factors affecting the detection period of aphid remains
659 in predators using ELISA. *Entomologia Experimentalis et Applicata*, **51**, 11-20.

660 Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A (2012) Navigating the
661 tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American*
662 *Journal of Botany*, **99**, 349-364.

663 Symondson WOC (2002) Molecular identification of prey in predator diets. *Molecular*
664 *Ecology*, **11**, 627-641.

665 Timmermans MJTN, Vogler AP (2012) Phylogenetically informative rearrangements in

666 mitochondrial genomes of Coleoptera, and monophyly of aquatic elateriform beetles
667 (Dryopoidea). *Molecular Phylogenetics and Evolution*, **63**, 299-304.

668 Thomas AP, Trotman J, Wheatley A, Aebi A, Zindel R, Brown PMJ (2012) Predation of
669 native coccinellids by the invasive alien *Harmonia axyridis* (Coleoptera: Coccinellidae):
670 detection in Britain by PCR-based gut analysis. *Insect Conservation and Diversity*, **6**, 20-
671 27.

672 Valentini A, Pompanon F, Taberlet P (2009a) DNA barcoding for ecologists. *Trends in*
673 *Ecology and Evolution*, **24**, 110-117.

674 Valentini A, Miguel C, Nawaz MA *et al.* (2009b) New perspectives in diet analysis based on
675 DNA barcoding and parallel pyrosequencing: The trnL approach. *Molecular Ecology*
676 *Research*, **9**, 51-60.

677 Vesterinen EJ, Lilley T, Laine VN *et al.* (2013) Next generation sequencing of fecal DNA
678 reveals the dietary diversity of the widespread insectivorous predator Daubenton's bat
679 (*Myotis daubentonii*) in southwestern Finland. *PLoS ONE*, **8** (11), e82168.

680 Vestheim H, Jarman SN (2008) Blocking primers to enhance PCR amplification of rare
681 sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs.
682 *Frontiers in Zoology*, **5**, 11.

683 Vilcinskas A, Stoecker K, Schmidtberg H, Röhrich CR, Vogel H (2013) Invasive harlequin
684 ladybird carries biological weapons against native competitors. *Science*, **340**, 862-3.

685 Weber DC, Lundgren JG (2009) Detection of predation using qPCR: effect of prey quantity,
686 elapsed time, chaser diet, and sample preservation on detectable quantity of prey DNA.
687 *Journal of Insect Science*, **9**, 1-12.

- 688 Weinert LA, Tinsley MC, Temperley M, Jiggins FM (2007) Are we underestimating the
689 diversity and incidence of insect bacterial symbionts? A case study in ladybird beetles.
690 *Biology Letters*, **3**: 678–681. doi:10.1098/rsbl.2007.0373
- 691 Wille BD, Hartman GL (2009) Two species of symbiotic bacteria present in the soybean
692 aphid (Hemiptera: Aphididae). *Environmental Entomology*, **38**, 110-115.
- 693 Yu P, Henry CM, Leung SM, *et al.* (2011) IDBA-UD: A *de novo* assembler for single-cell
694 and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, **28**, 1420-
695 1428.
- 696 Zeale MRK, Butlin RK, Barker GA, Lees DC, Jones G (2010) Taxon-specific PCR for DNA
697 barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*, **11**, 236-244.

698 **Data Accessibility**

699 The metagenomic data have been deposited in MG-RAST in project ??? (or Dryad ?). The
700 *Harmonia axyridis* mitogenome sequence was deposited at GenBank under the accession
701 code KJ778886. The Bioinformatics scripts are provided in the online Supporting
702 Information.

703 **Author Contributions**

704 DPP and APV designed the research. DAA conducted the field sampling, the feeding
705 bioassay and the decay statistical data analyses. DPP performed the molecular biology
706 experiments. DPP and BL analyzed the bioinformatics data. DPP, DAA, BL and APV wrote
707 the paper. DPP, DAA, BL, APV, CSSP and ERS revised it.

708 **Figure legends**

709

710 **Fig. 1** Coverage of *A. pisum* mtDNA for a single aphid in the prey feeding bioassay. The
711 tRNA genes are represented by amino acid single letter codes. The rRNA genes are
712 represented by “16S” and “12S”. The non-coding region (D-loop and AT rich) is represented
713 by “Misc”. Protein-coding genes are represented by their standard abbreviations.

714

715 **Fig. 2** Decay (analyte decay rate, analyte half-life and analyte D_{\max}) of the genetic material of
716 a single prey as a function of time after predation detected through metagenomics. A) the
717 mtDNA of *A. pisum*; B) nuclear genome of *A. pisum*; C) genome of the obligatory symbiont
718 *B. aphidicola*; D) genome of the facultative symbiont *R. insectiola*. The number of reads was
719 normalized by the library size. Heavy solid lines: expected decay process; Light solid lines:
720 95% confidence envelop for decay process; Solid circles: expected observed reads with 95%
721 credibility intervals based on posterior Bayesian distribution.

722

723 **Fig. 3** Mean number of reads of the prey bacterial symbionts found in the gut content of *H.*
724 *axyridis* as a function of time after aphid predation with 95% credibility intervals (from
725 posterior Bayesian distribution). The number of reads was normalized by the library size.

726

727 **Table 1** Number of reads obtained in the TruSeq libraries and MiSeq-Illumina sequencing for the
 728 feeding bioassay after quality control. The total number of reads in each library was used to normalize
 729 the data among the treatments for estimating the decay parameters

Reads number	Elapsed time after feeding					
	Pre	0 h	3 h	24 h	48 h	96 h
DNA ($\mu\text{g/mL}$)	24.39	28.73	24.81	28.01	20.70	22.70
Forward (R1)	1,751,599	1,967,870	1,664,734	2,072,981	2,115,223	1,602,152
Reverse (R2)	1,750,653	2,022,493	1,652,913	2,083,512	2,119,968	1,598,851
Predator mtDNA	7,427	10,849	9,165	13,442	10,963	7,191

730

731 **Table 2** Number of reads obtained for the mtDNA and nuclear genome for the prey, *A. pisum*, and for
 732 the complete genomes of the bacterial symbiont genera and species detected for the each time point in
 733 the feeding bioassay. The high *Serratia* spp. read numbers included an abundant species associated
 734 with the predator and the prey

	Elapsed time after feeding					
	Pre	0 h	3 h	24 h	48 h	96 h
<i>Acyrtosiphon. pisum</i> mtDNA	0	13	10	0	0	0
<i>Acyrtosiphon pisum</i> nuclear DNA	0	518	185	10	6	3
<i>Buchnera aphidicola</i>	0	1,651	171	2	0	0
<i>Arsenophonus</i> spp.	0	0	76	12	11	13
<i>Hamiltonella defensa</i>	0	0	577	0	0	4
<i>Regiella insecticola</i>	0	27	2	0	0	0
<i>Serratia</i> spp.	12,450	18,939	10,761	21,270	16,680	12,220
<i>Serratia symbiotica</i>	0	1	9	2	3	2

735

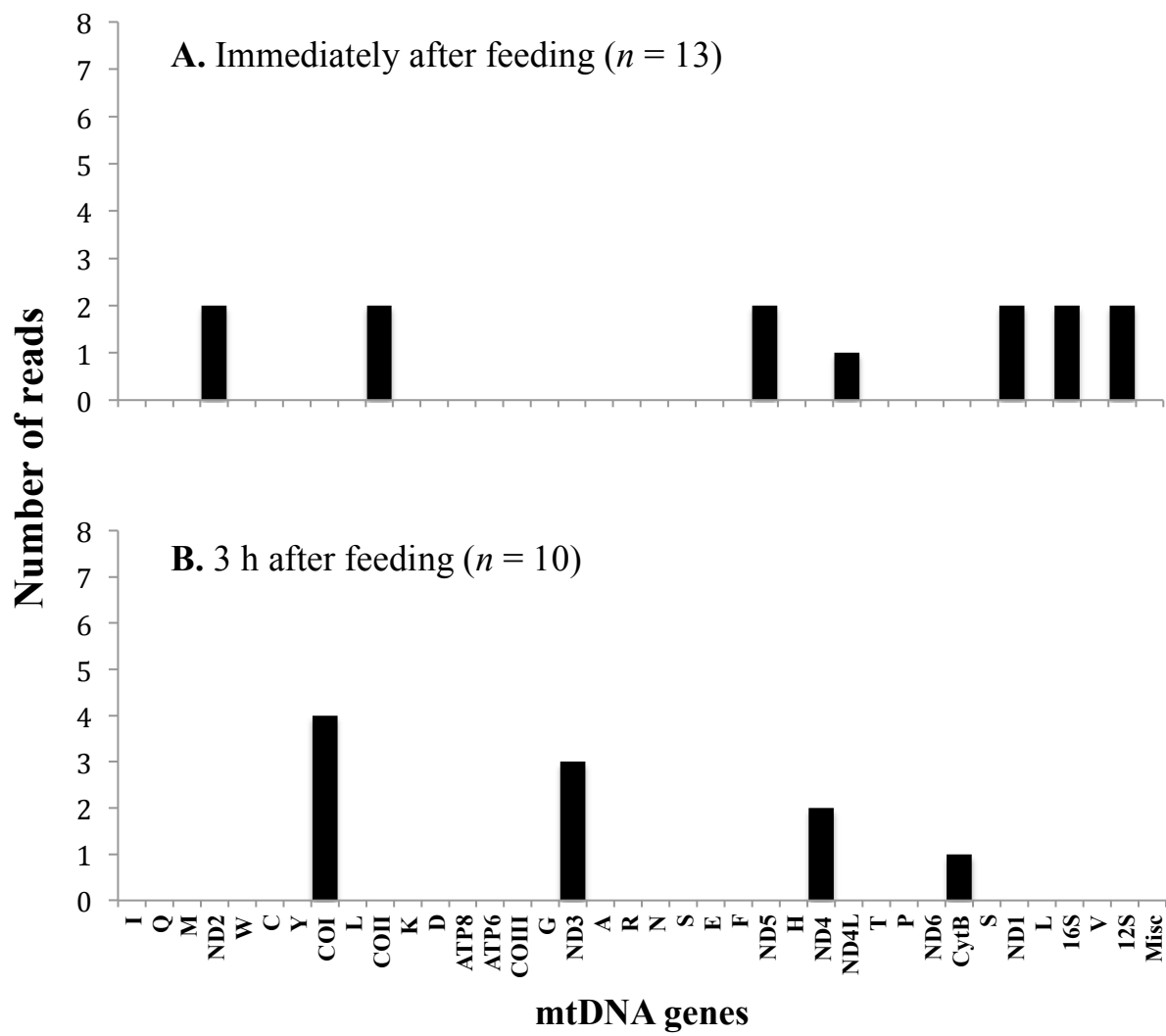


Fig. 1

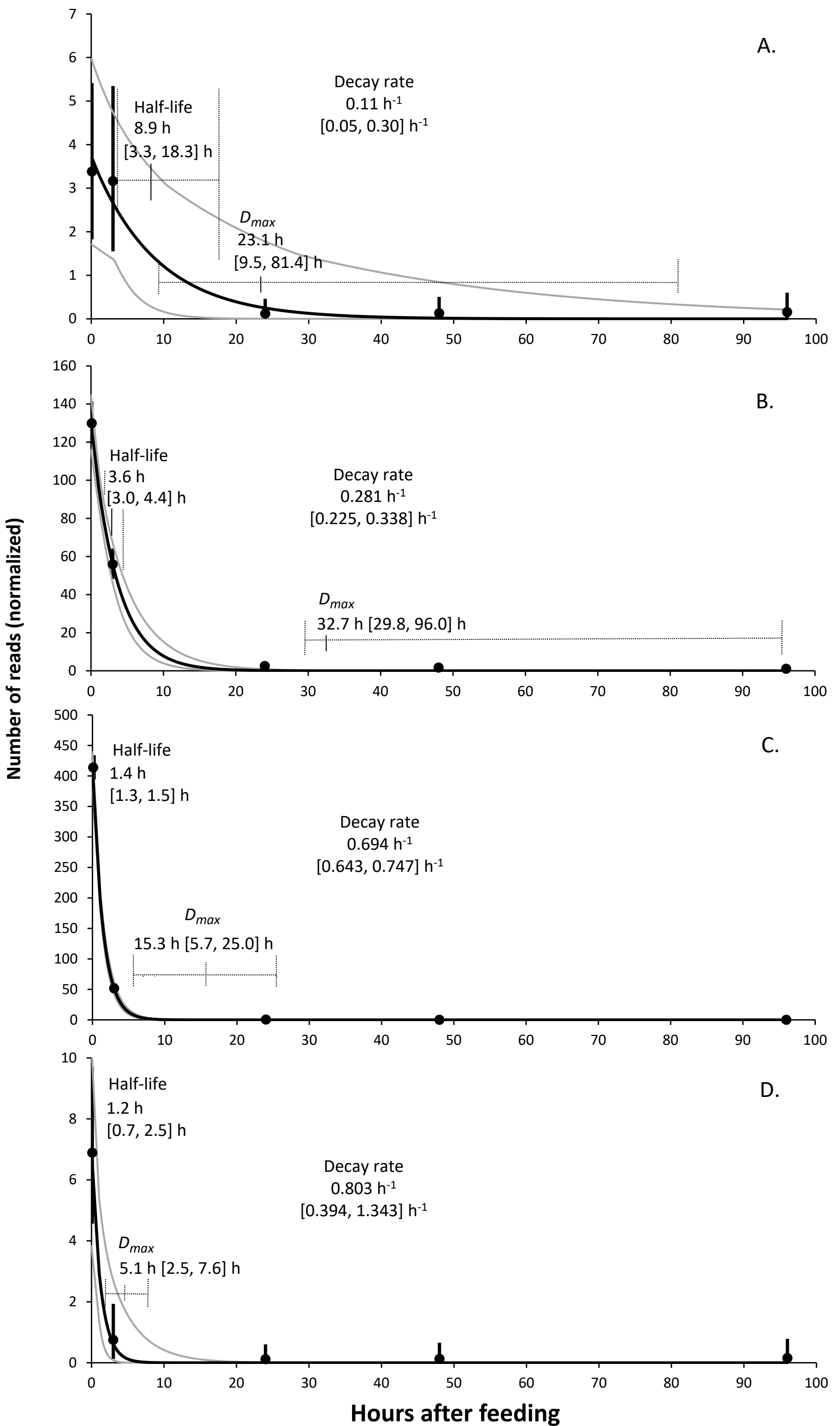


Figure 2

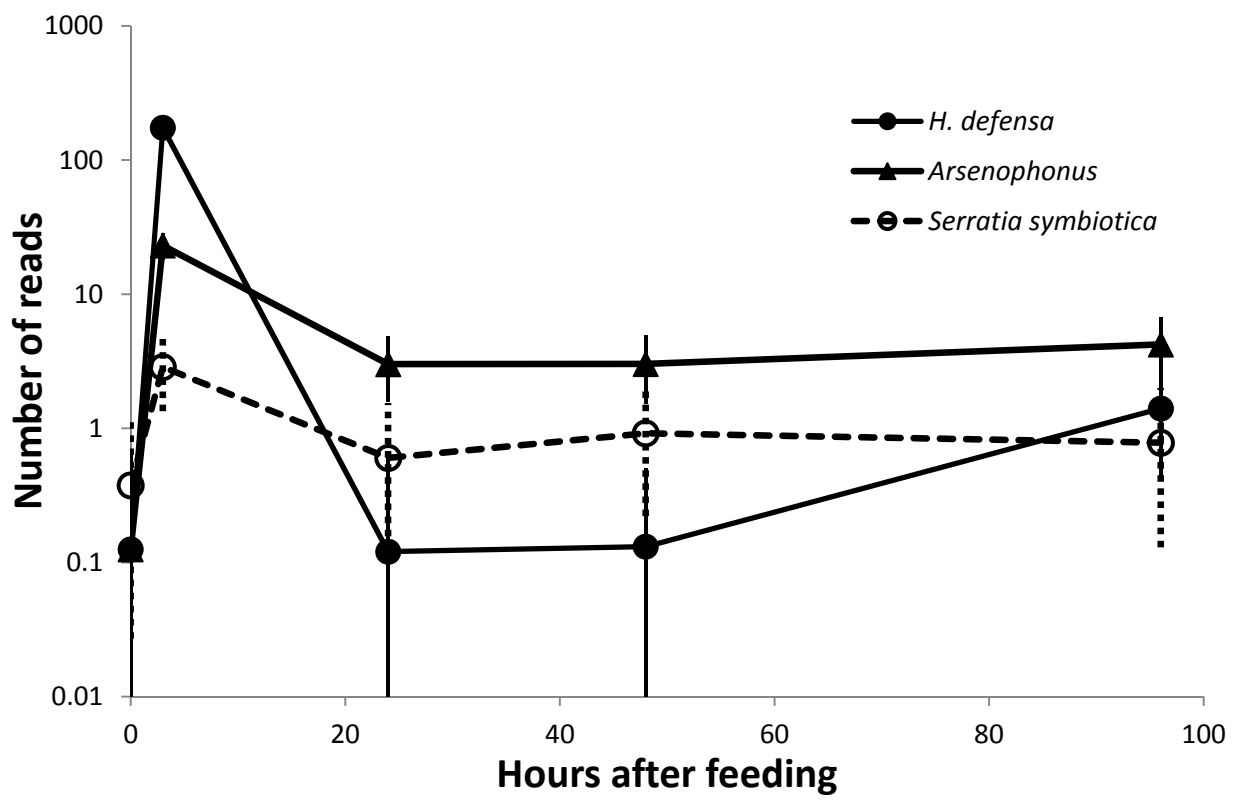


Figure 3

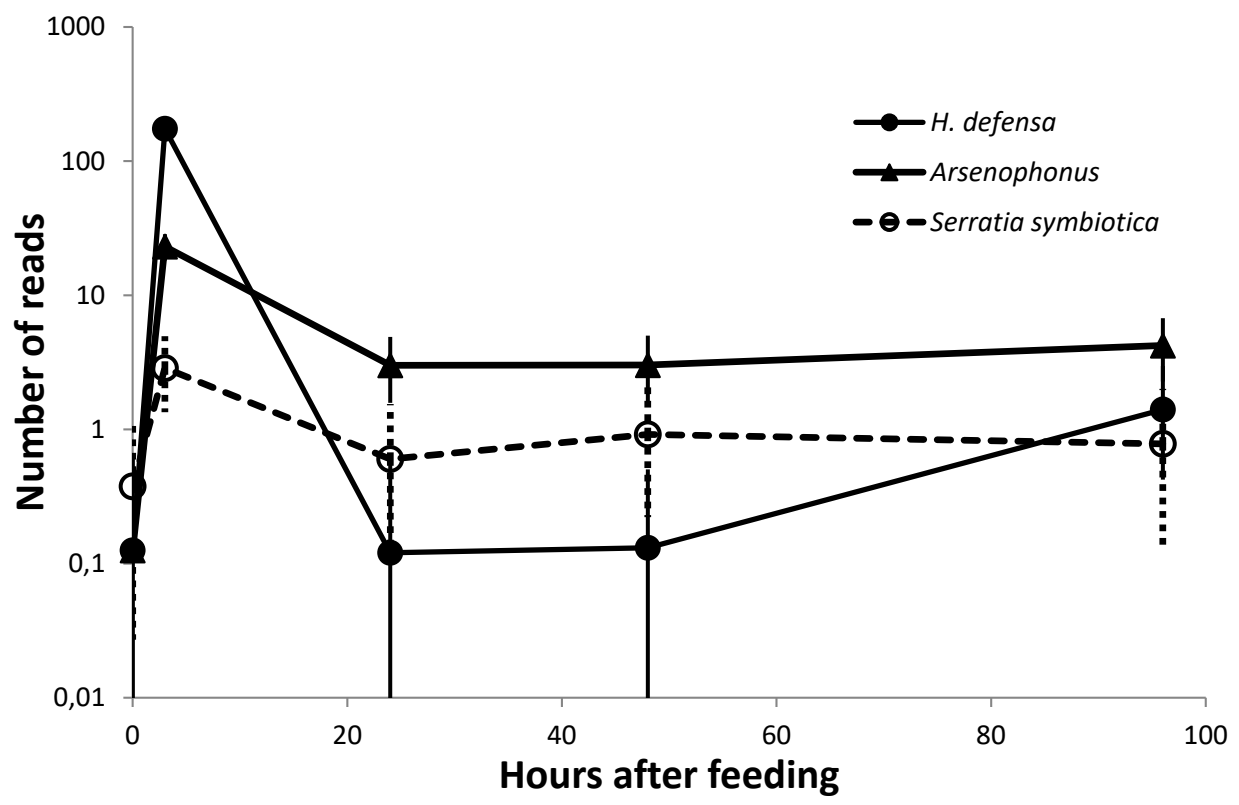


Figure 3