MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2015) **15**, 880–892 doi: 10.1111/1755-0998.12364

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* and ALFRIED P. VOGLER†,§

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

Abstract

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

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

Received 21 July 2014; revision received 15 December 2014; accepted 18 December 2014

Introduction

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

Since the advent of high-throughput DNA sequencing, diet analyses based on faeces have been assessed in several mammals, birds and insects through barcode region sequencing, known as metabarcoding (Valentini *et al.* 2009a,b; Deagle *et al.* 2010; Hereward & Walter 2012; Pompanon *et al.* 2012; Vesterinen *et al.* 2013). In these studies, DNA barcodes allow the detection of a spectrum of species against a set of DNA reference sequences, without need of cloning PCR products. However, despite being less time-consuming and very sensitive, there are still limitations, such as the need to design

Correspondence: Débora P. Paula, Fax: +55 (61) 34484929; E-mail: debora.pires@embrapa.br

© 2014 John Wiley & Sons Ltd

taxon-specific or group-specific primers that avoid predator DNA amplification (Jarman & Wilson 2004; Deagle *et al.* 2005), or to digest or block predator template DNA (Green & Minz 2005; Vestheim & Jarman 2008; Deagle *et al.* 2009; Deagle *et al.* 2010; Shehzad *et al.* 2012; but see Piñol *et al.* 2014). In addition, problems with nontarget template amplification (Zeale *et al.* 2010) or cross-amplification when predator and prey are phylogenetically close (Thomas *et al.* 2012) need to be addressed. Further, the amplification of target DNA limits the study of the decay dynamics of DNA inside the predators because of the difficulties of quantifying the amount of starting material with the PCR procedure, and because of the focus on a single gene region.

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

It is widely agreed that prey DNA susceptibility to predator digestion (Harwood & Obrycki 2005; Greenstone *et al.* 2007; Weber & Lundgren 2009) and the molecular technique (Greenstone *et al.* 2014) used for prey detection are important factors influencing the sensitivity of prey detection. Consequently, the prey detection system proposed here based on the detection of any part of the prey genomes (and on associated symbionts) and on shotgun sequencing of the DNA in the predator gut should be investigated more deeply by, for example, estimating the DNA decay rate and detectability period. These decay parameters indicate how long prey can be detected according to the speed and DNA susceptibility to digestion, providing a basis for comparison with other molecular techniques.

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

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

Material and methods

Insects and description of the study system

Harmonia axyridis (Coleoptera: Coccinellidae) was used as a model for prey detection in a common worldwide aphidophagous predator. Pupae (over 600) were collected in August 2013 (summer) in soybean fields in St. Paul, Minnesota, USA. Upon emergence, adults were transferred to individual petri dishes (35 × 10 mm) with moistened filter paper and held under controlled conditions (25 °C and 16:8 h L:D cycle) without food. After 24 h postemergence, the individuals were used in the feeding bioassay.

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

The presence of symbionts was tested against the genomes of the genera *Arsenophonus*, *Buchnera*, *Hamiltonella*, *Regiella*, *Rickettsia*, *Rickettsiella*, *Serratia*, *Spiroplasma* and *Wolbachia*. These genera were chosen because they are known to confer fitness advantages and costs either

© 2014 John Wiley & Sons Ltd

to aphids (Wille & Hartman 2009; Oliver *et al.* 2010; Jones *et al.* 2011; Jousselin *et al.* 2012) or to coccinellids (Majerus 2006; Weinert *et al.* 2007). *Buchnera* is an obligate symbiont occurring in high numbers in specialized host organs. Except for *Arsenophonus*, all of these symbionts have been reported in *A. pisum* (Simon *et al.* 2011; Russell *et al.* 2013). In addition, *Nosema* was included, as it could be associated with *H. axyridis* (Vilcinskas *et al.* 2013), and three insect, nonaphid and noncoccinellid symbionts, *Blattabacterium*, *Cardinium*, and *Midichloria*, were included as false-positive controls (Fein-Zchori & Bourtzis 2012).

Feeding bioassay

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

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

DNA sample preparation

The guts of the preserved predators were dissected out using clean forceps under a stereomicroscope to increase the chances of detecting prey DNA in the sample. Guts from the same time point were pooled into one sample. The total DNA of each sample was extracted with a DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) and quantified by fluorescence using the Qubit system (Invitrogen™) after quality checking spectrophotometrically (ratio $A_{260/280}$ nm). The total DNA concentration of each sample was normalized to 20 ng/μL and sonicated to construct TruSeq libraries of insert size of 450 bp (250-bp paired-end, 500 cycle kit). Each library was sequenced on MiSeq-Illumina using 17% of the flowcell.

Sequence quality controls

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

Predator mitogenome assembly

For the elucidation of the *H. axyridis* mitogenome, first the reads were filtered for similarity of E-value < 10^{-5} with NCBI Insecta mitogenomes that included partial and complete sequences of 587 species (taxonomic ID: 50557) using the BLASTn algorithm (Altschul *et al.* 1990). Filtering simplifies the assembly by reducing the size of the data set and enriching it with putative mitochondrial reads. The retained reads were assembled using Celera (Myers *et al.* 2000) and IDBA-UD (Yu *et al.* 2011), and for the latter, after quality control by PRINSEQ (Schmieder & Edwards 2011) with a minimum quality score and mean of 20, maximum ambiguous base N of 0 and trim quality from the right (3′) to a minimum of 20. The scaffolds generated by both assemblers were concatenated in GENEIOUS 7.0.5 (Biomatters, http://www.geneious.com/) using the parameters: no gaps allowed, minimum overlap 150, maximum mismatches per read 0, minimum overlap identity 99%, and maximum ambiguity 1. The mitogenome annotation was made by first annotating the tRNA genes using models based on the NCBI Insecta mitogenomes and the COVE software package (Eddy & Durbin 1994). The annotation process was finished manually in GENEIOUS 7.0.5. The nearly complete mitogenome sequence of 15 322 bp includes the expected two rRNA, 22 tRNA and 13 protein-coding genes arranged in the canonical gene order of Coleoptera (Timmermans & Vogler 2012). The control region was not completely sequenced. The mitogenome was deposited at GenBank under the Accession Code KJ778886.

Identification of aphid mtDNA

Good-quality reads from all time points were matched to the NCBI Insecta mtDNA reference database of 587 species, including pea aphid and five other aphid species (November 2013), and added to the sequenced mitogenome of *H. axyridis*. The matches were made by BLASTN with an E-value < 10^{-5}. Custom scripts (Supporting information) were used to associate the GenBank general identification (gi) number and its taxonomic identifica-

© 2014 John Wiley & Sons Ltd

tion with reads obtained by Illumina sequencing requiring sequence identity >98% over a minimum hit length of 225 bp (90% of read length). A species match was retained when it equalled or exceeded the thresholds for minimum length and identity. Preliminary analyses using lower identity thresholds indicated that all false positives and ambiguous identifications were eliminated at 98% identity. Many of these false positives were repetitive DNA with high AT content. The matched pea aphid mtDNA reads were mapped onto the prey mitogenome using GENEIOUS 7.0.5 to evaluate the overall coverage of the mtDNA. The map position of reads on the mitochondrial *A. pisum* genome was tested for a random distribution using the Poisson dispersion test.

Identification of aphid nuclear DNA

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

Identification of prey-associated symbionts

Thirteen bacterial genera with known insect symbiotic interactions were used to create a database of symbiont sequences. For each genus, we retrieved all available NCBI GenBank sequences and complete genomes to build the database (Table S1, Supporting information). DNA reads from each sample tested for the presence of these symbionts were aligned to this database with MEGABLAST, and all reads aligned over > 225 bp with >95% sequence identity and E-value $<10^{-9}$ were retained. Reads similar to the conservative rRNA sequences were removed to avoid misidentification due to insufficient sequence variability between related genera. The thresholds used discarded several reads that could be associated with one of the studied genera, but the need to discriminate several genera in a bacterial family (e.g. Enterobacteriaceae) required such a measure. The number of available reference genomes used to identify symbiont reads differed for each genus, which may affect the capacity for species detection (Table S1, Supporting information).

Statistical analyses to estimate decay parameters

An advantage of the metagenomic method is that the number of prey reads detected in the predator guts can be used to estimate the dynamics of analyte detectability. Although considerable work has been carried out with detectability half-lives, little use has been made of analyte detectability parameters (Greenstone *et al.* 2014). Here, we provide methods for estimating three analyte degradation parameters: analyte decay rate, analyte detectability half-life and analyte detectability period (D_{max}). Two critical points must be kept in mind. First, the detection of a prey or symbiont read is a stochastic process that combines random events associated with (i) the subsample of the total DNA in the gut sample and (ii) the subsample of reads sequenced from the resulting DNA library. This means that the number of reads observed at any time point is a random variable, and there is some probability that the actual number of reads was greater (or less) than the number observed. Specifically, a time point with zero observed reads must be treated as a random zero (i.e. there could have been one or more reads in the original sample, but the sampling and sequencing processes did not retain any of these reads), and not a true biological zero (i.e. there were no reads in the original sample), and is an important and meaningful datum. [Although similar random processes occur in PCR-based methods, in these methods, the sources of randomness simply add variance to the estimated probability of detecting a positive individual.] Second, because the bioassay used different individuals to evaluate digestion at each time point, the time points are statistically independent samples of the digestion process.

Treating digestion as a stochastic process makes explicit the uncertainty associated with the observed data. Assuming that all reads have the same probability of detection, the number of observed reads at each time point will follow a Poisson process. Further, the

© 2014 John Wiley & Sons Ltd

observed numbers can be used to estimate the underlying Poisson process and calculate the probability distribution for possible observed read numbers. This is carried out using Bayesian methods with a Jeffries prior. Bayesian methods were used because the observed read numbers are single realizations of the underlying random process. The number of reads and the Bayesian posterior distribution were normalized by the library size at each time point and multiplied by 10^6 for presentation purposes. Estimation of the detectability half-life usually assumes a first-order or exponential decay in the quantity of target DNA or protein degraded (Lövei *et al.* 1985; Sopp & Sunderland 1989; Weber & Lundgren 2009). Therefore, expected values from the posterior distributions (not the observed numbers) were used to fit an exponential decay model by nonlinear regression. This initial analysis allowed the identification of species and genera that did not decay exponentially in the predator guts.

For those that did decay exponentially, Monte Carlo simulation was used to estimate the analyte decay parameters. Three parameters were estimated: (i) instantaneous decay rate of the DNA (analyte decay rate), (ii) analyte detectability half-life and (iii) the maximum period during which DNA could be detected (analyte detectability period, which is analogous to D_{max}, Sunderland *et al.* 1987). Random read numbers were drawn from the normalized posterior distributions for each time point, an exponential decay model was fit to these values by nonlinear regression, and the estimated parameter values (analyte decay rate and initial number of reads) were saved. This was repeated 200 000 times to generate a joint probability distribution function (jpdf) of the 2-parameter values. The analyte decay rate and its 95% CI were estimated from the marginal distributions of the jpdf. The analyte detectability half-life and its 95% CI were estimated from the inverse of the decay rate. The jpdf was also used to estimate the 95% confidence region of the model parameters, and the border of this region was used to estimate the 95% confidence envelope of the nonlinear regression. Analyte D_{max} and its 95% CI were estimated using the original read numbers, the analyte decay rate and the 95% confidence envelope of the regression to calculate the time when only one read would be left. A similar method was used to estimate D_{max} from the original data published in McMillan *et al.* (2007), Kuusk *et al.* (2008) and Kerzicnik *et al.* (2012), who studied the detectability of single aphid prey using PCR. In these cases, we calculated the time when only one individual would test positive. All calculations were carried out in MATHEMATICA 7.0.

Results

Library basic statistics and recovery of predator DNA

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

Prey detection and decay parameters

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

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

	Elapsed time after feeding					
Read numbers	Pre	0 h	3 h	24 h	48 h	96 h
DNA (μg/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	7427	10 849	9165	13 442	10 963	7191

© 2014 John Wiley & Sons Ltd

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

	Elapsed time after feeding					
	Pre	0 h	3 h	24 h	48 h	96 h
Acyrthosiphon pisum mtDNA	0	13	10	0	0	0
Acyrthosiphon pisum nuclear DNA	0	518	185	10	6	3
Buchnera aphidicola	0	1651	171	2	0	0
Arsenophonus spp.	0	0	76	12	11	13
Hamiltonella defensa	0	0	577	0	0	4
Regiella insecticola	0	27	2	0	0	0
Serratia spp.	12 450	18 939	10 761	21 270	16 680	12 220
Serratia symbiotica	0	1	9	2	3	2

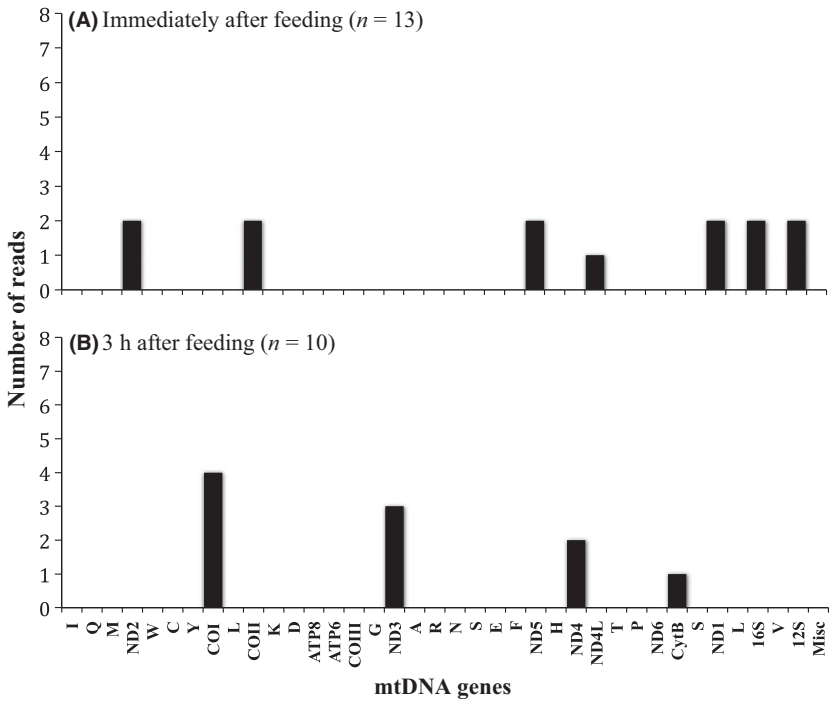

Fig. 1 Coverage of *A. pisum* mtDNA for a single aphid in the prey feeding bioassay. The tRNA genes are represented by amino acid single-letter codes. The rRNA genes are represented by '16S' and '12S'. The noncoding region (D-loop and AT rich) is represented by 'Misc'. Protein-coding genes are represented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

The decay of the mtDNA for a single *A. pisum* in *H. axyridis* fit the first-order exponential decay model extremely well ($p = 1.94 \times 10^{-3}$) with an adjusted $r^2 = 0.974$ (Fig. 2A). On average, the instantaneous analyte decay rate was 0.11 reads per hour with 95% CI of 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. The analyte D_{max} to detect a single *A. pisum* read based on mtDNA was 23.1 h with 95% CI of 9.5 to 81.4 h.

Nuclear genome. The number of reads with matches to the *A. pisum* nuclear genome exceeded the mtDNA reads by a factor of about 30, reaching over 500 reads at the moment of feeding (Table 2). No aphid sequences were detected in the prefeeding negative control. Aphid nuclear DNA detection continued for all time points, including the last one at 96 h after feeding. The latter was due to the recovery of three reads, which was unexpected given the already very low counts at the two earlier time points. Over the hundreds of reads showing a nearly perfect match in the pea aphid genome at 0 h, 48 matched 29 different pea aphid protein sequences retrieved from the NCBI RefseqP database (Table S2, Supporting information). Similarly, at 3 h, 13 reads matched 9 different aphid proteins. In many cases, both reads of the same pair matched the same aphid protein. Many aphid proteins are computational predictions based on the pea aphid genomes ('uncharacterized' and

© 2014 John Wiley & Sons Ltd

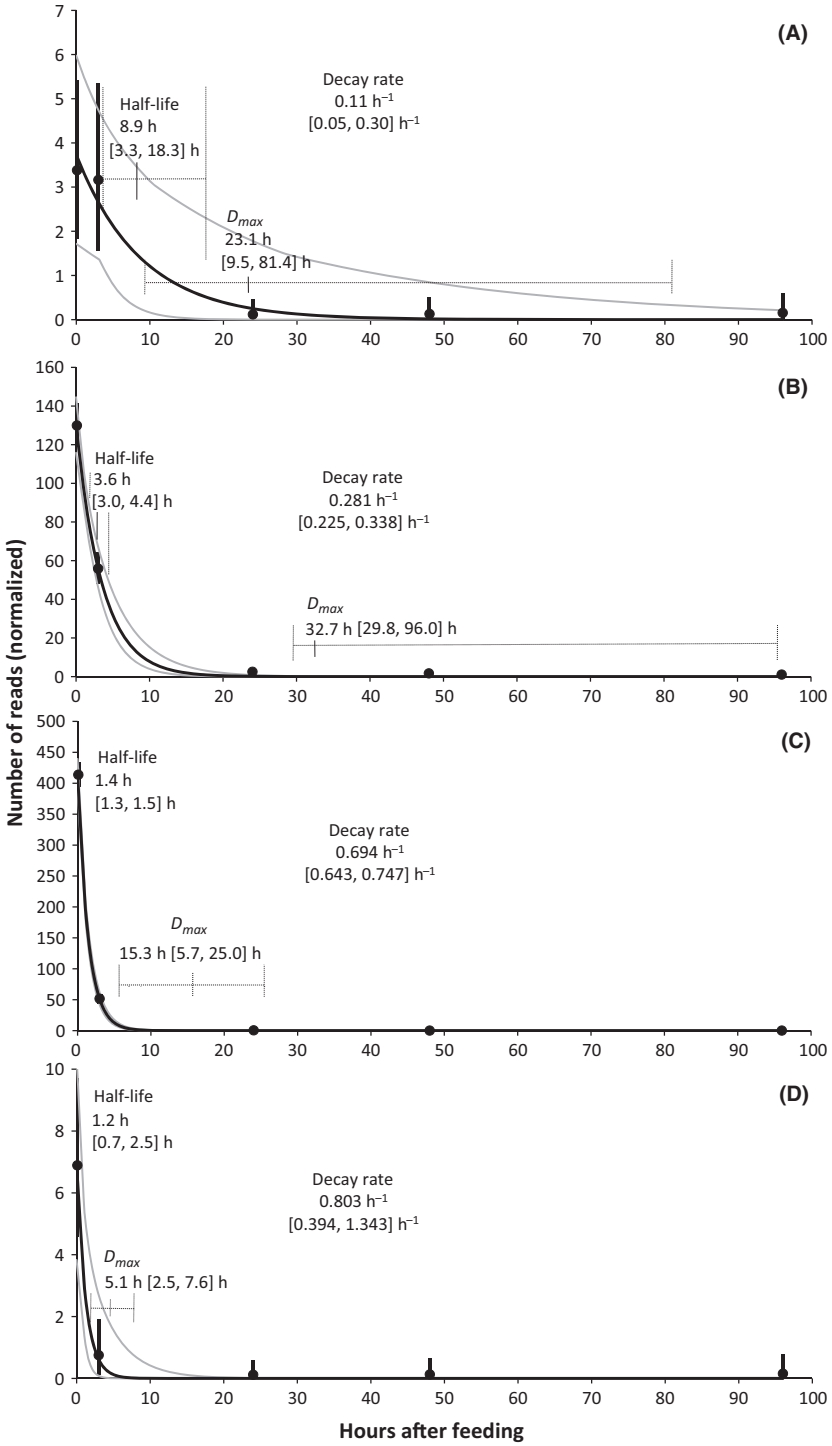

Fig. 2 Decay (analyte decay rate, analyte half-life and analyte D_{max}) of the genetic material of a single prey as a function of time after predation detected through metagenomics. A) the mtDNA of *A. pisum*; B) nuclear genome of *A. pisum*; C) genome of the obligatory symbiont *B. aphidicola*; D)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

'predicted' annotations) but they were nevertheless the closest hit in the database (which includes proteins from all domains of life). Some matches seem to be linked to integrated viral genomes (XP_008184955.1, an HIV Tat-specific factor-like element), but we also uncovered genes linked to specific functions. For instance, one of the reads matched an O-linked-mannose beta-1,2-N-acetylglucosaminyltransferase (XP_001948219.2, Table S2, Supporting information), a protein with a domain signature (NCBI domain cd13937) conserved in animals.

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

© 2014 John Wiley & Sons Ltd

analyte decay rate was 0.281 reads per hour with a 95% 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 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 significantly different from the corresponding parameters for *A. pisum* mtDNA, although the D_{max} was somewhat greater because many more nuclear reads were detected and reads were found at the final sampling time.

Detection characterization of prey symbionts

In addition to the detection of aphid nuclear and mitochondrial DNA, we identified reads homologous to known aphid bacterial symbionts, some of them in high numbers (Table 2). The symbionts *Buchnera aphidicola*, *Arsenophonus* spp., *Hamiltonella defensa*, *Regiella insecticola* and *Serratia symbiotica* were detected only after *H. axyridis* feeding, indicating that they were exclusively associated with the ingested pea aphids. The obligate symbiont *B. aphidicola* was present in the highest numbers, with an even read sampling over its whole genome, with 1651 reads at 0 h and 171 reads at 3 h (Fig. S2, Supporting information). Symbionts from the genera *Blattabacterium*, *Cardinium*, *Midichloria*, *Rickettsia*, *Rickettsiella*, *Spiroplasma* and *Wolbachia* were not detected at any time point, and neither was the microsporidian fungus *Nosema*. Reads matching the genomes of the genus *Serratia* (mainly *S. marcescens*) were detected in high numbers at all time points even before feeding, which indicates its association with the predator. In contrast, *S. symbiotica*, which is known to be an aphid symbiont (Lamelas *et al.* 2011), was observed only after feeding on pea aphids at all time points.

Two of the detected symbionts decayed according to the exponential decay model, and three of them did not. The first-order exponential decay model fit the data for the obligatory aphid symbiont *B. aphidicola* ($p = 2.39 \times 10^{-12}$; adjusted $r^2 = 1.000$) and the facultative *R. insecticola* ($p = 6.04 \times 10^{-7}$; adjusted $r^2 = 0.998$) very well. *Buchnera aphidicola* was detected in large numbers immediately after feeding and continued to be detected 24 h later (Table 2). The decay of *B. aphidicola* in *A. pisum* ingested by *H. axyridis* is presented in Fig. 2C. On average, the instantaneous analyte decay rate was 0.694 reads per hour with a 95% CI of 0.642 to 0.747 h^{-1}. The mean analyte detectability half-life was 3.4 h with 95% CI of 2.5 to 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 therefore was significantly faster than for either *A. pisum* mtDNA or nuclear DNA as there was no overlap in the 95% CIs.

Decay rate was similarly fast for *R. insecticola*, which however was detected in much smaller numbers than *B. aphidicola* (Table 2). On average, the instantaneous analyte decay rate was 0.80 reads per hour with a 95% CI of 0.39 to 1.34 h^{-1}. The analyte detectability half-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 2.4 to 7.6 h (Fig. 2D).

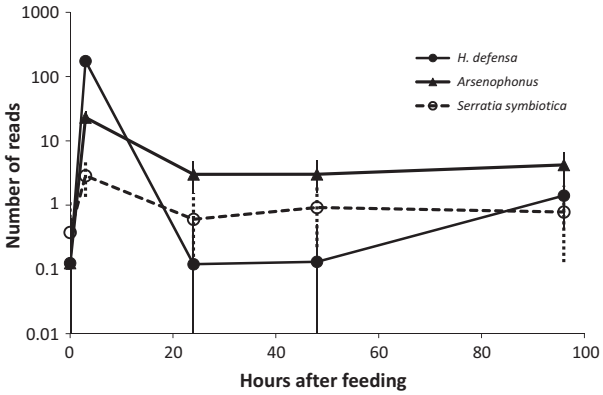

Fig. 3 Mean number of reads of the prey bacterial symbionts found in the gut content of *H. axyridis* as a function of time after aphid predation with 95% credibility intervals (from posterior Bayesian distribution). The number of reads was normalized by the library size.

The facultative symbionts *H. defensa*, *Arsenophonus* spp. and *S. symbiotica* did not fit the exponential decay model (*H. defensa*: $P = 0.245$, adjusted $r^2 = 0.007$; *Arsenophonus* spp.: $P = 0.185$, adjusted $r^2 = 0.115$; *S. symbiotica*: $P = 0.072$, adjusted $r^2 = 0.407$). All three exhibited a similar pattern, with no or almost no reads detected immediately after feeding, and a large, statistically significant increase in the number detected at 3 h after feeding, followed by a statistically significant decline in detection thereafter (Fig. 3). The rate of analyte decay 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} for *Arsenophonus* sp. and 0.04 [0.01, 0.08] h^{-1} for *S. symbiotica*. These values were significantly slower than for *B. aphidicola*, and *R. insecticola*, and equal to or slower than for *A. pisum* nuclear and mitochondrial DNA.

Discussion

Metagenomic approaches in gut analyses

This work showed that metagenomic approaches are sensitive enough to detect a single aphid prey and its associated bacterial symbionts without prior DNA amplification, based on dozens of mtDNA reads or hundreds of matches to the nuclear genome of the pea aphid. A key aspect for prey DNA recovery was the use of stringent thresholds, which not only ensured the use of high-quality reads but also limited false positives and established species identity of prey and symbionts with great precision. These parameters were clearly sufficient to discriminate the *A. pisum* mitogenome reads from

© 2014 John Wiley & Sons Ltd

Aph. glycines, which were provided as sustaining food later in the feeding trial. Available genome sequences serving as reference are an important resource for this approach. In the case of the pea aphid, both mitochondrial and nuclear genomes had been assembled (Richards *et al.* 2010). In addition, the NCBI pea aphid scaffold archive, containing many genome segments that remained unplaced in the final genome assembly, was an important source for aphid read identifications. Thirty-four per cent of complex-repeat families are in the unplaced scaffold archive and produced a greater number of hits than the placed scaffolds. The complex-repeat families need characterization, because they can be very powerful nuclear markers for species recognition (Dodsworth *et al.* 2014). Conceivably, similar databases can be created readily for other aphid species that lack these genomic resources, for example by low-coverage genomic sequencing ('genome skimming', Straub *et al.* 2012) from which scaffolds of repetitive regions are readily assembled as a potentially large source of taxon-specific markers.

A fraction of the selected aphid reads corresponded to potential non-species-specific reads, that is highly conserved regions such as rRNAs or simple sequence repeats (SSRs). A small proportion (3.5%) of SSRs was present in the NCBI pea aphid scaffold archive, but they generally did not produce matches to our read-to-genome BLAST-based mapping. Nevertheless, any detected rRNAs and SSRs matches were excluded, and therefore, species misidentifications based on these sequences are unlikely.

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

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

Prey decay in the predator gut

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

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

Symbiont detection and population dynamics in the predator gut

The secondary detection of several genera and species of prey symbionts in this work was possible because we

© 2014 John Wiley & Sons Ltd

could construct a symbiont reference database from GenBank. Secondary detection is defined here as the detection of exogenous DNA that was inside the first source of exogenous DNA (prey, in this case). Usually, the detection of insect symbionts has been carried out with PCR-based methods, including metabarcoding through 16S rRNA (Jones *et al.* 2011; Hirsch *et al.* 2012). As found here, metagenomics can be used to monitor symbiont population fluctuations after prey ingestion.

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

Two kinds of decay patterns of prey symbionts were detected. One kind was for *B. aphidicola* and *R. insecticola,* which decayed according to the first-order exponential decay model similar to prey DNA. This result suggests that the population dynamics of *B. aphidicola* and *R. insecticola* in the gut of *H. axyridis* could be characterized as a pure death process, where they are introduced into the predator gut via their aphid host and then die and are digested at a fixed rate. *Buchnera aphidicola* was detected in large numbers immediately after feeding, and up to 24 h after feeding, but not thereafter. A similar dynamic was found for *R. insecticola*, which is only known from aphids (Oliver *et al.* 2010). Interestingly, both decayed at a faster rate than *A. pisum* nuclear or mitochondrial DNA. Although *Aph. glycines* aphids were provided once a day, starting four hours after *A. pisum* aphid feeding, no *B. aphidicola* or *R. insecticola* were found at 48 and 96 h after feeding on *A. pisum*. Their decay rates may have been so fast that any *B. aphidicola* or *R. insecticola* DNA introduced via *Aph. glycines* aphids was already degraded by the time the predators were collected at 48 and 96 h in the bioassay.

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

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

All three symbionts (*H. defensa*, *Arsenophonus* spp. and *S. symbiotica*) started with small or undetectable numbers immediately after *H. axyridis* fed on *A. pisum*, and by 3 h later, their populations grew in the predator guts by 1–2 orders of magnitude. Subsequently, they declined at different rates, with *H. defensa* declining fastest and *S. symbiotica* declining slowest. The predator gut appears to be suitable for initial high rates of reproduction of these symbionts, suggestive of an infection attempt during the 3 h after *A. pisum* ingestion. Indeed, Degnan *et al.* (2009) found that *H. defensa* had abundant putative pathogenicity loci and regulatory genes that may be important for infecting new hosts. In addition, Costopoulos *et al.* (2014) fed the coccinellid *Hippodamia convergens* with aphids containing either *H. defensa* or *S. symbiotica* which, compared to a control diet, reduced coccinellid survival and increased adult size. The transient increase in symbiont populations reported here could account for how a prey symbiont could affect the predator. The observed decline in symbiont populations later in the bioassay indicates deterioration of the predator gut environment, possibly caused by the predator immunity defence and increased competition from other gut bacteria.

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

© 2014 John Wiley & Sons Ltd

and *S. symbiotica* seem to be directly correlated with these different levels of symbiosis.

Broader implications

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

Acknowledgements

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

References

Altschul SF, Gish W, Miller W *et al.* (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**, 573–580.

Chen Y, Giles KL, Payton ME, Greenstone MH (2000) Identifying key cereal aphid predators by molecular gut analysis. *Molecular Ecology*, **9**, 1887–1898.

Costopoulos K, Kovacs JL, Kamins A, Gerardo NM (2014) Aphid facultative symbionts reduce survival of the predator lady beetle *Hippodamia convergens*. *BMC Ecology*, **14**, 5.

Deagle BE, Tollit DJ, Jarman SN, Hindell MA, Trites AW, Gales NJ (2005) Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. *Molecular Ecology*, **14**, 1831–1842.

Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology*, **18**, 2022–2038.

Deagle BE, Chiaradia A, McInnes J, Jarman SN (2010) Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? *Conservation Genetics*, **11**, 2039–2048.

Degnan PH, Moran NA (2008) Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin-encoding bacteriophage. *Molecular Ecology*, **17**, 916–929.

Degnan PH, Yub Y, Sisneros N, Wing RA, Moran NA (2009) *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceeding of the National Academy of Sciences, USA*, **106**, 9063–9068.

Dodsworth S, Chase MW, Kelly LJ *et al.* (2014) Genomic repeat abundances contain phylogenetic signal. *Systematic Biology*, doi:10.1093/sysbio/syu080.

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

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

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

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

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

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

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

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

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

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

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

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, **110**, 462–467.

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

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

Lamelas A, Gosalbes MJ, Manzano-Marín A, Peretó J, Moya A, Latorre A (2011) Serratia symbiotica from the aphid Cinara cedri: a missing link from facultative to obligate insect endosymbiont. *PLoS Genetics*, **7**, e1002357.

Lohse M, Bolger AM, Nagel A *et al.* (2012) ROBINA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research*, **40**, W622–W627.

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

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

© 2014 John Wiley & Sons Ltd

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

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

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

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

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

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

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

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

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

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

Russell JA, Weldon S, Smith AH *et al.* (2013) Uncovering symbiont-driven genetic diversity across North American pea aphids. *Molecular Ecology*, **22**, 2045–2059.

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

Shehzad WM, Riaz T, Nawaz MA *et al.* (2012) Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*, **21**, 1951–1965.

Simon J-C, Boutin S, Tsuchida T *et al.* (2011) Facultative symbiont infections affect aphid reproduction. *PLoS ONE*, **6**, e21831.

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

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

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

Sunderland KD, Crook NE, Stacey DL, Fuller BJ (1987) A study of feeding by polyphagous predators on cereal aphids using ELISA and gut dissection. *Journal of Applied Ecology*, **24**, 907–933.

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

Tarailo-Graovac M, Chen N (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, **4**, 4.10.1–4.10.14.

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

Timmermans MJTN, Vogler AP (2012) Phylogenetically informative rearrangements in mitochondrial genomes of Coleoptera, and monophyly of aquatic elateriform beetles (Dryopoidea). *Molecular Phylogenetics and Evolution*, **63**, 299–304.

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

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

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

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

Vilcinskas A, Stoecker K, Schmidtberg H, Röhrich CR, Vogel H (2013) Invasive harlequin ladybird carries biological weapons against native competitors. *Science*, **340**, 862–863.

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

Wheeler TJ, Clements J, Eddy SR *et al.* (2013) Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Research*, **41**, D70–D82.

Weinert LA, Tinsley MC, Temperley M, Jiggins FM (2007) Are we underestimating the diversity and incidence of insect bacterial symbionts? A case study in ladybird beetles. *Biology Letters*, **3**, 678–681. doi:10.1098/rsbl.2007.0373.

Wille BD, Hartman GL (2009) Two species of symbiotic bacteria present in the soybean aphid (Hemiptera: Aphididae). *Environmental Entomology*, **38**, 110–115.

Yu P, Henry CM, Leung SM *et al.* (2011) IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, **28**, 1420–1428.

Zeale MRK, Butlin RK, Barker GA, Lees DC, Jones G (2010) Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*, **11**, 236–244.

D.P.P. and A.P.V. designed the research. D.A.A. conducted the field sampling, the feeding bioassay and the decay statistical data analyses. D.P.P. performed the molecular biology experiments. D.P.P. and B.L. analysed the bioinformatics data. D.P.P., D.A.A., B.L. and A.P.V. wrote the manuscript. D.P.P., D.A.A., B.L., A.P.V., C.S.S.P. and E.R.S. revised it.

Data accessibility

The metagenomic data are available at Dryad (doi:10.5061/dryad.n6278). The *Harmonia axyridis* mitogenome sequence is available at GenBank (Accession Code KJ778886). The Bioinformatics scripts are provided in the online Supporting Information.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Symbiont database obtained from NCBI for symbiont detection. A sample of the GenBank accession numbers is provided for each taxon.

© 2014 John Wiley & Sons Ltd

Table S2. Genes detected in the aphid reads matching to the *A. pisum* nuclear genome.

Table S3. Repeats detected in the selected aphid reads matching to *A. pisum* nuclear genome.

Table S4. Minimum number of infected aphids (of the 10 consumed by beetles at each time point) predicted from the read patterns.

Fig. S1. Coverage of the predator mtDNA, *H. axyridis* (15,319 bp, GenBank accession number KJ778886). The tRNA genes are represented by amino acid single letter codes.

Fig. S2. Detection of the obligate aphid symbiont *B. aphidicola*.

© 2014 John Wiley & Sons Ltd