A new primer for metabarcoding of spider gut contents

Lafage D.[1,2*], Elbrecht V.[3], Cuff J.[4], Steinke D.[3], Hambäck P.A.[5], Erlandsson A[1].

1: Department of Environmental and Life Sciences / Biology, Karlstad University. Universitetsgatan 2, 65188 Karlstad, Sweden.
2: UMR ECOBIO, Université de Rennes 1, 263 Avenue Général Leclerc, 35700 Rennes, France
3: Centre for Biodiversity Genomics, University of Guelph, 50 Stone Road East, Guelph, Ontario, N1G 2W1, Canada
4: School of Biosciences, Cardiff University, Cardiff, CF10 3AT, UK
5: Department of Ecology, Environment and Plant Sciences, Stockholm University, SE-106 91 Stockholm, Sweden

*: Corresponding author: denis.lafage@kau.se

Running title: Primer for spider gut content metabarcoding

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27854v1 | CC BY 4.0 Open Access | rec: 12 Jul 2019, publ: 12 Jul 2019

Abstract

1. As a key predator group, spiders have received a lot of attention by food web ecologists in diverse fields such as pest control, pollutant transfers, and cross-ecosystem fluxes. The difficulty involved in studying their diet has led to the use of new technologies such as metabarcoding of gut contents. The amplification of a broad range of spider prey without amplifying spiders themselves is challenging and, until now, an efficient universal primer purposed for this has not existed.
2. We developed a novel forward primer (NoSpi2) targeting the cytochrome c oxidase subunit I gene. The primer was designed not to amplify spiders of the oval calamistrum clade (Lycosidae and closely related species) while still amplifying most other invertebrates. NoSpi2 was tested together with the reverse primer BR2 *in silico*, *in vitro* on single specimens of prey and spiders, on mock and malaise trap communities, and in an ecological application.
3. *In silico* evaluation predicted high primer bias for spiders of the oval calamistrum clade and low bias for all other invertebrates. These results were largely confirmed by *in vitro* tests. Additionally, some spider families were not amplified contrary to our expectations. We demonstrated a high efficiency for the primer pair NoSpi2/BR2 which recovered up to 94% of taxa in the mock community and 85% of the taxa detected by the best invertebrate primer pair known (BF3+BR2) for the malaise trap community. The field experiment showed that Lycosidae spider DNA is not amplified by the NoSpi2 primer set. It also demonstrated a broad range of detectable prey species. We found prey from 12 orders, 67 families and 117 species.
4. The ability of the NoSpi2/BR2 primer combination to reliably amplify prey species, without amplifying any predator reads, makes it an ideal choice for gut-content analysis for spider species of lycosids and closely related species, even enabling the homogenization of entire spider specimens without dissection. Given that the detected prey species included other spiders and carabid beetles, this primer could be used for not only dietary and biological control studies, but also to study intra-guild predation.

Keywords

Lycosidae; Zoropsidae; Ctenidae; Senoculidae; Oxyopidae; Pisauridae; Trechaleidae; Psechiridae; diet; gut content metabarcoding; molecular diet analysis; prey detection.

Introduction

The ecological interactions of spiders, the most abundant terrestrial arthropod predator group found on every life-supporting land mass, are pivotal in many ecological networks (Turnbull, 1973; Riechert & Lockley, 1984). As such, they have received a lot of attention by food web ecologists of diverse fields including pest control (Holland et al., 2016), pollutant transfers (Kraus et al., 2016; Walters, Otter, Kraus, & Mills, 2018) and cross-ecosystem fluxes (Lafage et al., 2019). Spiders feed on pre-digested fluids of their prey, restricting field studies of spider diet to direct observation, stable isotope analysis and molecular analysis (Symondson, 2002; Pompanon et al., 2012; Birkhofer et al., 2017). The inherent bias and the laborious nature of direct observation have led to the increased use of molecular techniques to investigate spider diet, of which DNA metabarcoding is currently among the most accurate and efficient for analysis of polyphagous generalist diets (Piñol, Senar, & Symondson, 2018).

General metabarcoding-based analyses of diet require PCR primers that amplify a broad range of potential prey species, but ideally without amplifying predator DNA. Given the degraded nature of gut content or faecal DNA in comparison to the relatively intact DNA of the predator, the latter can outcompete prey DNA in both PCR and sequencing (Vestheim & Jarman, 2008). The selection of primers is thus the most critical step for dietary metabarcoding studies (Piñol, Senar, & Symondson, 2018). Many studies use primer combinations that target specific species (Greenstone & Shufran, 2003; Kuusk, Cassel-Lundhagen, Kvarnheden, & Ekbom, 2008; Boreau de Roincé, Lavigne, Mandrin, Rollard, & Symondson, 2013) while others focus on a limited number of closely-related groups (e.g. Hosseini, Keller, Schmidt, & Li, 2011; Chapman, Schmidt, Welch, & Harwood, 2013; Hambäck, Weingartner, Dalén, Wirta, & Roslin, 2016). Whilst this approach does eliminate amplification of predator DNA, it also requires pre-existing knowledge of the diet and reduces the possibility of finding unexpected prey species.

Due to the degraded nature of prey DNA, gut content samples are often dominated by DNA of the predator (Shehzad et al., 2012). While universal primers can detect a wide range of prey items, detectability is far reduced if predator DNA is not excluded from the amplification process, resulting in much of the sequenced DNA being the predator (Piñol 2014, Vestheim 2008). Many studies thus employ blocking probes, which are predator-specific oligonucleotides that bind to the DNA without extension thereby inhibiting predator amplification (Vestheim & Jarman, 2008; Piñol, San Andrés, Clare, Mir, & Symondson, 2014). This can, however, also block closely-related prey species and increase amplification bias (Piñol, San Andrés, Clare, Mir, & Symondson, 2014). Other ways to reduce predator DNA include extraction from faeces (Sint, Thurner, Kaufmann, & Traugott, 2015), spider webs (Xu, Yen, Bowman, & Turner, 2015) or regurgitates. However, these methods either reduce the concentration and quality of DNA relative to gut content extraction (Agustí et al., 2003) or are not possible due to the fluid-feeding behaviour of spiders (Waldner, Sint, Juen, & Traugott, 2013; Kamenova et al., 2018). The amount of predator DNA in spider gut content extractions can also be limited by extraction of just the abdomen, which has a higher proportion of prey DNA, but nevertheless a majority of the extracted DNA will be from the predator (Krehenwinkel, Rödder, & Tautz, 2015). Similarly, size selection with magnetic beads can be used to limit the amount of intact longer predator DNA

present in an extract, leaving only shorter degraded prey DNA (Krehenwinkel, Rödder, & Tautz, 2015), but this removal is based wholly on size and may limit the prevalence of more recently ingested prey and does not remove degraded predator DNA. A better approach would be to use primers which do not amplify the DNA of the predator while still amplifying prey DNA. Lineage-specific primers have previously been designed for multiplex amplification of different prey lineages (Krehenwinkel et al., 2019), but no single primer pair has yet been developed for universal amplification of spider prey whilst avoiding predator amplification.

In the present study a novel forward PCR primer was designed to amplify spider prey DNA from spider gut content extracts without amplifying spider predator DNA. We specifically targeted lycosid spiders (and phylogenetically closely related species) which is the dominant spider group in the world both with respect to diversity and local abundance (Murphy et al., 2006). This primer together with a previously designed reverse primer are relevant for future metabarcoding-based analyses of spider diet.

Methods

Primer Development and *In Silico* Evaluation

Mitochondrial cytochrome c oxidase subunit 1 (COI) sequences for 15 freshwater invertebrate groups (see Vamos, Elbrecht, & Leese, 2017) and one spider genus (*Pardosa*) were downloaded and clustered using the 'PrimerMiner' package v0.18 (Elbrecht & Leese, 2017a). All following analyses were run in R 3.6 (R Core Team, 2019). Sequences were aligned in Geneious 8.1.7 (Kearse et al., 2012) using MAFFT v7.017 (Katoh, 2002). PrimerMiner's "selectivetrim" function was used to trim 26 bp of the HCO and 25 bp of the LCO binding sites and the alignment for each group was visualized with PrimerMiner to visually identify suitable primer binding sites. Sites conserved among target spider prey taxa but differing in *Pardosa* sequences were selected, and a primer was designed based on the most optimal combination of sites (Table 1 and Fig. 1). The resulting forward primer (NoSpi2) (Table 1) contains a high base degeneracy and binds at position 233 255 of the COI Folmer region (Folmer et al. 1994), taking advantage of sequence polymorphisms to enable amplification of target prey species and avoiding amplification of spiders of the genus *Pardosa*. Amplicon length in combination with the reverse primer BR2 (Table 1) (Elbrecht & Leese, 2017b) was 403bp.
PrimerMiner was also used to evaluate the primer pair (NoSpi2 and BR2) against alignments of 30 aquatic and terrestrial invertebrate groups and 103 spider families, using default tables for mismatch scoring. As a reference, penalty scores were also computed for the standard barcoding primers LCO1490 and HCO2198 (Folmer, Black, Hoeh, Lutz, & Vrijenhoek, 1994).

Table 1: Primer characteristics as determined with Primer3 2.3.7 implemented in Geneious v2019.0.4 with default settings.

Primer	Sequence (5'-3')	Source	Average Tm °C	Average GC %	Length bp
NoSpi2	TTYCCHCGWATAAAYAAYATAAG	This study	51.9	29.7	23
BR2	TCDGGRTGNCCRAARAAYCA	Elbrecht 2017	59.0	49.2	20

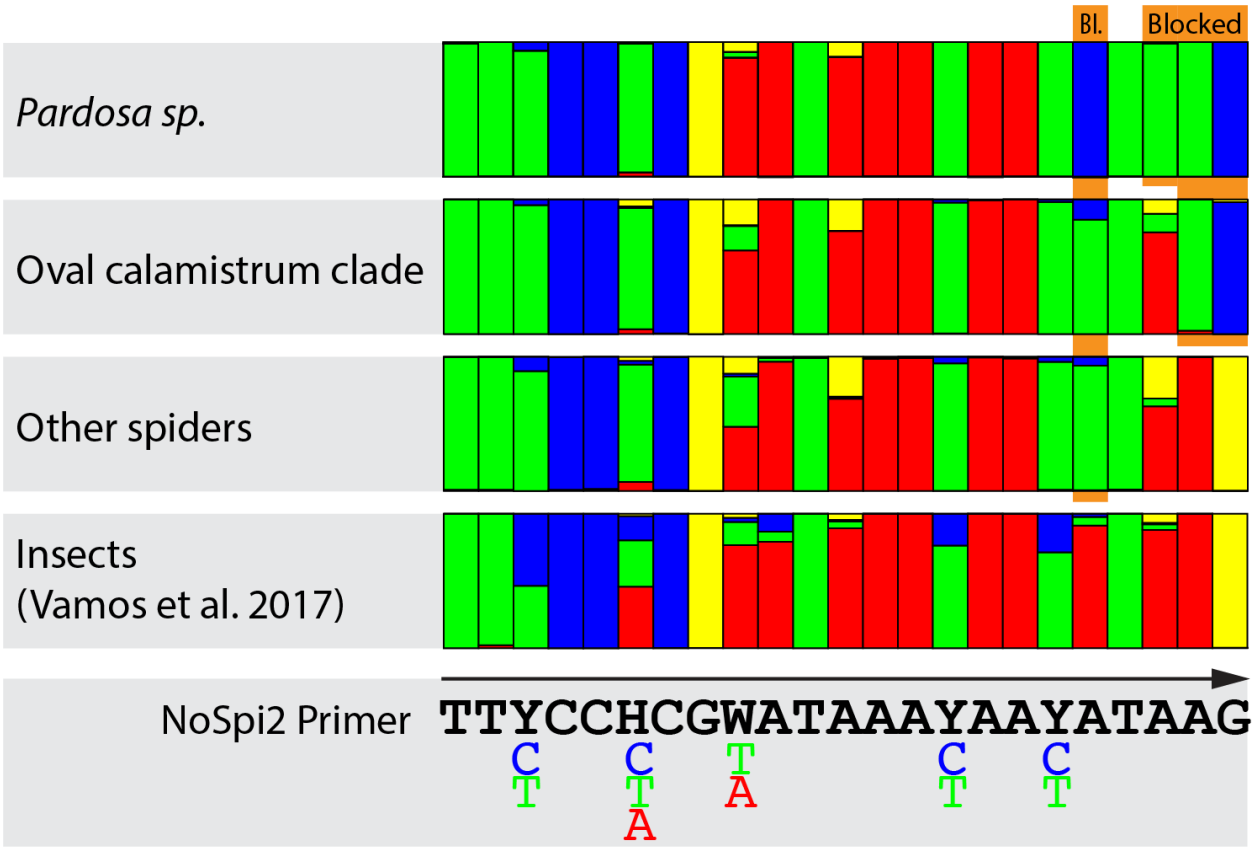

Figure 1: Bar plot showing NoSpi2 binding site variability of *Pardosa* sp., spiders of the oval calamistrum clade, other spider families and other insects, generated with PrimerMiner. Amplification of spiders of the genus *Pardosa* and the oval calamistrum clade is expected to be unlikely, due to mismatches at the 3" end of the NoSpi2 primer. Insect data from Vamos et al. (2017), base composition, green: thymine, blue: cytosine, red: adenine, yellow: guanine, blocked positions highlighted in orange.

In vitro evaluation

The primer pair was tested *in vitro* on 6 freshwater and 13 terrestrial invertebrate families (total

28 species), and 20 spider families (32 species) separately (Appendix S1 and S2). Freshwater invertebrates were selected from samples collected in 2017 from four Swedish rivers using drift nets. Spiders were selected from samples collected in 2017 in the riparian zone of the same four Swedish rivers using a vacuum sampler. Terrestrial invertebrates were selected from samples collected in 2016 in the United Kingdom from arable crops via vacuum sampler, and from decaying beech wood via hand-sorting. For herbivorous invertebrates, whole bodies were used, for predators only legs to restrict possible amplification of gut contents. To investigate possible competitive effects (interference of predator DNA with amplification of prey DNA), sample mixes with equal DNA concentrations (39 ng/µL) of starved Lycosidae (*Pardosa amentata*) and Collembola (Poduridae) or Trichoptera (Limnephiliidae) or Plecoptera (Leutridae) were also tested. In addition, DNA extracts from a starved Lycosidae (*Pardosa amentata*) at a concentration of 90.8 ng/µL supplemented with Trichoptera (Limnephiliidae) DNA in a decreasing concentration (lowest tested addition 0.007 ng/µL) were analysed.
DNA extraction of tissue samples was performed using the QIAamp DNA Micro Kit (Qiagen, Hilden, Germany), following manufacturer instructions. For amplification, the combination NoSpi2 (see table 1) and BR2 was used (Elbrecht & Leese, 2017b). PCR reactions were carried out in 25 µL reaction volumes containing 2 µL of DNA extract, 12.5 µL of PCR Multiplex mastermix (Qiagen, Hilden, Germany) and 2.5 µM of each primer. Thermocycler conditions were: initial denaturation at 95°C for 15 min; 30 cycles of: 30 sec at 94°C, 90 sec at 48°C and 90 sec at 72°C; and a final extension for 10 min at 72°C. Positive amplifications were confirmed by visual inspection of PCR products in 2% agarose gels. For positive amplifications PCR products were purified using ExoSAP-IT™ PCR Product Cleanup Reagent (Thermofisher Scientific, Massachusetts, US). In case of negative amplification, a second PCR with LCO1490/HCO2198 primers was run using the same conditions (different annealing temperature: 40°C) to ensure that the sample contained amplifiable DNA. Purified PCR products (positive samples with NoSpi2/BR2 and positive samples with LCO1490/HCO2198) were then Sanger sequenced and the resulting sequences processed using the sangeranalyse R package (v. 0.1) https://github.com/roblanf/sangeranalyseR. Low quality ends of sequences were automatically trimmed based on their quality (default cutoff of 0.0001). Subsequently, forward and reverse sequences were merged into a consensus sequence. If the quality of one of the sequences was low, only the other sequence was used. Consensus sequences were queried against NCBI GenBank using the blastn algorithm (Camacho et al., 2009) to retrieve species identity.

Primer assessment

In order to assess the arthropod detection efficiency of the primer pair, it was also used to metabarcode an insect mock sample (Braukmann et al., 2019) and a malaise trap sample from Ontario, Canada, both previously tested with 21 primer sets (Elbrecht et al., 2019, PrePrint). A two-step PCR was used to amplify and tag the DNA fragments. The first PCR was performed in 25 µL reaction volumes containing: 1 µL of DNA extract (12.5 ng) 12.5 µL of PCR Multiplex plus kit (Qiagen, Hilden, Germany), 0.5 µM of each primer (NoSpi2+BR2) and DNA-free water. The conditions for these PCRs were: initial denaturation at 95 °C for 5 min; 30 cycles of: 30 sec at

95 °C, 30 sec at 48 °C and 50 sec at 72 °C; and a final extension of 5 min at 72 °C. The second PCR was performed using fusion primers in 25 µL reaction volumes: 1 µL product from PCR1, 12.5 µL of PCR Multiplex (Qiagen, Hilden, Germany) 0.2 µM of each primer (P7_NoSpi2 CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCT GATTYCCHCGWATAAAYAAYATAAG combined with mock sample: P5_BR2_F, malaise sample: P5_BR2_X, Elbrecht & Steinke 2018). The second PCR conditions were: initial denaturation at 95°C for 5 min; 20 cycles of: 30 sec at 95°C, 30 sec at 48°C, 2 min at 72°C; and a final extension of 5 min at 72°C. PCR products were normalized and pooled using Sequal prep (Thermofisher Scientific, Massachusetts, US). Primer dimer were subsequently removed using SPRIselect cleanup according to the manufacturer protocol (0.76x SPRIselect used) (Beckman Coulter, Indianapolis, US). The final library contained 96 additional samples of mostly malaise samples and macrozoobenthos kick samples. Sequencing was carried out by the AAC Genomics Facility at the University of Guelph, Canada on an Illumina MiSeq using the 600 v3 cycle kit with 5% PhiX spike in. As fusion primer in-line tags were used, indexing was skipped and an additional 16 bp were added to the first read, leading to paired-end sequencing of 316 + 300 bp length. Bioinformatic processing was done following Elbrecht et al. (2019, PrePrint) using the JAMP pipeline.

Field experiment

To ascertain its performance in an ecological application, the primer pair was used to amplify spider gut DNA from a field experiment conducted in 2018. Lycosid spiders were sampled at three riparian sites along the Klarälven River (Sweden). Samples were collected weekly between May 15th and June 19th. In each location, spiders were collected from recently-flooded areas (n = 20) and non-flooded areas (n = 10). Individuals were collected using clean forceps and stored individually in 99.6% ethanol before freezing at -20°C.
Each adult was identified to species-level before sterile removal of its abdomen for DNA extraction using the QIAamp DNA Micro Kit, as described above. Extra care was taken to avoid contamination: each spider was manipulated in a sterile petri-dish and dissected with sterilized forceps and blades. Each step of the follow-up procedures such as DNA extraction, PCR reaction mixes as well as PCR product detection and isolation was also performed at separate locations to avoid contamination. PCR was performed with the primer pair NoSpi2/BR2 extended with universal Turesq Illumina adaptors (33 and 21 bp respectively) using the PCR conditions described (see in vitro evaluation part). Positive amplifications were confirmed by visual inspection of PCR products in 2% agarose gels. Samples exhibiting bands of the expected size were purified using Agencourt AMPure XP beads following manufacturer specifications (Beckman Coulter, Indianapolis, US). DNA concentration of the cleaned PCR products was determined using a Qubit fluorometer (Thermofisher, Massachusetts, US). A second PCR was performed to barcode the samples with dual indexes (13 forward x 10 reverse) extended with Illumina adapters (Hugerth et al., 2014). PCR reactions were carried out in 25 µL reaction volumes containing: 2 µL of the cleaned PCR product, 12.5 µL of KAPA Taq ReadyMix (2X) (Sigma-Aldrich, Missouri, US), 10µM of each primer. Thermocycler conditions were: an initial denaturation at 98°C for 2 min; 10 cycles of: 20 sec at 98°C, 30 sec at 62°C and 30 sec at 72°C; and a final extension for 2 min at 72°C. PCR product concentrations were

measured using a Qubit fluorometer before equimolar pooling (20.1 ng/μL). The pooled product was then purified using Agencourt AMPure XP beads following manufacturer specifications (Beckman Coulter, Indianapolis, US). The DNA library was sequenced in a flow cell on an Illumina MiSeq v3, PE 2x300, at the Science for Life laboratory, Sweden (www.ScilifeLab.se).

Sequences were processed using the 'dada2' package (Callahan et al., 2016) in R. The pipeline includes demultiplexing, filtering, trimming, dereplication, correction of errors, merging of forward and reverse sequences and clustering into amplicon sequence variant (ASV). All ASVs produced were screened against BOLD (Ratnasingham & Hebert 2007). Sequences with no match were subsequently queried against NCBI Genbank. Sequences were attributed at species level with a similarity match ≥ 98%, to the genus level with a similarity match ≥ 95%, to the family level with a similarity match ≥ 90% and to the order level with a similarity match ≥ 85% (Elbrecht, Vamos, Meissner, Aroviita, & Leese, 2017). If several species obtained identical similarity, the species with more occurrences in Sweden, based on GBIF data (http://www.gbif.org/), was selected. When several ASVs attributed to the same sample referred to the same species, the number of differences between sequences was computed. If the number of differences was higher than 50, the two ASVs were considered two different individuals. Finally, only species which are likely preyed upon by spiders were kept (i.e. bacteria and fungi sequences were discarded).

Results

Primer development and *in silico* evaluation

We designed a primer NoSpi2, which in combination with the primer BR2 (Elbrecht & Leese, 2017b) targets the Folmer region of the COI gene to generate a sequence of 403 bp. *In silico* evaluation of the primer pair showed high penalty scores for all spider families of the oval calamistrum clade (Fig. 2) with the strongest values obtained for Lycosidae.

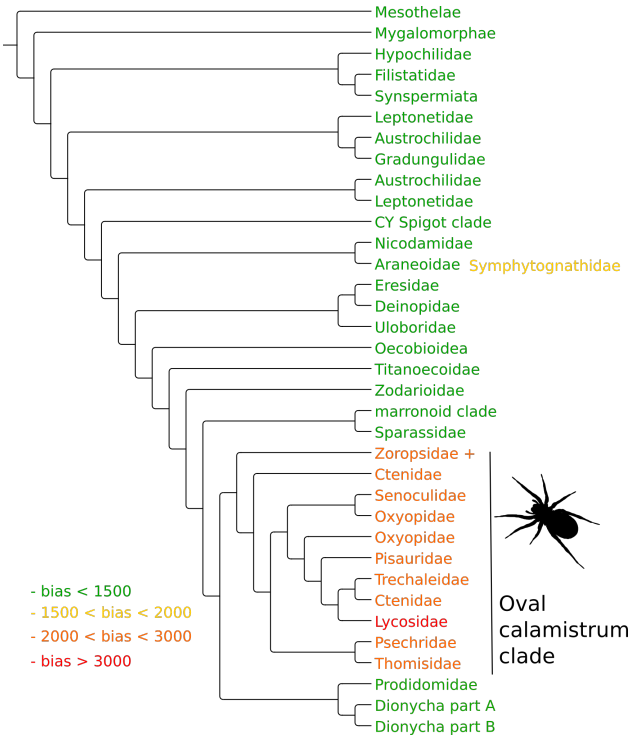

Figure 2: Summary of the phylogenetic tree of the world's spider clades/families from Wheeler et al. (2016). Colours correspond to penalty scores calculated with PrimerMiner for NoSpi2. The higher the penalty score the lower the likelihood of amplification.

In silico evaluation of the potential prey (8 aquatic and 22 terrestrial) showed mostly lower penalty values for NoSpi2 compared to LCO1490 and for BR2 compared to HCO2198 (Fig. 3). Among aquatic prey, NoSpi2 and BR2 always had lower penalty scores. Among terrestrial prey, LCO showed lower penalty scores for Embioptera, Strepsiptera and Archeognatha. BR2 showed lower penalty scores for all other terrestrial prey tested.

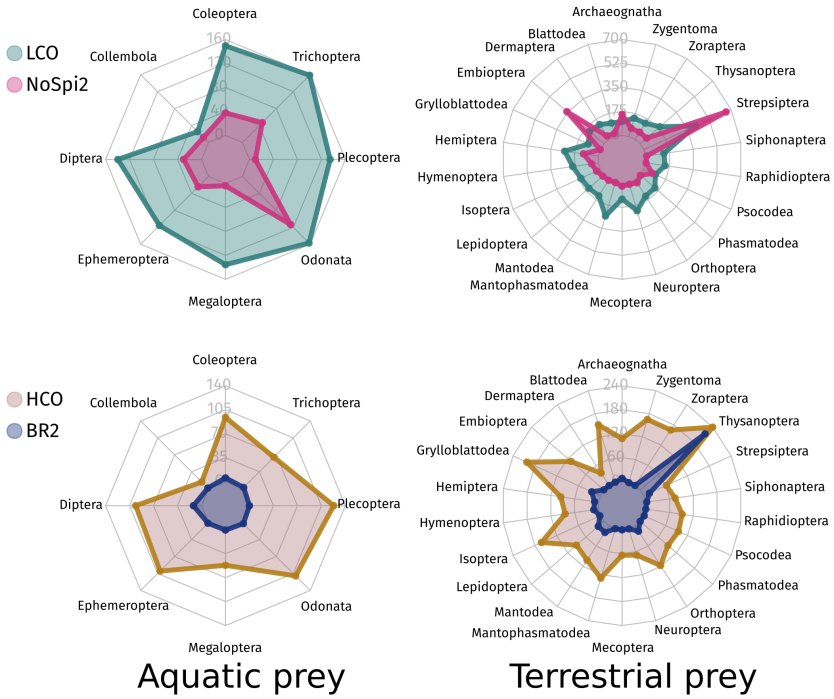

Figure 3: Radar charts of penalty scores for each forward and reverse primer. Scores are computed using PrimerMiner (Elbrecht & Leese, 2017a). Low scores correspond to high amplification efficiency.

Single specimen tests

The primer pair NoSpi2/BR2 was evaluated in vitro against 19 invertebrate families (28 species) and 20 spider families (32 species). It was able to amplify the correct sequence from 25 invertebrate species and 16 spider species (Table 2 and Appendix S3-8). Only one Hymenoptera species (*Tricopria* sp.) was not amplified. A sequence obtained from the PCR with LCO1490, identified as *Pardosa palustris,* was not amplified with NoSpi2/BR2. Two species (one Collembola and one Diptera) could not be identified by sequencing due to the low quality of the reads. The primer pair NoSpi2/BR2 produced contrasting results for spiders (Table 2). Several families were not amplified: Lycosidae, Amaurobidae, Clubionidae, Ctenidae, Dysderidae, Oonopidae, Oxyopidae, Philodromidae, Pholicidae, Pisauridae and Thomisidae. It appears that the Dysderidae sample was also negative with LCO/HCO2198 suggesting that no amplifiable DNA was present (i.e. no DNA present or PCR inhibition). The Oonopidae and Pisauridae samples were amplified by LCO1490/HCO2198 but the species could not be identified. Among the Linyphiidae species tested, only one sample was not amplified but the genus mismatch between the morphological and DNA identification suggests a sample error. All mixes for the assessment of a potential inhibitory effect of predator DNA and for the detection threshold of the primer were successfully amplified by NoSpi2/BR2 and provided accurate sequencing results.

Table 2: Summary of *in vitro* tests. +N/B: band present in the electrophoresis gel after PCR with NoSpi2/BR2 primers,- NB + LH: band absent in the electrophoresis gel after PCR with NoSpi2/BR2 but present with LCO1490/HCO2198. \% match: % similarity when the sequence is blasted against NCBI database. N/A corresponds to sequences with matching < 85\%. All hits have an e-value of $1e^{-6}$, and coverage > 96\%

Type	Organisms	PCR	Blast id	% match	Remark / GenBank nb
Preys	Acari	+N/B	Damaeidae sp.	100	
	Chilopoda - *Cryptops hortensis*	+N/B	*Cryptops hortensis*	99.3	
	Coleoptera - *Petrosticus melaniarius*	+N/B	*Pterostichus melanarius*	98.6	
	Coleoptera – *Bembidion bruxellense*	+N/B	*Bembidion bruxellense*	100	
	Coleoptera – *Bembidion* sp.	+N/B	*Bembidion gilvipes*	98.3	
	Coleoptera – *Elaphrus riparius*	+N/B	*Elaphrus riparius*	100	
	Collembola - *Sminthurus viridis*	+N/B	N/A		
	Collembola – Poduridae	+N/B	*Pogonognathellus flavescens*	98.2	
	Diptera – Chaboridae	+N/B	*Chaoborus flavicans*	100	
	Diptera – Chironomidae	+N/B	Chironomidae	89.7	MN106227
	Diptera – Simuliidae	+N/B	*Simulium intermedium*	98.5	
	Diptera – Simuliidae	+N/B	N/A		
	Ephemeroptera – Baetidae	+N/B	*Baetis niger*	96.3	MN106226
	Ephemeroptera – Leptophlebidae	+N/B	*Leptophlebia marginata*	99.6	Cutoff = 0.001
	Gastropoda - *Discus rotundatus*	+N/B	*Discus rotundatus*	92.7	Cutoff = 0.001
	Hemiptera – *Metopolophium* sp.	+N/B	*Metopolophium dirhodum*	100	
	Hemiptera - *Anthocoris nemorum*	+N/B	*Anthocoris nemorum*	100	
	Hymenoptera - *Trichopria* sp.	- NB + LH	*Pardosa palustris*	100	
	Isopoda - *Porcellio scaber*	+N/B	*Porcellio scaber*	100	
	Lepidoptera - *Euproctis similis*	+N/B	*Euproctis similis*	100	
	Megaloptera	+N/B	*Sialis fuliginosa*	100	
	Neuroptera - *Chrysoperla* sp.	+N/B	*Chrysoperla* sp.	99.1	
	Plecoptera – Leutridae	+N/B	*Leuctra fusca*	100	
	Pseudoscorpiones - *Chernes cimicoides*	+N/B	*Dinocheirus panzeri*	100	Possible misidentification
	Trichoptera – Limnephilidae	+N/B	*Chaetopteryx villosa*	100	MN106224
	Trichoptera – Phryganeidae	+N/B	*Oligotricha striata*	100	MN106225

	Trichoptera – Polycentropodidae	+N/B	*Plectrocnemia conspersa*	99.7	
Spiders	Agelenidae - *Eratigena arctica*	+N/B	*Eratigena atrica*	100	
	Amaurobiidae - *Amaurobius similis*	- NB - LH	*Amaurobius similis*	99.8	
	Araneidae - *Zygiella X-notata*	+N/B	Araneidae	86.7	
	Araneidae – *Araneus alsine*	+N/B	*Araneus alsine*	91.5	
	Clubionidae - *Clubiona corticalis*	- NB + LH	*Clubiona corticalis*	99.7	
	Ctenidae	- NB + LH	*Cupiennius salei*	94.1	
	Cybaeidae – *Cryphoeca sylvicola*	+N/B	*Cryphoeca silvicola*	97.1	
	Dysderidae - *Dysdera crocata*	- NB + LH			
	Hahniidae – *Hahnia nava*	+N/B	N/A		
	Linyphiidae - *Erigone atra*	+N/B	*Erigone atra*	99.7	
	Linyphiidae - *Oedothorax fuscus*	+N/B	N/A		
	Linyphiidae - *Tenuiphantes tenuis*	+N/B	*Tenuiphantes tenuis*	100	
	Linyphiidae - *Walckenaeria* sp.	- NB + LH	*Oedothorax fuscus*	99.4	
	Linyphiidae – *Bathyphantes nigrinus*	+N/B	*Bathyphantes nigrinus*	100	
	Linyphiidae – *Tenuiphantes cristatus*	+N/B	*Tenuiphantes cristatus*	97.5	
	Lycosidae – *Pardosa amentata*	- NB + LH	*Pardosa amentata*	100	
	Lycosidae – *Pardosa lugubris*	- NB + LH	*Pardosa lugubris*	100	
	Lycosidae – *Pardosa prativaga*	- NB + LH	*Pardosa prativaga*	99.8	
	Lycosidae – *Pardosa pullata*	- NB + LH	*Pardosa pullata*	98.8	
	Lycosidae – *Piratula hygrophila*	+N/B	*Neriene montana*	100	Prey
	Oonopidae - *Oonops pulcher*	- NB + LH	N/A		
	Oxyopidae – *Oxyopes salticus*	- NB + LH	*Oxyopes salticus*	99.5	
	Philodromidae - *Philodromus* sp.	- NB + LH	*Philodromus aureolus*	99.5	
	Pholcidae - *Psilochorus simoni*	- NB + LH	*Psilochorus simoni*	100	
	Pisauridae – *Pisaura mirabilis*	- NB + LH	N/A		
	Salticidae - *Neon reticulatus*	+N/B	*Neon reticulatus*	100	
	Tetragnathidae – *Pachygnatha degeeri*	+N/B	*Pachygnatha degeeri*	100	
	Tetragnathidae – *Tetragnatha* sp.	+N/B	*Homo sapiens*	100	Contamination
	Tetragnathidae – *Metellina merianae*	+N/B	*Metellina merianae*	100	
	Theridiidae – *Enoplognatha ovata*	+N/B	*Enoplognatha ovata*	99.7	
	Thomisidae - *Xysticus audax*	- NB + LH	*Xysticus audax*	100	
	Thomisidae – *Diaea dorsata*	+N/B	*Sialis fuliginosa*	96.2	Prey

Mix	Pardosa + Collembola	+N/B	*Pogonognathellus flavescens*	98
	Pardosa + Trichoptera	+N/B	*Chaetopteryx villosa*	100
	Pardosa + Plecoptera	+N/B	*Leuctra fusca*	100
	1 Pardosa + 1/30 Trichoptera	+N/B	*Chaetopteryx villosa*	100
	1 Pardosa + 1/3000 Trichoptera	+N/B	*Chaetopteryx villosa*	100

Metabarcoding of mock and malaise samples

The NoSpi2+BR2 primer set did recover up to 94.17% of taxa of the mock community, with an average of 310.7 (SD = 4.22) and 352.2 (SD = 2.51) taxa recovered at 10,000 and 100,000 reads sequencing depth, respectively (subsampled with 1,000 iterations). The same primer set recovered an average of 445.7 (SD = 9.60) and 678.7 (SD = 6.17) taxa of the malaise sample (again at 10,000 and 100,000 reads sequencing depth). This recovery represents approximately 10-15% lower taxon recovery than the currently most efficient universal invertebrate primer pair (e.g. BF3+BR2) in Elbrecht et al. (2019).

Field experiment

Among the 388 spiders sampled, 126 led to positive amplification results. We obtained 540 ASVs corresponding to 12 orders, 67 families and 117 species (Table 3 and Fig. 4, detailed table in Appendix S9 and raw ASVs in Appendix S10). Fifteen spider samples resulted in ASVs that had no match in BOLD nor NCBI GenBank. One ASV corresponded to a spider but not to predator DNA (*Clubiona lutescens*). No reads corresponded to predator DNA.

Table 3: Summary table showing the number of species and reads for each order found in 126 spider gut contents.

Class or Order	Number of species	Number of reads
Acari	3	7706
Araneae	1	1027
Coleoptera	5	33063
Collembola	4	6568
Diptera	42	242192
Hemiptera	32	116518
Hymenoptera	14	137023
Lepidoptera	6	16838
Opiliones	2	521
Stylommatophora	1	48
Thysanoptera	2	43
Trichoptera	2	4256

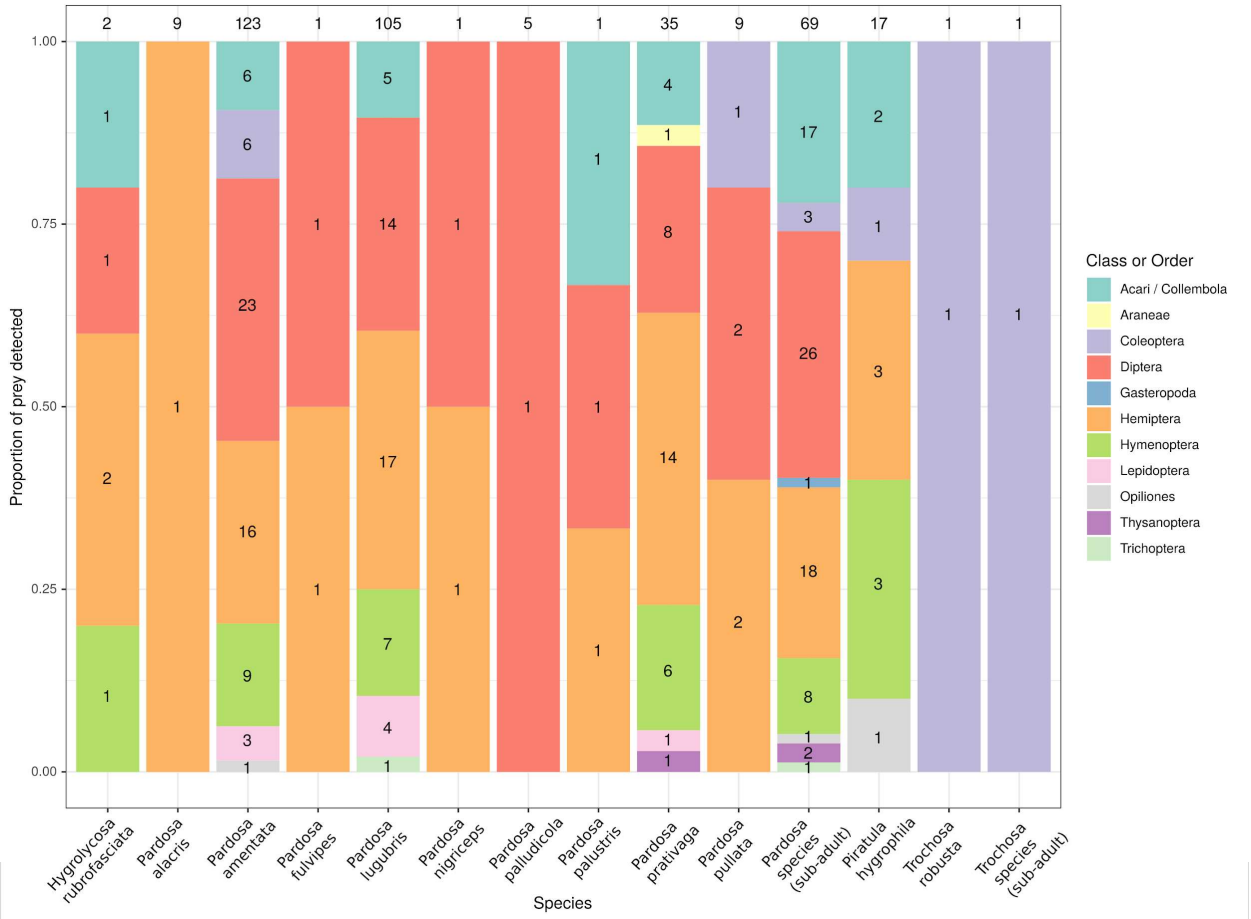

Figure 4: Proportion of prey orders detected per spider species. Numbers within bars represent the total number of ASVs per species. Numbers above the bars show the number of individuals tested per species.

Discussion

In this study, we developed a forward primer (NoSpi2) specifically designed for metabarcoding of gut contents of lycosid spiders and phylogenetically closely-related families of the oval calamistrum clade. Together with the previously-designed reverse primer BR2 (Elbrecht & Leese, 2017b), it detected DNA from a broad range of aquatic and terrestrial arthropods. By using a mock sample and a malaise trap sample from a prior study evaluating primer performance (Elbrecht et al., 2019, PerPrint), we were able to confirm that NoSpi2 recovered about 94% of taxa. Taking design constraints into account, the primer should detect most prey taxa.

In vitro tests for single specimen were also very successful for prey, with most species being amplified. In some cases, results were less good but this is likely not linked to primer performance itself. For example, low matching scores for some Chironomids is very likely the result of the under-representation of inland chironomids in GenBank and BOLD. The

performance of NoSpi2/BR2 was also compared *in silico* to the barcoding primer pair LCO1490/HCO2198 (Folmer et al. 1994) which it outperformed for metabarcoding all of the groups tested except Embioptera, which are typically tropical insects.

As expected, NoSpi2/BR2 *in vitro* tests for spiders showed no amplification for members of the oval calamistrum clade. Only the amplification of *Piratula hygrophila* was contrary to our expectation, but Sanger sequencing revealed that the DNA originated from a very common Linyphiidae (*Neriene montana*), demonstrating the ability of our primer pair to detect intraguild predation among spiders. Similarly, the amplification of *Diaea dorsata* was unexpected but Sanger sequencing again demonstrated that the DNA originated from a prey species: *Sialis fuliginosa* (Megaloptera). This result further demonstrates the ability of NoSpi2/BR2 to amplify prey DNA without strong amplification of predator DNA. The detection of prey is surprising as only DNA from legs was used, but spider digestive caeca extend into femurs making the detection of prey possible, even when using only legs.

We expected members of the families Amaurobidae, Clubionidae, Oonopidae, Philodromidae and Pholcidae to be amplified by NoSpi2/BR2 given their low primer bias value (<320). This was not the case, however, only one sample was tested per family. More tests are required to verify the capacity of NoSpi2/BR2 to amplify these families. Dysderidae were neither amplified with NoSpi2/BR2 nor with LCO1490/HCO2198. The low bias values (306 for NoSpi2 and 6 for HCO) suggest the presence of PCR inhibitors or DNA nucleases in the sample.

In vitro tests with different mixes of spider and prey (Trichoptera) suggested no inhibitory effect or lower detection threshold through the inclusion of predator DNA. Indeed, prey was successfully amplified in all mixes and was identifiable to species level with matches greater than 99%, even at very low concentrations in the mix (2.8^{-6} ng/µl).

The field test confirmed that NoSpi2/BR2 can detect a large number of arthropod species (up to 12 orders and 117 species). Interestingly, no predator DNA was amplified, confirming that there is no amplification for Lycosidae species (genera tested: *Pardosa*, *Trochosa*, *Hygrolycosa*, *Xerolycosa* and *Pirata*) using NoSpi2.

Studies focusing on spider diet in natural habitats using metabarcoding are rare. Wirta, Weingartner, Hambäck, & Roslin (2015) found *Pardosa glacialis* to consume mainly Diptera and Lepidoptera in the High Arctic. Hambäck et al. (2016) also found Diptera as the main prey of *Pardosa prativaga* followed by Lepidoptera, Coleoptera and Heteroptera in the Baltic shoreline. In both cases the authors used primers designed to amplify specifically Diptera and Lepidoptera, potentially resulting in an underestimation of other orders. The main orders constituting Lycosid diets in our study were Diptera, Hemiptera, Hymenoptera and Lepidoptera. Interestingly, spiders, harvestmen and carabid beetles were also found in spider gut contents which confirms the potential of NoSpi2 for the investigation of intraguild predation, which has often been documented among spiders and for carabid beetles feeding on spiders (Lang, 2003; Davey et al., 2012; Sitvarin & Rypstra, 2014). Although coleopterans have been reported as spider prey (Hambäck et al., 2016), predation of carabid beetles by spiders has not been observed so far.

Snail (Stylommatophora) DNA was also detected in spider gut content, showing that despite gastropods generally being considered carabid and harvestmen prey, they are indeed a constituent of the diet of spiders (Nyffeler & Symondson, 2001).

Despite these very promising results, the fragments generated by NoSpi2/BR2 are relatively long (403 bp). This length ensures a high taxonomic resolution but it could also limit DNA detectability (Symondson, 2002) due to the higher likelihood of missing shorter degraded sequences (Symondson, 2002; Deagle, Eveson, & Jarman, 2006). Further studies, particularly feeding experiments, could help to determine the dynamics of DNA degradation enhancing our understanding of spider diet.

Conclusion

Molecular techniques that enable the simple and accurate analysis of spider diet are of great importance for studies investigating the impact of environmental change on ecological processes affecting the most abundant arthropod predators. The primer NoSpi2 developed and extensively validated for this study is ideal for gut content analysis of predator spider species of the oval calamistrum clade. It is enabling dietary and biological control studies as well as intra-guild predation studies. NoSpi2 amplifies a large number of prey but not the predator species and thereby renders dissection unnecessary because an entire homogenized spider can be used as sample.

Data accessibility

Trimmed sequences from the Sanger sequencing (*in vitro* evaluation) are available under the NOSPI project on BOLD.
Raw sequence data from the mock community metabarcoding and malaise sample are available under the SRA accession numbers SRX5975749 and SRX5975748. Data from the complete run with the 96 additional samples not analysed in this study is available under accession PRJNA546583.
Raw sequence data from the field test metabarcoding is available under SRA accession number PRJNA550981.
The code used for data analysis is available on GitLab under the project ID: 9419382 (https://gitlab.com/DenisLafage/nospi)

References

Agustí, N., Shayler, S. P., Harwood, J. D., Vaughan, I. P., Sunderland, K. D., & Symondson, W. O. C. (2003). Collembola as alternative prey sustaining spiders in arable ecosystems: Prey detection within predators using molecular markers. *Molecular Ecology*, *12*(12), 3467–3475. doi:10.1046/j.1365-294X.2003.02014.x

Birkhofer, K., Bylund, H., Dalin, P., Ferlian, O., Gagic, V., Hambäck, P. A., … Jonsson, M. (2017). Methods to identify the prey of invertebrate predators in terrestrial field studies. *Ecology and Evolution*, *7*(6), 1942–1953. doi:10.1002/ece3.2791

Boreau de Roincé, C., Lavigne, C., Mandrin, J.-F., Rollard, C., & Symondson, W. O. C. (2013). Early-season predation on aphids by winter-active spiders in apple orchards revealed by diagnostic PCR. *Bulletin of Entomological Research*, *103*, 148–154.

Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Elbrecht, V., Steinke, D., Ratnasingham, S., … Hebert, P. D. N. (2018). Revealing the complexities of metabarcoding with a diverse arthropod mock community. *BioRxiv*, 433607. doi:10.1101/433607

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. doi:10.1038/nmeth.3869

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1), 421. doi:10.1186/1471-2105-10-421

Chapman, E. G., Schmidt, J. M., Welch, K. D., & Harwood, J. D. (2013). Molecular evidence for dietary selectivity and pest suppression potential in an epigeal spider community in winter wheat. *Biological Control*, *65*(1), 72–86. doi:10.1016/j.biocontrol.2012.08.005

Davey, J. S., Vaughan, I. P., King, A. R., Bell, J. R., Bohan, D. A., Bruford, M. W., … Symondson, W. O. C. (2012). Intraguild predation in winter wheat: prey choice by a common epigeal carabid consuming spiders. *Journal of Applied Ecology*, 50(1) 271-279. doi:10.1111/1365-2664.12008

Deagle, B. E., Eveson, J. P., & Jarman, S. N. (2006). Quantification of damage in DNA recovered from highly degraded samples – a case study on DNA in faeces. *Frontiers in Zoology*, *3*(1), 11. doi:10.1186/1742-9994-3-11

Elbrecht, V., Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Hajibabaei, M., Wright, M., … Steinke, D. (2019). *Validation of COI metabarcoding primers for terrestrial arthropods* (preprint). PeerJ Preprints. doi:/10.7287/peerj.preprints.27801v1

Elbrecht, V., & Leese, F. (2017a). PrimerMiner: an R package for development and *in silico* validation of DNA metabarcoding primers. *Methods in Ecology and Evolution*, *8*(5), 622–626. doi:10.1111/2041-210X.12687

Elbrecht, V., & Leese, F. (2017b). Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science*, *5*(April), 1–11. doi:10.3389/fenvs.2017.00011

Elbrecht, V., & Steinke, D. (2018). Scaling up DNA metabarcoding for freshwater

macrozoobenthos monitoring. *Freshwater Biology*. http://doi.org/10.7287/peerj.preprints.3456v4

Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, *8*(10), 1265–1275. doi:10.1111/2041-210X.12789

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299. doi:10.1371/journal.pone.0013102

Greenstone, M. H., & Shufran, K. A. (2003). Spider predation: species specific identification of gut contents by polymerase chain reaction. *Journal of Arachnology*, *31*, 131–134.

Hambäck, P. A., Weingartner, E., Dalén, L., Wirta, H., & Roslin, T. (2016). Spatial subsidies in spider diets vary with shoreline structure: Complementary evidence from molecular diet analysis and stable isotopes. *Ecology and Evolution*, *6*(23), 8431–8439. doi:10.1002/ece3.2536

Holland, J. M., Bianchi, F. J. J. A., Entling, M. H., Moonen, A.-C., Smith, B. M., & Jeanneret, P. (2016). Structure, function and management of semi-natural habitats for conservation biological control: A review of European studies. *Pest Management Science*. doi:10.1002/ps.4318

Hosseini, R., Keller, M. A., Schmidt, O., & Li, D.-M. (2011). Design and assessment of primers to detect insect pests of Brassica crops in the gut contents of arthropod predators. *Munis Entomology and Zoology*, *6*(1), 477–492.

Hugerth, L. W., Wefer, H. A., Lundin, S., Jakobsson, H. E., Lindberg, M., Rodin, S., … Andersson, A. F. (2014). DegePrime, a program for degenerate primer design for broad-taxonomic-range PCR in microbial ecology studies. *Applied and Environmental Microbiology*, *80*(16), 5116–5123.

Kamenova, S., Mayer, R., Rubbmark, O. R., Coissac, E., Plantegenest, M., & Traugott, M. (2018). Comparing three types of dietary samples for prey DNA decay in an insect generalist predator. *Molecular Ecology Resources*, *18*(5), 966–973. doi:10.1111/1755-0998.12775

Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. doi:10.1093/nar/gkf436

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. doi:10.1093/bioinformatics/bts199

Kraus, J. M., Pomeranz, J. F., Todd, A. S., Walters, D. M., Schmidt, T. S., & Wanty, R. B. (2016). Aquatic pollution increases use of terrestrial prey subsidies by stream fish. *Journal of Applied Ecology*, *53*(1), 44–53. doi:10.1111/1365-2664.12543

Krehenwinkel, H., Rödder, N., & Tautz, D. (2015). Eco-genomic analysis of the poleward range expansion of the wasp spider Argiope bruennichi shows rapid adaptation and genomic admixture. *Global Change Biology*, *21*, 4320–4332.

Krehenwinkel, H., Kennedy, S.R., Adams, S.A., Stephenson, G.T., Roy, K., Gillespie, R.G.

(2019). Multiplex PCR targeting lineage-specific SNPs: A highly efficient and simple approach to block out predator sequences in molecular gut content analysis. *Methods in Ecology and Evolution*, 00, 1– 12. https://doi.org/10.1111

Kuusk, A.-K., Cassel-Lundhagen, A., Kvarnheden, 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*(6), 718–725. doi:10.1016/j.baae.2007.08.012

Lafage, D., Bergman, E., Eckstein, R. L., Österling, E. M., Sadler, J. P., & Piccolo, J. J. (2019). Local and landscape drivers of aquatic-to-terrestrial subsidies in riparian ecosystems: a worldwide meta-analysis. *Ecosphere*, *10*(4), e02697. doi:10.1002/ecs2.2697

Lang, A. (2003). Intraguild interference and biocontrol effects of generalist predators in a winter wheat field. *Oecologia*, *134*(1), 144–153. doi:10.1007/s00442-002-1091-5

Murphy, N. P., Framenau, V. W., Donnellan, S. C., Harvey, M. S., Park, Y.-C., & Austin, A. D. (2006). Phylogenetic reconstruction of the wolf spiders (Araneae: Lycosidae) using sequences from the 12S rRNA, 28S rRNA, and NADH1 genes: implications for classification, biogeography, and the evolution of web building behavior. *Molecular Phylogenetics and Evolution*, *38*(3), 583–602.

Nyffeler, M., & Symondson, W. O. C. (2001). Spiders and harvestmen as gastropod predators. *Ecological Entomology*, *26*(6), 617–628. doi:10.1046/j.1365-2311.2001.00365.x

Piñol, J., San Andrés, V., Clare, E. L., Mir, G., & Symondson, W. O. C. (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*(1), 18–26. doi:10.1111/1755-0998.12156

Piñol, J., Senar, M. A., & Symondson, W. O. C. (2018). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology*. doi:10.1111/mec.14776

Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology*, *21*(8), 1931–1950. doi:10.1111/j.1365-294X.2011.05403.x

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular Ecology Notes*, *7*(3), 355–364.

Riechert, S. E., & Lockley, T. (1984). Spiders as biological control agents. *Annual Review of Entomology*, *29*, 299–320.

Shehzad, W., Riaz, T., Nawaz, M. A., Miquel, C., Poillot, C., Shah, S. A., … Taberlet, P. (2012). Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (Prionailurus bengalensis) in Pakistan. Molecular Ecology, 21(8), 1951–1965.

Sint, D., Thurner, I., Kaufmann, R., & Traugott, M. (2015). Sparing spiders: faeces as a non-invasive source of DNA. *Frontiers in Zoology*, *12*(1), 3. doi:10.1186/s12983-015-0096-y

Sitvarin, M. I., & Rypstra, A. L. (2014). The importance of intraguild predation in predicting emergent multiple predator effects. *Ecology*, *95*(10), 2936–2945. doi:10.1890/13-2347.1

Symondson, W. O. C. (2002). Molecular identification of prey in predator diets. *Molecular Ecology*, 15.

Turnbull, A. L. (1973). Ecology of the True Spiders (Araneomorphae). *Annual Review of Entomology*, *18*(1), 305–348. doi:10.1146/annurev.en.18.010173.001513

Vamos, E., Elbrecht, V., & Leese, F. (2017). Short COI markers for freshwater macroinvertebrate metabarcoding. *Metabarcoding and Metagenomics*, *1*, e14625–e14625. doi:10.3897/mbmg.1.14625

Vestheim, H., & Jarman, S. N. (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*(1), 12. doi:10.1186/1742-9994-5-12

Waldner, T., Sint, D., Juen, A., & Traugott, M. (2013). The effect of predator identity on post-feeding prey DNA detection success in soil-dwelling macro-invertebrates. *Soil Biology and Biochemistry*, *63*, 116–123. doi:10.1016/j.soilbio.2013.03.030

Walters, D. M., Otter, R. R., Kraus, J. M., & Mills, M. A. (2018). Riparian spiders indicate the magnitude and sources of polychlorinated biphenyl contamination at a large contaminated sediment site. *Environmental Toxicology and Chemistry*, (37), 2467–2474.

Wheeler, W. C., Coddington, J. A., Crowley, L. M., Dimitrov, D., Goloboff, P. A., Griswold, C. E., … Zhang, J. (2016). The spider tree of life: Phylogeny of Araneae based on target-gene analyses from an extensive taxon sampling. *Cladistics*, 1–43. doi:10.1111/cla.12182

Wirta, H. K., Weingartner, E., Hambäck, P. A., & Roslin, T. (2015). Extensive niche overlap among the dominant arthropod predators of the High Arctic. *Basic and Applied Ecology*, *16*(1), 86–92. doi:10.1016/j.baae.2014.11.003

Xu, C. C. Y., Yen, I. J., Bowman, D., & Turner, C. R. (2015). Spider web DNA: a new spin on noninvasive genetics of predator and prey. *PLOS ONE*, *10*(11), e0142503. doi:10.1371/journal.pone.0142503

Acknowledgement

We also thank the RIVEM group (Karlstad University) for funding the *in vitro* and field tests. V.E. and D.S. as well as the metabarcoding experiments were supported by funding through the University of Guelph's Canada First Research Excellence Fund project Food from Thought.

Authors contribution

D.L. conceived the idea and led the . V.E. designed the primer. D.L., V.E., J.C., P.H. and AE contributed to the lab work, data processing and analyses. D.L., J.C. and V.E. wrote the draft manuscript. All authors contributed critically to the drafts and gave final approval for publication.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27854v1 | CC BY 4.0 Open Access | rec: 12 Jul 2019, publ: 12 Jul 2019