This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: http://orca.cf.ac.uk/52924/

This is the author’s version of a work that was submitted to / accepted for publication.

Citation for final published version:


Publishers page: http://dx.doi.org/10.1111/1755-0998.12156 <http://dx.doi.org/10.1111/1755-0998.12156>

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
A pragmatic approach to the analysis of diets of generalist predators: The use of next generation sequencing with no blocking probes

J. Piñol\textsuperscript{1,2,3}, V. San Andrés\textsuperscript{3}, E. L. Clare\textsuperscript{3,4}, G. Mir\textsuperscript{5}, W.O.C. Symondson\textsuperscript{3}

(1) Univ. Autònoma Barcelona, Cerdanyola del Vallès, 08193, Spain
(2) CREAF, Cerdanyola del Vallès, 08193, Spain
(3) Cardiff School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue, Cardiff CF10 3AX, UK
(4) School of Biological and Chemical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK
(5) Centre de Recerca en Agrigenòmica (CRAG) CSIC IRTA UAB UB, Cerdanyola del Vallès 08193, Spain

Keywords: crop ecology, gut content analysis, linyphiid spiders, predator-prey interactions

Corresponding Author:
Josep Piñol
CREAF, Univ. Autònoma Barcelona, Cerdanyola del Vallès 08193, Spain
E-mail: Josep.Pinol@uab.es

Running title: Diets through NGS with no blocking probes
Abstract

Predicting whether a predator is capable of affecting the dynamics of a prey species in the field implies the analysis of the complete diet of the predator, not simply rates of predation on a target taxon. Here we employed the Ion Torrent Next Generation Sequencing technology to investigate the diet of a generalist arthropod predator. A complete dietary analysis requires the use of general primers, but these will also amplify the predator unless suppressed using a blocking probe. However, blocking probes can potentially block other species, particularly if they are phylogenetically close. Here, we aimed to demonstrate that enough prey sequence could be obtained without blocking probes. In communities with many predators this approach obviates the need to design and test numerous blocking primers, thus making analysis of complex community food webs a viable proposition. We applied this approach to the analysis of predation by the linyphiid spider *Oedothorax fuscus* in an arable field. We obtained over two million raw reads. After discarding the low quality and predator reads, the libraries still contained over 61,000 prey reads (3% of the raw reads; 6% of reads passing quality control). The libraries were rich in Collembola, Lepidoptera, Diptera, and Nematoda. They also contained sequences derived from several spider species and from horticultural pests (aphids). *Oedothorax fuscus* is common in UK cereal fields and the results showed that it is exploiting a wide range of prey. Next Generation Sequencing using general primers but without blocking probes provided ample sequences for analysis of the prey range of this spider and proved to be a simple and inexpensive approach.
Introduction

Critical questions in trophic ecology centre around the choices that predators and herbivores make. For example, what a predator eats depends upon its ability to capture and subdue a target prey species, with predation rates being affected by relative densities of predators and prey, prey handling times and the shape of functional response curves (Symondson et al. 2002). However, equally important is the availability of ‘non-target’ alternative prey (Harwood et al. 2004). Prey ‘choice’ in the field is highly complex, where the densities and diversity of prey available to the predator are constantly changing over time and space (Symondson et al. 2002). At its simplest, what a predator eats in a given environment at any moment in time depends upon what is available (quantitatively and qualitatively) and accessible to the predator.

Understanding and predicting whether a predator is capable of affecting the dynamics of a prey species, and under what circumstances it manages to do so in the field, often require techniques capable of analysing the complete diets of the predators, not simply rates of predation on a target taxon, such as a particular crop pest. The introduction of DNA-based techniques (Asahida et al. 1997; Agustí et al. 1999; Zaidi et al. 1999) has substantially improved the study of predation, and such approaches are now widely used to study a range of trophic relationships (reviewed in Symondson 2002; King et al. 2008; Pompanon et al. 2012). This technology provides precise information on whether a predator species feeds on a particular target prey or which parasitoid is present within a host (Agustí et al. 2005; Traugott et al. 2008). However, the now classical approach, based on prey-specific primers, is of limited utility for generalist predators as the
analysis has to be repeated for a large number of potential, and often unknown, prey. The use of multiplexing can make this process more efficient (Harper et al. 2005, King et al. 2010, Traugott et al. 2012, Davey et al. 2013) but still limits the analysis to predicted prey rather than neutrally screening for all prey consumed.

An alternative approach is to use general primers that amplify a range of dietary components simultaneously. Initially this was followed by cloning and sequencing, and has been applied to studies of both herbivory (e.g. Poinar et al. 1998, Passmore et al. 2006, Bradley et al. 2007) and predation (e.g. Sutherland 2000, Jarman et al. 2004, Blankenship & Yayanos 2005, Deagle et al. 2007). The recent development of Next Generation Sequencing (NGS) technologies provide a more efficient means of rapidly gathering a mass of information on the dietary ranges of generalist predators and herbivores (Pompanon et al. 2012). Using NGS it is possible, at least in theory, to sequence all prey species present in predator gut or faecal samples simultaneously.

This technology has mainly been used to study the diets of vertebrate predators and herbivores, such as seals (Deagle et al. 2009), penguins (Deagle et al. 2010), bats (Bohmann et al. 2011), chamois (Raye et al. 2011), snow leopards (Shehzad et al. 2012), lizards (Brown et al. 2012) and bison (Kowalczyk et al. 2011), but has also been tested on invertebrates (Valentini et al. 2009; Boyer et al. 2012).

The use of NGS to study predator diets makes use of universal primers, able to amplify a wide range of prey. As these primers also amplify predator DNA, especially when prey and predator are phylogenetically close, the PCR product may be dominated by predator
instead of prey DNA. To overcome this problem predator-specific blocking primers are applied (Vestheim & Harman, 2008; Deagle et al., 2009, 2010). However, 'universal' primers do not in practice amplify everything equally while blocking primers are rarely specific. All 'universal' primers used in species-rich mixtures are likely to have more affinity for some species than for others. Similarly, 'specific' blocking primers could block to some degree prey species in addition to the targeted predator, especially if prey and predators are closely related (e.g. spiders feeding on other spiders). The difference in affinity between the 'universal' primer and the 'specific' blocking-primer for different species may be small, but after a number of cycles the PCR product is going to be enriched in some species and impoverished in others, in comparison to the original mixture. Thus, the use of universal primers and predator-specific blocking primers introduces two different biases in the analysis of predator diets.

Here we propose the use of NGS to study animal diets with no predator-specific blocking primers. By doing so, we get rid of one of the two biases described above. The final result will be dominated by reads of the predator, but the problem is overcome by the formidable sequencing capacity of modern NGS platforms. This 'brute force' method obviates the need to design and extensively test blocking primers. The latter may be worthwhile for ecological studies involving single predator species and have been used effectively (reviewed in Pompanon et al. 2012). However, our aim was to develop an approach that would be applicable to complex invertebrate food webs involving multiple predator species. Designing and testing blocking primers for dozens of different predators would be both impractical and expensive.
Spiders, especially very small species such as the linyphiids, are a group for which knowledge of their diets is often particularly difficult to obtain. They are fluid feeders, precluding the slow and taxonomically challenging classical approach of microscopically examining prey remains in predator gut samples. Many species hunt nocturnally, making direct observations of predation events difficult if not impossible. Some species build webs, from which prey can be collected, but not all captured prey are actually eaten. However, many spiders are not web builders while others hunt away from their webs, particularly at night (De Keer & Maelfait 1987; Alderweireldt 1994). There is a clear need for objective analytical methods, both for a better understanding of spider ecology generally and for applied research into the role of spiders as predators of major crop pests.

Here we tested the ‘brute force’ approach through the analysis of the diet of the linyphiid *Oedothorax fuscus* in an arable field. This small epigeal spider is widespread in Europe, living amongst thick vegetation in a range of habitats including field margins and arable crops. Most spiders are generalist predators of insects and other spiders (Pekár et al. 2012), making them appropriate targets for NGS of their complete dietary ranges. The abundance of *O. fuscus* in crops makes it a potentially useful biocontrol agent, particularly of aphids. As discussed in Pompanon et al. (2012), NGS provides an excellent tool for initial screening of predators or herbivores. Though not quantitative (Pompanon et al. 2012), it can provide an invaluable guide to the composition and range of species consumed. Our aim was to evaluate the technology and to determine the
dietary range of this elusive predator, revealing the breadth of its feeding ecology and hence its potential as a biocontrol agent in cereal fields.

**Methods**

*Collection of samples and DNA extraction*

Spiders were collected from a grass-dominated boundary strip surrounding a crop of winter barley at Burdons Farm, Wenvoe (51.439°N, 3.271°W), near Cardiff, Wales, UK. We sampled spiders six times from August to November 2011 using between six and ten quadrats (0.25 m$^2$) 10 m apart on each sampling date. We collected spiders with an entomological pooter and transferred them immediately to separate micro-centrifuge tubes. Back at the laboratory, the spiders were transferred to new Eppendorf tubes, in absolute ethanol, and kept at -80°C until needed.

The spiders were identified under a binocular microscope following Roberts (1996). We extracted the DNA from whole homogenized *O. fuscus* according to the “animal tissue” protocol of the Qiagen DNeasy tissue kit (Qiagen Ltd, Crawley, UK) following manufacturer instructions. For each batch of extractions a negative control was included. DNA quality was checked on 2% agarose gel. We then prepared a single pooled sample consisting of 5 µL from each individual extraction (15 to 20 individuals from each one of the six sampling events, a total of 109 individual extractions). The analysis of this sample was intended to assess the diet breadth of *O. fuscus* at a population level over a four month period.
Selection of universal invertebrate primers

We amplified arthropod DNA from spider extracts using the general invertebrate COI primers ZBJ-ArtF1c and ZBJ-ArtR2c (Zeale et al. 2011; Bohmann et al. 2011). These yielded a 157 bp amplicon located within the COI barcode region (Folmer et al. 1994) and have been shown to amplify spiders and members of a wide range of insect orders (Table S1).

Lab procedure

We prepared the samples with fusion primers following Ion Torrent recommendations for bidirectional sequencing (Ion Amplicon Library Preparation, Fusion Method). Briefly, two pairs of primers were designed (i) Ion Torrent primer A linked to the specific forward primer with Ion Torrent primer trP1 linked to the specific reverse primer (from now on “forward”), and the opposite (ii) Ion Torrent primer trP1 linked to the specific forward primer with Ion Torrent primer A linked to the specific reverse primer (from now on “reverse”) (Table S2).

The DNA was amplified in 50 μl reaction volumes containing 3 μl of template DNA, 45 μl of Platinum PCR Supermix High Fidelity (Invitrogen Corporation), and 1 μl of each pair of 10 μM fusion primers. We ran 40 cycles at 94°C for 30 s, 45°C for 45 s and 68°C for 45 s following an initial denaturation step at 94°C for 5 min and before a final extension step at 68°C for 10 min. The PCR product was purified using the QIA-quick PCR Purification Kit (Qiagen). Two PCR replicates of each forward and reverse reactions
were conducted.

We checked the PCR products from the forward and reverse design separately. We selected fragments of the expected size (E-Gel® Size Select 2% Agarose Gel, Invitrogen), quantified the size-selected DNA (DNA High Sensitivity kit, Bioanalyzer 2100, Agilent Technologies) and prepared an equimolar pool of the forward and reverse libraries. Then, we diluted, amplified (template amplification, Ion One Touch System) and sequenced this pool on a PGM as described by the manufacturer (Ion Torrent). As a first test of the technique we used the smallest Ion Torrent 314 chip, that should provide approximately $10^5$ DNA sequences. Based on the obtained results, we next ran a 318 chip in order to increase the sequencing depth and to lower the cost per read (the 318 chip can potentially increase the yield 50 times over the 314 chip). The two forward runs were conducted with the product of two different PCR reactions using the same DNA template and the two reverse runs with the product of a single third PCR reaction using the same DNA template as in the forward reactions. The first run was amplified with the Ion One Touch System whereas the second one was amplified with the upgraded version Ion One Touch System DL. In all runs we used the sequencing chemistry for 200 bp read length and version 2.2 of the Torrent Suite software for base calling (Ion Torrent, Life Technologies).

**Processing and analysis of data**

We divided the process into three steps: (1) Quality control and preprocessing; (2) comparison of the reads with sequence databases; and (3) taxonomic assignment of
We separated the forward and reverse reads into two FASTQ files using a purpose-made perl script. We eliminated the primer sequence from the 5' end of each read using TagCleaner (Schmieder et al. 2011). Then we trimmed (truncated to 150 bp at the end opposite of the sequencing primer), discarded those with a mean quality score lower than 25 or shorter than 150 bp, and downloaded the sequences in FASTA format (all using PRINSEQ, Schmieder & Edwards 2011). Both TagCleaner and PRINSEQ were run as web services at http://edwards.sdsu.edu/tagcleaner and http://edwards.sdsu.edu/cgi-bin/prinseq/prinseq.cgi, respectively.

To make downstream computation simpler, the FASTA files obtained above were first visually inspected with a general-purpose text editor for common sequences. We then used another purpose-made perl script to separate in two different files the common sequence from the rest (all relevant FASTQ files and perl code are available in Dryad). This step was iterated several times until the file containing the rest of the sequences was small enough (< 5,000 reads) to be BLASTed directly at the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). In particular, we used the nucleotide BLAST (blastn; Zhang et al. 2000) optimized for very similar sequences (megablast) on the nucleotide collection (nr/nt) that includes all GenBank+EMBL+DDBJ+PDB sequences. The BLAST parameters were those provided by default.

We imported the output from the BLAST alignment into MEGAN (MEtaGenomics
ANalyzer; Huson et al. 2007) to compute and explore the taxonomical content of the
data set, employing the NCBI taxonomy to summarize and order the results. In order to
assign a taxon to each sequence in the BLAST-files, MEGAN processes the BLAST
data of each sequence to determine all the hits. We discarded hits below the threshold
for the bit-score of hits (min. score = 50). We also discarded hits due to the threshold for
the maximum percentage (top percentage = 10) by which the score of a hit may fall
below the best score achieved for the sequence. After collecting the hits that exceeded
the thresholds, MEGAN found the lowest node that encompassed all these hits using
the LCA-assignment algorithm (LCA = Lowest Common Ancestor) to assign sequences
to taxa. We set the minimum support to a conservative value of 5, implying that a
minimum of five sequences had to be assigned to a taxon to appear in the final
cladogram.

The files (or a subset of them if the file was too big) containing the common sequences
were also BLASTed and then imported to MEGAN to explore their taxonomic content.
The reported result (Table 1) is the sum for each taxon of the number of reads provided
by MEGAN from each one of the files in which the original FASTA file was divided.

Results

Overall performance.
The 314 run produced a FASTQ file with 60,771 reads (52% forward and 48% reverse).
The quality control process reduced the number of “good” reads to 17,096 forward and
10,157 reverse reads. As expected, most reads belonged to the predator itself (14,020
forward and 6,905 reverse reads) and some produced no hit in BLAST or the result was uninformative. Overall, the 314 run produced 2,362 forward and 827 reverse reads useful to describe the diet of *O. fuscus*.

The 318 run produced a FASTQ file with 2,162,102 reads (50% forward and 50% reverse). The quality control process reduced the number of “good” reads to 580,131 forward and 390,976 reverse reads. Again, most reads belonged to the predator itself or were uninformative, so the overall result were 38,274 forward and 19,804 reverse reads useful to describe the diet of *O. fuscus*.

**Oedothorax fuscus diet**

The sequences of the *O. fuscus* diet obtained with the forward sequencing on the 314 chip were dominated by Collembola (69%; Table 1). There were also many sequences of spiders (other than *O. fuscus*), mostly of the thomisid *Xysticus* sp. There were also reads of lepidopterans, aphids, and nematodes. The forward reads obtained with the 318 chip were also dominated by Collembola (91%; Table 1), but there were very few spiders that could be assigned unambiguously to taxa that were different from *O. fuscus*. Lepidopteran, dipteran, and nematode reads were abundant, but aphids were absent. Finally, there were also a several reads of a cockroach and of an Auchenorrhyncha hemipteran (probably a leafhopper).

The two reverse runs on the 314 and 318 chips were conducted on the same PCR product, so they produced very similar results and are described together. Again, most
sequences belonged to Collembola (66%, considering both 314 and 318 runs; Table 1).

Spiders of three different linyphiid (other than O. fuscus) were also present: Tenuiphantes tenuis, Bathyphantes gracillis, and Centromerita bicolor. Dipterans were the second most abundant group in number of sequences (18%), but they were different from those obtained in the forward runs. Nematodes, lepidopterans and ephemerupterans were also present, but hemipterans and cockroaches were absent.

In summary, the four sequencing runs (two forward, two reverse) performed on the product of three different PCR reactions from the same DNA template, showed that Collembola, Nematoda, Araneae, and Lepidoptera were always present in the diet of O. fuscus. The Collembola belonged to at least three families, Isotomidae, Tomoceridae, and Sminthuridae, but most of them could only be resolved to Entomobryoidea. By contrast, the common springtail Parisotoma notabilis (Isotomidae) was clearly identified in all runs. Five genera of three different families of spiders (Thomisidae, Therididae, and Linyphidae) were found in the diet of O. fuscus, but nematodes and lepidopterans could only be resolved to higher groups. Dipterans were also present in all but one reaction, whereas Aphididae, Auchenorrhyncha, Blattodea and Ephemeroptera were only present in one reaction each.

Discussion

These results demonstrate that NGS technology using general primers without a predator-specific blocking primer can be an effective way to study the diet of invertebrates such as spiders. This method is also simple and inexpensive and
particularly effective when used with the 318 Ion Torrent chip that provided more than
two million DNA sequences. As expected, most of these belonged to the predator itself,
but there were still ca. 58,000 sequences from prey, far more than enough to describe
the diet of this spider.

The analysis provided evidence of predation on pests, non-pest prey and intraguild
predation (see below) while the species detected in the diet of *O. fuscus* were those
commonly found in UK cereal fields. Because the identification of these sequences
relies on the existence of taxonomically validated reference libraries, this work can only
be as precise as the sequence databases available. As such, more work needs to be
done on the barcoding of organisms found in arable crops to improve our ability to fully
identify these sequences.

**Oedothorax fuscus diet**

The diets of linyphiid spiders have been studied previously in cereal crops using other
techniques and these too show the importance of Collembola in their diets. Agustí *et al.*
(2003), using species-specific primers, reported the consumption of Collembola by a
guild of linyphiid spiders that included *Oedothorax* spp. We confirmed here that
Collembola are common prey of *O. fuscus*. Amongst the Linyphiidae, the Erigoninae,
which includes *O. fuscus*, were shown to build their webs in areas of high Collembola
density (Harwood *et al.* 2003). When their webs were substituted with sticky traps,
Collembola were the dominant prey captured.
Of great interest was the extent of intraguild predation, a widespread phenomenon amongst spiders (Pekár et al. 2012). Such intraguild effects are likely to negatively affect control of a shared prey, such as aphids. San Andres et al. have confirmed that O. fuscus preys on other spiders, as do other species in the Linyphiidae (unpublished data).

There were some taxa that have to be considered with caution: (i) All runs showed high numbers of nematode reads, a group of organisms certainly abundant in soil habitats but which are unlikely prey of spiders. They could be internal or external parasites of the spider or of some of their prey (Noordam et al. 1998). Another likely explanation may be that this is secondary predation, as nematodes are major prey of many species of Collembola (Read et al. 2006). (ii) The Auchenorryncha reported in Table 1 was in fact attributed by BLAST plus MEGAN to the cicada Psithyristria crassinervis, but this genus is endemic from Luzon, in the Phillipines (Lee & Hill, 2010). For this reason, we reported the reads to the suborder level. (iii) Similarly, the Blattodea reported in Table 1 was unambiguously attributed (100% coverage; 100% identity; Genbank accession code HM996892.1 ) to the German cockroach Blatella germanica. Whilst this species can be found in Wales, its preferred habitat are buildings rather than cereal fields. For this reason, we reported the reads to the order level.

The reason for not attributing many sequences to genus or species levels is the lack of similar taxa in Genbank. This happened in this study with most nematodes and collembolans. Considering the enormous amount of Lepidoptera sequences stored in
Genbank, and previous success assigning species level taxonomy to sequences from this region (Razgour et al. 2011) it comes as a surprise that Lepidoptera reads could only be resolved to order level and that those reads were very similar to many different species of Lepidoptera. For example, a group of reads was very similar to *Glyphodes* nr. *stolalis* (Pyraloidea: Crambidae; 97% coverage; 98% identity; Genbank accession code JX970289.1), but also to *Eois chrysocraspedata* (Geometroidea: Geometridae; 97%; 97%; JX150916.1), and to *Hippia pronax* (Noctuoidea: Notodontidae; 97%; 97%; JN806871.1), among others. In this case, MEGAN attributed the sequence to order Lepidoptera. It remains unclear why we were unable to resolve species in MEGAN but it may be caused by differences in BLAST vs. the HMM assignment algorithm of BOLD (Ratnasingham & Hebert 2007). The approach of using short amplicons (necessary in many molecular analyses of diet to ensure DNA survival in the guts of predators and herbivores) could, paradoxically, be problematic in groups like Lepidoptera. In these groups, a longer amplicon or more than one genomic region should be considered to improve taxonomical resolution. Alternatively, a thorough knowledge of the local arthropod fauna, plus construction of a local barcode database should help to assign DNA sequences to species.

**Methodological issues - blocking primers**

As expected, a high proportion of the sequencing effort (>90%) was lost to co-amplification of the predator, but even so we obtained around 58,000 reads with the 318 chip that could be attributed to prey (or parasites). Clearly, the use of a blocking-primer of *O. fuscus* would have greatly increased the number of prey reads, but at the cost of
possible co-blocking of other prey, especially other spiders. Suppressing amplification of predator DNA may well improve detection of rare, low copy number prey (O’Rorke et al. 2012). This is not always an advantage, especially for invertebrate food webs, where weak links may provide little nutrition to a predator and are unlikely to affect overall predator-prey dynamics.

The use of predator-specific blocking primers in studies of diets has pros and cons and so it is open to debate. In some instances, it can be highly beneficial. For example, when the predator is phylogenetically distant from prey (seals or penguins eating fish or cephalopods; Deagle et al., 2009, 2010), the use of predator-specific blocking primers increases the yield of the sequencing reaction without co-blocking (it is hoped) any prey. However, when prey and predator are phylogenetically close the risk of non-specific co-blocking cannot be disregarded. Here we detected other spiders in O. fuscus diet, some even from the same family (i.e. Tenuiphantes, Bathyphantes), which could have been blocked. Careful design of the blocking primer considering all possible prey could solve the problem, but in a field setting this is impossible because the range of potential prey is often unknown. This also assumes that a sufficiently specific blocking site exists adjacent to the primer region, which is by no means certain.

We obtained large numbers of prey sequences despite the fact that the whole spider was homogenised. With larger invertebrate predators, such as carabid beetles or earwigs from which the contents of the fore-gut can be extracted, predator DNA is likely to be less dominant. In a previous study using general invertebrate primers and TGGE
(temperature gradient gel electrophoresis) separation of PCR products on gels, prey DNA was not swamped by predator DNA and dietary components were clearly distinguished (Harper et al. 2006). Thus the whole-spider approach provided a ‘worst case’ scenario; if it can work on spiders it should work on other small predators that cannot be dissected easily.

Another critical point for many is the cost per read. Here we conducted two runs, one with the low-end Ion Torrent chip (314) and one with the more powerful 318 chip. The total cost (at internal rates of the University) was, approximately, 400 euros for the 314 and 900 euros for the 318. Considering the total number of prey reads obtained (Table 1), the price per read was 0.13 euros/read for the 314 and 0.02 euros/read for the 318. These prices are massively lower than what could be achieved by cloning and sequencing, and much lower still should labour costs be included in the calculations.

It is likely that the proportions of reads for each taxon only loosely reflect the actual proportion of prey in the diet (Pompanon et al. 2012) but few empirical tests to demonstrate this have been done. DNA analyzed from faeces of captive penguins (Deagle et al., 2010) or seals (Deagle et. al., 2013) fed with a constant diet of three fish species did not reproduce the original proportions of the three species after NGS. Murray et al. (2011) found a good correspondence between the proportions of four species of fish eaten by captive penguins obtained by NGS vs. qPCR, suggesting that NGS does not necessarily introduce any new sources of bias (though qPCR has also
been problematic in some diet studies, see McCracken et al. 2012). Many of the sources of bias (e.g. amplification efficiency, differences in mtDNA copy number) affect both approaches, and indeed cloning and sequencing. The effect is likely to be exacerbated as the diversity of prey increases. Even the short primer tags used to identify individual samples and the stringency in the bioinformatic quality control step seems to induce biases in the quantitative final result (Deagle et al., 2013). With increased prey diversity, the chance of primer bias also increases, thus any attempt at quantification within a sample/library should be avoided. In this context, the numbers of reads of each taxon in Table 1 cannot be interpreted as the relative abundance of each prey in the diet of O. fuscus, but mainly as a record of the richness of taxa in the diet. While there is likely to be a trend such that higher numbers of reads will result from prey being eaten in greater quantities, this can be influenced by the species-specific survival of DNA during digestion (e.g. soft bodied prey may degrade while hard bodied are better preserved), the relative DNA content of a prey species (which can differ between genders, ages and reproductive states) and the actual size of prey (larger prey may yield more DNA).

**Methodological issues - repeatability**

Here we sequenced three different PCR products obtained with the same template DNA. All samples showed the importance of Collembola as well as a number of spiders (albeit different in each sample) in the diet of O. fuscus. All samples also contained sequences of lepidopterans and nematodes. However, dipterans, aphids, leafhoppers, cockroaches, and ephemeropterans were absent in at least one sample. It might be predicted that a greater diversity would be obtained with the 318 chip. This is generally true, but there
were exceptions such as the 157 reads for aphids and the 158 of *Xysticus* obtained with 314 forward sequences that did not appear in any other sample. In summary, the method showed a degree of repeatability for the most abundant prey items, but there was a lot of variation for other taxa. We attribute the observed differences to the stochastic nature of the PCR, with some instability of PCR dominance between runs. The results do suggest that (funds permitting) maximum information can be obtained by running the same DNA through NGS more than once.

The forward runs produced approximately the same number of raw reads as the reverse ones. However, the quality of the reads was poorer in the reverse runs, so the proportion of forward:reverse reads increased to ca. 3:2 after the quality control step. The elimination of predator and uninformative reads then increased the ratio to 2:1. Apart from this fact, the reverse runs did not differ more from the forward ones than the two forward runs (with the 314 and 318 chips) differed from each other.

Concluding remarks

Our proposal is therefore disarmingly simple. The protocol involves DNA extraction, PCR amplification with universal primers, deep-sequencing on a NGS platform, bioinformatic disposal of predator reads, BLAST matching and MEGAN taxon assignment. There is no need to sequence the region of interest in all prey to design blocking primers, saving time and money. It is not even necessary to sequence the predator itself, as it is going to appear many times in the output file after NGS. Consequently, when the price per read comes down in the future, we envisage this method as a template to resolve real food
webs in which a high number of species interact, a goal that is today unattainable at a reasonable cost in terms of time and money. This method has the advantage of eliminating the danger that predator blocking probes co-block DNA from other species. The loss of sequencing power caused by co-sequencing of dominant predator DNA is compensated for by the extremely high capacity of modern NGS platforms. NGS of prey remains is essentially an exploratory technique which, after application to pooled samples, allows the primary prey to be identified. This can be followed by rapid screening of large numbers of individuals using appropriate species-specific primers (discussed in Pompanon et al. 2012) and multiplexing (Harper et al. 2005) to obtain a quantitative measure of the numbers of predator individuals testing positive for each prey. The same thing could be done with individual tags, but where thousands of predators need to be screened this would currently be very expensive in terms of primer costs. Our data suggest that the ‘brute force’ approach for NGS is viable, but probably not always so. Further work with other predators and general primers would be needed to be sure that this approach could be applied more widely.

**Acknowledgements**

We thank several anonymous reviewers and Bruce Deagle for useful suggestions on the initial version of the manuscript. Financial help was provided by the MICINN grant CGL2010-18182. This work was developed during a sabbatical leave of JP to Cardiff University financed by the Ministry of Education of Spain. Núria Agustí participated in the early stages of the project. Enrique Blanco, Xavier Espadaler, and Laia Mestre kindly commented on early versions of the manuscript. We thank Mr Robert Reader of Burdons
Farm, Wenvoe, for allowing us to sample spiders on his land.

References


Davey JS, Vaughan IP, King RA, Bell JR, Bohan DA, Bruford MW, Holland JM, Symondson WOC (2013) Intraguild predation in winter wheat: prey choice by a common


**Authors contribution**

JP prepared the DNA for NGS, conducted the bioinformatics and drafted the manuscript. VSA designed and carried out the field sampling, identified the spiders and extracted the DNA from the spiders. ELC was an advisor on the bioinformatics. GM was responsible for the Ion Torrent sequencing. WOCS was overall supervisor of the project, coordinated the group and co-drafted the manuscript. All authors played a role in editing the final version of the paper.

**Data Accessibility**


**Supporting information**

Additional supporting information may be found in the online version of this article.

**Table S1.** Alignment of the Zeale et al. (2011) primers with some representative insects, spiders, and collembolans.

**Table S2.** Fusion primers for the forward and reverse sequencing in the Ion Torrent PGM.
Table 1. Informative sequences amplified from the predator *Oedothorax fuscus* using an Ion Torrent PGM sequencer. Both forward (fwd) and reverse (rvr) reads using the 314 or the 318 chips are given, as well as the sum of all them.

<table>
<thead>
<tr>
<th>Broader group</th>
<th>Intermediate group</th>
<th>Family</th>
<th>Genus or species</th>
<th>314-fwd</th>
<th>318-fwd</th>
<th>314-rvr</th>
<th>318-rvr</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nematoda</td>
<td>Chromadorea</td>
<td></td>
<td></td>
<td>257</td>
<td>873</td>
<td>9</td>
<td>206</td>
<td>1345</td>
</tr>
<tr>
<td>Nematoda</td>
<td>Rhabditida</td>
<td></td>
<td></td>
<td>24</td>
<td>81</td>
<td>36</td>
<td>1353</td>
<td>1494</td>
</tr>
<tr>
<td>Collembola</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>220</td>
<td>0</td>
<td>0</td>
<td>230</td>
</tr>
<tr>
<td>Collembola</td>
<td>Entomobryoidea</td>
<td></td>
<td></td>
<td>1527</td>
<td>32594</td>
<td>491</td>
<td>11445</td>
<td>46057</td>
</tr>
<tr>
<td>Collembola</td>
<td>Entomobryoidea</td>
<td>Isotomidae</td>
<td><em>Parisotoma notabilis</em></td>
<td>102</td>
<td>2186</td>
<td>71</td>
<td>1417</td>
<td>3776</td>
</tr>
<tr>
<td>Collembola</td>
<td>Entomobryoidea</td>
<td>Isotomidae</td>
<td><em>Sminthurus viridis</em></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Collembola</td>
<td>Entomobryoidea</td>
<td>Tomoceridae</td>
<td></td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Araneae</td>
<td>Thomisidae</td>
<td></td>
<td></td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>Araneae</td>
<td>Thomisidae</td>
<td></td>
<td><em>Xysticus</em></td>
<td>158</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>158</td>
</tr>
<tr>
<td>Araneae</td>
<td>Therididae</td>
<td></td>
<td><em>Theridion</em></td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Araneae</td>
<td>Linyphiidae</td>
<td></td>
<td><em>Tenuiphanter tenuis</em></td>
<td>1</td>
<td>0</td>
<td>11</td>
<td>303</td>
<td>315</td>
</tr>
<tr>
<td>Araneae</td>
<td>Linyphiidae</td>
<td></td>
<td><em>Centromerita bicolor</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Araneae</td>
<td>Linyphiidae</td>
<td></td>
<td><em>Bathyphantes gracillis</em></td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>127</td>
<td>130</td>
</tr>
<tr>
<td>Araneae</td>
<td>Linyphiidae</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>74</td>
<td>1322</td>
<td>1396</td>
</tr>
<tr>
<td>Insecta</td>
<td>Diptera</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>42</td>
<td>1041</td>
<td>1083</td>
</tr>
<tr>
<td>Insecta</td>
<td>Diptera (Schizofora)</td>
<td></td>
<td></td>
<td>0</td>
<td>1066</td>
<td>0</td>
<td>0</td>
<td>1066</td>
</tr>
<tr>
<td>Insecta</td>
<td>Diptera</td>
<td>Cecidomyiidae</td>
<td></td>
<td>0</td>
<td>0</td>
<td>48</td>
<td>1009</td>
<td>1057</td>
</tr>
<tr>
<td>Insecta</td>
<td>Diptera</td>
<td>Anthomyiidae</td>
<td><em>Delia</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>148</td>
<td>148</td>
</tr>
<tr>
<td>Insecta</td>
<td>Lepidoptera</td>
<td></td>
<td></td>
<td>97</td>
<td>693</td>
<td>32</td>
<td>1251</td>
<td>2073</td>
</tr>
<tr>
<td>Insecta</td>
<td>Hemiptera</td>
<td>Aphididae</td>
<td><em>Aphis</em></td>
<td>157</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>157</td>
</tr>
<tr>
<td>Insecta</td>
<td>Hemiptera</td>
<td></td>
<td></td>
<td>0</td>
<td>274</td>
<td>0</td>
<td>0</td>
<td>274</td>
</tr>
<tr>
<td>Insecta</td>
<td>Blattodea</td>
<td></td>
<td></td>
<td>0</td>
<td>273</td>
<td>0</td>
<td>0</td>
<td>273</td>
</tr>
<tr>
<td>Insecta</td>
<td>Ephemeroptera</td>
<td>Heptageniidae</td>
<td><em>Afronurus</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>
**Table S1.** Alignment of the universal Zeale primers (ZBJ-ArtF1c and ZBJ-ArtR2c; Zeale et al., 2011) with some representative sequences of Collembola, Araneae, and Insecta. The mismatches between the primers and the sequence are highlighted in yellow. The forward primer is only showed partially, because their first 10 nucleotides (in the 5’ end; see Table s2 for the complete primer sequence) do not belong to the barcoding Folmer region. Genbank accession number, species, and family of all the reported sequences are the following ones:

<table>
<thead>
<tr>
<th>species family</th>
<th>species name</th>
<th>accession</th>
<th>Genbank accession number</th>
<th>species</th>
<th>family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collembola</td>
<td>Parisotoma notabilis</td>
<td>JN298134</td>
<td>JN298134</td>
<td>Parisotoma notabilis</td>
<td>Isotomidae</td>
</tr>
<tr>
<td>Collembola</td>
<td>Tomocerus minor</td>
<td>HM398042</td>
<td>HM398042</td>
<td>Tomocerus minor</td>
<td>Tomoceridae</td>
</tr>
<tr>
<td>Araneae</td>
<td>Erigone atra</td>
<td>HQ924416</td>
<td>HQ924416</td>
<td>Erigone atra</td>
<td>Linyphiidae</td>
</tr>
<tr>
<td>Araneae</td>
<td>Philodromus rufus vibrans</td>
<td>HQ979349</td>
<td>HQ979349</td>
<td>Philodromus rufus vibrans</td>
<td>Philodromidae</td>
</tr>
<tr>
<td>Hemiptera</td>
<td>Sitobion avenae</td>
<td>EU701907</td>
<td>EU701907</td>
<td>Sitobion avenae</td>
<td>Aphididae</td>
</tr>
<tr>
<td>Diptera</td>
<td>Drosophila simulans</td>
<td>HQ979110</td>
<td>HQ979110</td>
<td>Drosophila simulans</td>
<td>Drosophilidae</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>Coccinella septempunctata</td>
<td>JF296222</td>
<td>JF296222</td>
<td>Coccinella septempunctata</td>
<td>Coccinellidae</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>Cydia pomonella</td>
<td>JF859814</td>
<td>JF859814</td>
<td>Cydia pomonella</td>
<td>Tortricidae</td>
</tr>
<tr>
<td>Psocoptera</td>
<td>Psocoptera sp</td>
<td>HQ978921</td>
<td>HQ978921</td>
<td>Psocoptera sp</td>
<td>Psocoptera sp</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>Aphidius sp.</td>
<td>JN300351</td>
<td>JN300351</td>
<td>Aphidius sp.</td>
<td>Braconidae</td>
</tr>
<tr>
<td>Ephemeroptera</td>
<td>Baetis tricaudatus</td>
<td>JN297844</td>
<td>JN297844</td>
<td>Baetis tricaudatus</td>
<td>Baetidae</td>
</tr>
</tbody>
</table>
Table S1. Alignment of the universal Zeale primers (ZBJ-ArtF1c and ZBJ-ArtR2c; Zeale et al., 2011) with some representative sequences of Collembola, Araneae, and Insecta. The mismatches between the primers and the sequence are highlighted in yellow. The forward primer is only showed partially, because their first 10 nucleotides (in the 5' end; see Table s2 for the complete primer sequence) do not belong to the barcoding Folmer region. Genbank accession number, species, and family of all the reported sequences are the following ones:

<table>
<thead>
<tr>
<th>Genbank Accession</th>
<th>Species</th>
<th>Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>JN298134</td>
<td>Parisotoma notabilis</td>
<td>Isotomidae</td>
</tr>
<tr>
<td>HM398042</td>
<td>Tomocerus minor</td>
<td>Tomoceridae</td>
</tr>
<tr>
<td>HQ924416</td>
<td>Erigone atra</td>
<td>Linyphiidae</td>
</tr>
<tr>
<td>HQ979349</td>
<td>Philodromus rufus vibrans</td>
<td>Philodromidae</td>
</tr>
<tr>
<td>EU701907</td>
<td>Sitobion avenae</td>
<td>Aphididae</td>
</tr>
<tr>
<td>HQ979110</td>
<td>Drosophila simulans</td>
<td>Drosophilidae</td>
</tr>
<tr>
<td>JF296222</td>
<td>Coccinella septempunctata</td>
<td>Coccinellidae</td>
</tr>
<tr>
<td>JF859814</td>
<td>Cydia pomonella</td>
<td>Tortricidae</td>
</tr>
<tr>
<td>HQ978921</td>
<td>Psocoptera sp</td>
<td>Psocopterae sp</td>
</tr>
<tr>
<td>JN300351</td>
<td>Aphidius sp.</td>
<td>Braconidae</td>
</tr>
<tr>
<td>JN297844</td>
<td>Baetis tricaudatus</td>
<td>Baetidae</td>
</tr>
</tbody>
</table>