



High-throughput sequencing for community analysis: the promise of DNA barcoding to uncover diversity, relatedness, abundances and interactions in spider communities

Susan R. Kennedy¹ · Stefan Prost^{2,3} · Isaac Overcast^{4,5} · Andrew J. Rominger⁶ · Rosemary G. Gillespie⁷ · Henrik Krehenwinkel⁸

Received: 28 November 2019 / Accepted: 29 January 2020 / Published online: 10 February 2020
© The Author(s) 2020

Abstract

Large-scale studies on community ecology are highly desirable but often difficult to accomplish due to the considerable investment of time, labor and, money required to characterize richness, abundance, relatedness, and interactions. Nonetheless, such large-scale perspectives are necessary for understanding the composition, dynamics, and resilience of biological communities. Small invertebrates play a central role in ecosystems, occupying critical positions in the food web and performing a broad variety of ecological functions. However, it has been particularly difficult to adequately characterize communities of these animals because of their exceptionally high diversity and abundance. Spiders in particular fulfill key roles as both predator and prey in terrestrial food webs and are hence an important focus of ecological studies. In recent years, large-scale community analyses have benefitted tremendously from advances in DNA barcoding technology. High-throughput sequencing (HTS), particularly DNA metabarcoding, enables community-wide analyses of diversity and interactions at unprecedented scales and at a fraction of the cost that was previously possible. Here, we review the current state of the application of these technologies to the analysis of spider communities. We discuss amplicon-based DNA barcoding and metabarcoding for the analysis of community diversity and molecular gut content analysis for assessing predator-prey relationships. We also highlight applications of the third generation sequencing technology for long read and portable DNA barcoding. We then address the development of theoretical frameworks for community-level studies, and finally highlight critical gaps and future directions for DNA analysis of spider communities.

Keywords Metabarcoding · Portable sequencing · Third generation sequencing · Gut content analysis · Community assembly

This article is part of the Special Issue “Crossroads in Spider Research - evolutionary, ecological and economic significance”

Communicated by Matthias Pechmann

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00427-020-00652-x>) contains supplementary material, which is available to authorized users.

✉ Henrik Krehenwinkel
krehenwinkel@uni-trier.de

¹ Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology, Onna, Okinawa, Japan

² LOEWE-Centre for Translational Biodiversity Genomics, Senckenberg Museum, Frankfurt, Germany

³ National Zoological Garden, South African National Biodiversity Institute, Pretoria, South Africa

⁴ Graduate Center of the City University New York, New York, NY, USA

⁵ Ecole Normale Supérieure, Paris, France

⁶ Santa Fe Institute, Santa Fe, NM, USA

⁷ Environmental Sciences Policy and Management, University of California Berkeley, Berkeley, CA, USA

⁸ Department of Biogeography, Trier University, Trier, Germany

Introduction

Ecological communities are defined by both the organisms that persist within habitats, and the interactions that shape the assembly and diversity patterns of these organisms. Historically, characterizations of abundance, richness, relatedness, and interactions across entire communities have been limited to taxa that are readily identifiable or have been done on a sufficiently small scale that the laborious process of quantifying all community members and their interactions has been feasible (Gruner 2004; Krushelnycky et al. 2007). Predator-prey interactions have largely been based on observation (Binford 2001; Hiruki et al. 1999), detailed morphological examination of gut contents (Grey et al. 2002; Lafferty and Page 1997), or the analysis of stable isotope data (Wise et al. 2006). The recent advent of molecular metabarcoding approaches is just starting to revolutionize our ability to characterize biological communities (Cristescu 2014). In particular, data on small invertebrates, which make up the foundation of food webs and play central roles in ecosystem function, can be obtained on larger scales and in greater detail than ever before. We use spiders, some of the most phylogenetically and ecologically diverse predators on Earth (Foelix 2011), to illustrate the potential of such approaches for understanding community assembly.

In the last two decades, DNA barcoding, the sequencing of short species-specific amplicons, has considerably simplified community analyses (Hebert et al. 2003). DNA barcodes can provide information on genetic variation within and between species, rapidly assign taxonomic status across divergent lineages (Hebert and Gregory 2005), and identify the prey composition of predators' gut contents (Agustí et al. 2003; Greenstone and Shufan 2003). However, traditional Sanger sequencing-based DNA barcoding protocols can be prohibitively expensive and laborious when large community samples have to be processed. The emergence of high-throughput sequencing technologies (HTS) has been a significant step forward in recent years, greatly reducing both the cost and the labor required for biodiversity studies (Bohmann et al. 2014; Taberlet et al. 2012). These technologies enable simultaneous processing of DNA barcodes for thousands of specimens (Shokralla et al. 2015; Srivathsan et al. 2019; Meier et al. 2016) with considerably improved phylogenetic resolution (Krehenwinkel et al. 2019a). Metabarcoding makes it possible to characterize the species composition of whole communities (Cristescu 2014; Yu et al. 2012) and identify the makeup of the predators' diets in an unprecedented detail (Piñol et al. 2014; Verschut et al. 2019). Recent developments even enable mobile DNA barcoding under remote field conditions (Menegon et al. 2017; Pomerantz et al. 2018).

Here, we provide an overview of available HTS-based methods, focusing specifically on the use of genetic and genomic data for characterizing community structure and

function in spiders. Within this context, we first discuss DNA barcoding for taxonomic and phylogenetic assignments, metabarcoding for community analysis, recent developments in long-read sequencing technology, and portable field barcoding solutions. We then review the application of DNA barcoding for gut content analysis to assess predator-prey associations. We additionally discuss the development of theoretical models to apply to the DNA-based community analyses. Finally, we address promising avenues of future research.

DNA barcoding and metabarcoding for community analysis

DNA barcoding in spiders: An overview

As major predators of invertebrates, spiders are a central element of terrestrial food webs and perform key roles in community function and assembly (Nyffeler and Birkhofer 2017). They provide important ecosystem services, such as pest control (Riechert and Lockley 1984; Thomson and Hoffmann 2010) and at the same time make up much of the diet of higher order predators such as birds (Nyffeler et al. 2018). Most habitats harbor diverse communities of spiders with complex ecological interrelationships (Kennedy et al. 2019; Raso et al. 2014). Consequently, the diversity of spiders and their manifold interactions with other species must be understood in order to characterize community assembly in terrestrial ecosystems. Spider communities are usually composed of ecologically and morphologically distinct taxa, stemming from deeply divergent evolutionary lineages (Cardoso et al. 2011). The identification of different groups often requires specialized taxonomic expertise, a skillset which is rapidly disappearing (Agnarsson and Kuntner 2007).

The task of characterizing the entire spider communities has been greatly simplified by DNA barcoding (Barrett and Hebert 2005; Čandek and Kuntner 2015; Crespo et al. 2018; Fig. 1). DNA barcoding of spiders is usually based on the 650-bp "barcode region" of the mitochondrial COI gene, which provides good taxonomic resolution in this group (Čandek and Kuntner 2015). As part of the mitochondrial genome, COI is maternally inherited and not affected by recombination. Mitochondria occur in most tissues in high copy numbers and are thus easily accessible for PCR amplification, even in degraded samples. The gene also evolves relatively quickly, making it suitable to distinguish even recently diverged species and recover intraspecific variation (Hajibabaei et al. 2007).

DNA barcodes garnered much enthusiasm after their initial establishment, with some authors even suggesting that traditional taxonomic methods be entirely replaced with a DNA sequence divergence-based taxonomy (Meierotto et al. 2019; Tautz et al. 2003). However, a divergent barcode sequence recovered from an unknown specimen is not enough to

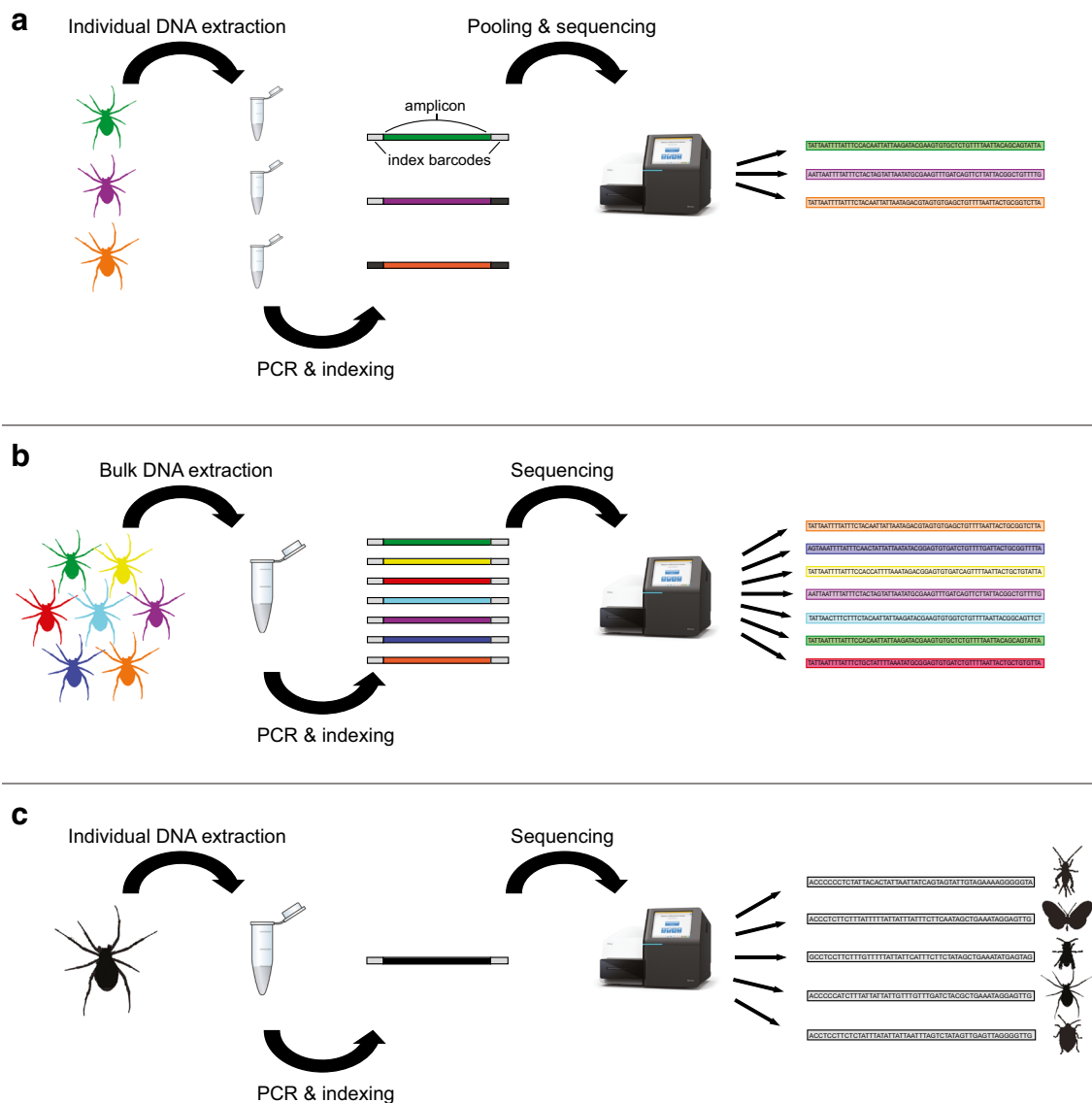


Fig. 1 Summary of applications of Illumina amplicon sequencing for DNA barcoding and metabarcoding of spiders. A) In individual DNA barcoding, DNA from each specimen is extracted, then the desired fragment is amplified in a PCR and tagged with a unique combination of index barcodes before all samples are pooled and sequenced. B) In bulk metabarcoding, DNA extraction is performed on pools of multiple

specimens. This greatly reduces the number of PCRs and index combinations needed per specimen. C) For molecular gut content analysis, DNA is extracted from individual predator specimens, and PCR primers are chosen to amplify prey taxa while minimizing amplification of the predator itself

indicate the species status (Moritz and Cicero 2004; Obertegger et al. 2018). Instead, DNA barcodes can serve as a valuable complement to traditional taxonomy by facilitating the identification of divergent lineages, including cryptic species (Wang et al. 2018). This has proven to be useful in some spiders with ambiguous morphological differentiation, which show sufficiently deep genetic divergence to be considered separate species (Leavitt et al. 2015; Starrett and Hedin 2007; Crespo et al. 2018).

Multiple primers with various levels of taxonomic specificity are available for DNA barcoding of spiders (Blagoev et al. 2016; Krehenwinkel et al. 2018; Supplementary Table 1). The

resulting barcode sequences can be compared to reference databases to assign specimens to species (Barrett and Hebert 2005; Robinson et al. 2009). Alternatively, the presence of a so-called DNA barcode gap between interspecific and intra-specific genetic divergences in the COI gene (Hebert et al. 2004a, b) can be used to identify putative species from the DNA barcode data. There is no universal rule for genetic distances to warrant “species” status; instead, the barcode gap must be evaluated on a lineage-specific basis. This approach was demonstrated to work well in orb-weaver and wolf spiders (Čandek and Kuntner 2015). Automated approaches aiding in the discovery of barcode gaps and resulting species

are available (Puillandre et al. 2012). Yet another approach is the grouping of barcodes from a community into clusters of similarity, the so-called operational taxonomic units (OTUs) (Edgar 2013). These clusters, usually based on a maximum sequence divergence of 3%, are then treated as biological entities. Even though OTU clusters do not necessarily correspond to actual species, this approach can be very useful when reference libraries are incomplete (Dopheide et al. 2019) and large numbers of sequences need to be processed. Clustering approaches can also be phylogenetically informed, resulting in more accurate approximations of real species (Fujita et al. 2012; Zhang et al. 2013).

Problems with DNA barcoding, and approaches to mitigate problems

A significant obstacle to DNA barcoding is the incompleteness of the barcode reference libraries. Because many species have not yet been added to these libraries, often specimens can only be identified to a relatively coarse taxonomic level such as order or family. Even though substantial contributions to reference databases have been made in recent years (Astrin et al. 2016; Blagoev et al. 2016), given the sheer taxonomic diversity of spiders, a large proportion of species is still not represented. There are two major bottlenecks in the generation of barcode reference libraries. First, the identification of species is time-consuming and requires taxonomic expertise, which may not be available (Agnarsson and Kuntner 2007). Misidentifications or sequencing of contaminant DNA can then lead to erroneously assigned barcode sequences in the database. Second, many species are rare or difficult to collect, and thus represented by little more than type material. Museum collections are therefore an indispensable resource for DNA barcoding. In spiders, this is particularly feasible because the standard storage medium for spiders – ethanol – is an effective DNA preservative. Only slight modifications of DNA extraction and PCR protocols are needed to recover the reduced and fragmented DNA from historical specimens (Krehenwinkel and Pekár 2015; Miller et al. 2013). Several primer combinations are available to target short, so-called mini-barcodes, which are suitable for amplification of older specimens (Supplementary Table 1). In spiders, barcode analysis of historical specimens has provided valuable insights into the taxonomic assignment of species (Cotoras et al. 2017) and historical changes in genetic variation (Krehenwinkel and Tautz 2013).

Another challenge for DNA barcoding is that short-mitochondrial amplicons such as COI can yield biased biodiversity assessments when used in isolation (Krehenwinkel et al. 2018). Mitochondrial divergence patterns do not necessarily parallel species divergence but are influenced by numerous different factors. For example, male-biased gene flow can

lead to highly divergent mitochondrial genomes in the absence of nuclear differentiation (Krehenwinkel et al. 2016). Conversely, introgression can result in complete homogenization of the mitochondrial gene pools, despite divergent nuclear genomes (Irwin et al. 2009). Infections with endosymbiotic bacteria can mimic various demographic scenarios of over- and under-differentiation of mitochondrial genomes compared to the nuclear background (Hurst and Jiggins 2005). Moreover, nuclear mitochondrial pseudogenes (NUMTs) can be recovered as barcode sequences, leading to incorrect taxonomic assignments and biased phylogenetic inferences (Bensasson et al. 2001). To avoid these pitfalls, it is often recommended to use multiple loci for DNA barcoding (Dupuis et al. 2012). Information from the unlinked loci in the nuclear genome is particularly important and can aid with DNA barcode-assisted taxonomic discoveries, e.g., for testing hypotheses on cryptic species (Satler et al. 2013). Although many popular nuclear markers evolve much more slowly than COI, they still show comparable patterns of genetic divergence when intraspecific and interspecific divergence rates are compared (Supplementary Fig. 1). Multilocus data can also increase the phylogenetic resolution of DNA barcoding, which is very limited when analyses are based on a single mitochondrial amplicon (Krehenwinkel et al. 2018).

High throughput sequencing-based DNA barcoding

DNA barcoding is traditionally based on Sanger sequencing, requiring separate sequencing reactions for every sample. With a total cost of \$ 5–10 per sequence, this method can be prohibitively expensive for community-level studies. A cost-efficient alternative is high-throughput amplicon sequencing (Kozich et al. 2013). Illumina technology, for example the MiSeq, with its maximum read length of 2×300 bp, is highly suitable for DNA barcode generation (Shokralla et al. 2015). Due to limitations in read length, HTS-based DNA barcoding usually relies on shorter amplicons than the complete 650 bp barcode (Leray et al. 2013). Alternatively, the complete barcode can be recovered by sequencing multiple overlapping amplicons.

Illumina amplicon sequencing is distinguished by a very simple library preparation process. Most commonly, a two-step PCR is used, in which the target sequence is amplified in the first round of PCR (Fig. 2). Dual indexes for unique sample tagging and the necessary adapters for sequencing are then incorporated in the second round of PCR (Lange et al. 2014). This approach accommodates thousands of samples in a single sequencing run. Multiplex PCRs targeting multiple unlinked loci can additionally reduce the necessary number of PCRs (Krehenwinkel et al. 2018; Macías-Hernández et al. 2018), and inline barcodes attached to the first-round PCR primers allow for a further increase in sample number (Sternes et al. 2017). Alternatively, fusion primers including sample tags and sequencing adapters can be used. This allows

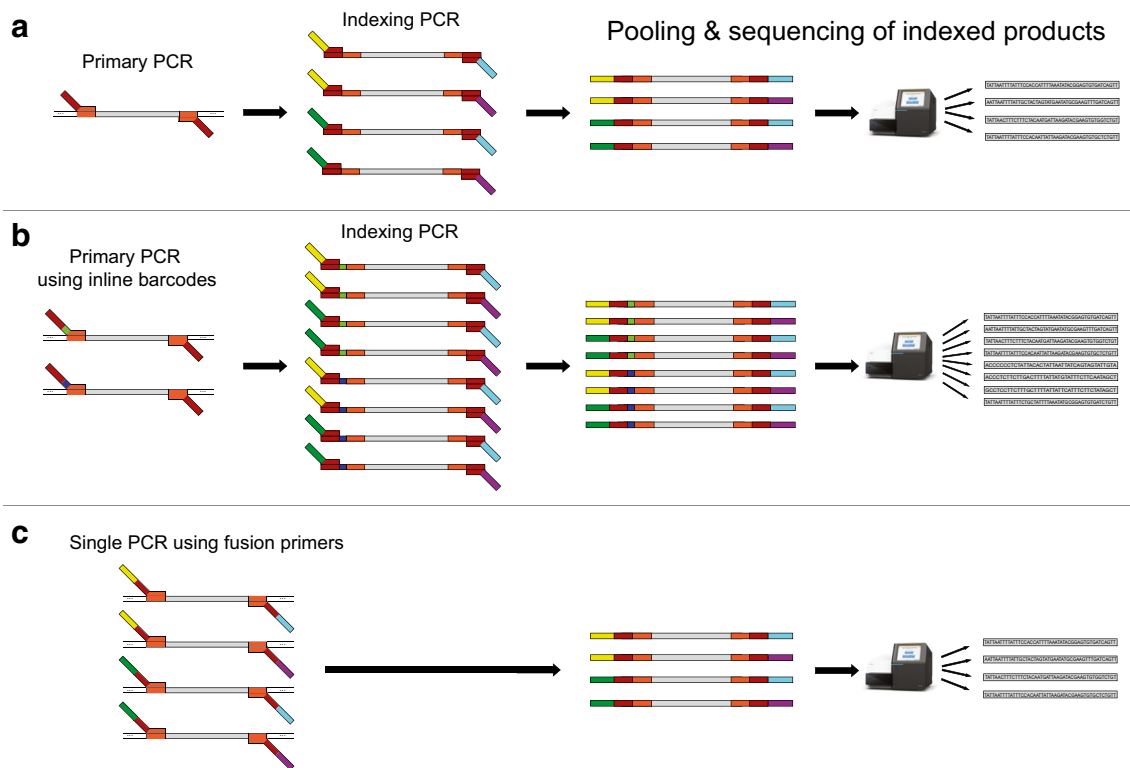


Fig. 2 Dual indexing strategies for Illumina sequencing. A) Library preparation can be accomplished in two separate PCRs. In the first PCR, the DNA barcode specific primers contain added tails (in brown). Second PCR primers then bind to those tails and incorporate unique barcode identifiers as well as sequencing adapters to each sample. A unique barcode combination is used for each sample. B) Throughput can be increased with the use of inline barcodes (light green and indigo)

attached to the 5'-end of the first-round PCR primer. They multiply the number of unique barcode combinations available. Here, we show inline barcodes only on the forward primer, but they can also be incorporated into the reverse primer. C) Fusion primers can be used so that only one round of PCR is necessary. The desired fragment is amplified and indexed simultaneously

library preparation to be accomplished in a single PCR (Kozich et al. 2013; Fadrosch et al. 2014) but limits the flexibility to target multiple amplicons. Further reductions in processing cost can be achieved by limiting the number of DNA extractions. This can be done by pooling specimens of divergent lineages, then performing bulk extractions (de Kerdrel et al. 2020). PCR and library preparation are then performed on the pooled extract, and the resultant sequences are assigned back to their specimens using a reference database. One other option is to omit DNA extraction entirely and instead use direct PCR (Wong et al. 2014). Here, specimens are dropped directly into the PCR buffer, and the traces of DNA they release are sufficient for barcode amplification. This method has been tested and established in different insect groups (Thongjued et al. 2019; Yeo et al. 2018) but has yet to be optimized for spiders.

By limiting the number of DNA extractions, using multiplex PCRs, and using multiple levels of sample indexing, barcodes can now be generated at a cost of \$ 0.2–1 each and with a considerable reduction of workload (de Kerdrel et al. 2020; Meier et al. 2016; Srivathsan et al. 2019). This enables researchers to generate barcode sequences for thousands of specimens, thereby allowing estimates of both abundance

and taxonomic richness within a community (see below). Recent reductions in cost have even led to the suggestion of a reverse DNA barcoding workflow, in which all specimens in a collection are barcoded and only divergent lineages are selected for further morphological analysis (Wang et al. 2018).

Using multiplex PCR approaches, sequences for multiple independent loci can be generated in parallel, greatly improving the phylogenetic resolution of the generated data. Knowing the evolutionary relationships among taxa in a community is critical for understanding the processes underlying community assembly (Barker 2002). Currently, phylogenetic analyses often rely on information from hundreds or thousands of loci, for example, inferred from whole transcriptome sequencing (Foley et al. 2019) or the targeted enrichment of ultra-conserved elements (Kulkarni et al. 2020). While such data offer unprecedented phylogenetic resolution, their generation is expensive and laborious. These methods are thus not feasible for phylogenetic analyses at the community level, where information for thousands of specimens has to be generated in parallel. This makes multiplexed amplicon sequencing an attractive alternative to phylogenomic approaches for community phylogenetic analyses.

Community metabarcoding

Reductions in cost and processing effort have made DNA barcoding suitable even for the analysis of large community samples. However, processing all specimens individually still amounts to a considerable workload and cost. Metabarcoding offers a simple alternative to a single-specimen DNA barcoding and is therefore quickly increasing in popularity (Gibson et al. 2014; Yu et al. 2012). In metabarcoding, bulk samples are extracted and DNA barcode sequences generated for the pooled community. Based on the sequence similarity, the recovered barcodes are then clustered into OTUs. Community diversity is estimated based on the number of recovered OTU sequences. In order to achieve a comprehensive taxon recovery in metabarcoding experiments, the use of more than one amplicon is advisable (Krehenwinkel et al. 2018; Zhang et al. 2018). Recent developments in clustering algorithms (Edgar 2018) also allow the inference of haplotypic information from bulk metabarcoding data. This way, even the intraspecific genetic variation can be estimated within whole biological communities (Elbrecht et al. 2018).

Due to its speed, accuracy, and cost efficiency, metabarcoding is now often the method of choice for arthropod community analysis. A major drawback of this approach, however, is that it only yields a list of OTU sequences, which cannot be linked back to individual specimens because the DNA is bulk-extracted from mixed samples. As sequences cannot be assigned back to specimens, individual sequences from multilocus barcoding cannot be linked together, limiting the phylogenetic resolution of this approach unless the specimens can be linked to an extensive reference library (de Kerdrel et al. 2020). Metabarcoding can also lead to inflated diversity estimates, as spurious sequences coamplify with the targeted specimen's DNA barcodes. These include NUMTs, non-target species such as parasitic fungi or nematodes associated with the specimen, and chimeras resulting from linking of incomplete PCR products of different taxa (Elbrecht et al. 2017). Chimeras can be removed efficiently with appropriate software solutions (Edgar 2013) and non-target taxa by comparing the resulting data against a reference database. However, the removal of NUMTs is more challenging and has not been fully resolved, especially when NUMTs retain an intact reading frame.

Another issue with metabarcoding is that many current protocols are performed destructively, i.e., specimens are crushed in order to maximize the amount of DNA released for extraction. This inevitably leads to the loss of morphological information. This issue, however, can be circumvented by subsampling tissue from specimens before extraction. In spiders, it is common to extract DNA from one or more legs while leaving the rest of the specimen intact (Gillespie et al. 2018; Krehenwinkel et al. 2018). In addition, several nondestructive protocols have been developed to isolate DNA from

specimens without compromising their morphological integrity, e.g., via a brief soak in lysis buffer (Andersen and Mills 2012; Porco et al. 2010). DNA may even be extracted directly from the collection medium (e.g., ethanol), because specimens leave trace amounts of DNA in the medium (Hajibabaei et al. 2012; Martins et al. 2019). However, the DNA recovered from nondestructive extraction of community samples may be biased toward those taxonomic groups that release DNA more readily than others, e.g., soft-bodied animals (Carew et al. 2018; Marquina et al. 2019).

Quantitative metabarcoding and PCR-free metagenomics

While metabarcoding has been able to provide a highly accurate overview of a community's species composition, it has not been possible to obtain accurate measures of abundance using this approach. This is because varying PCR efficiencies between different taxa inevitably lead to biased recovery of species abundances, sometimes by several orders of magnitude (Elbrecht and Leese 2015). Besides simple primer-template mismatches (Piñol et al. 2015), the GC content of the template and even the polymerase type can bias taxon recovery (Nichols et al. 2018). The effects of these biases are condition dependent and difficult to quantify. Even the effect of primer-template mismatches does not simply accumulate with mismatch number. Instead, position and type of mismatch can have widely different effects on amplification efficiency (Kwok et al. 1990). Nevertheless, the accurate quantification of relative abundances of taxa is critical for many biodiversity analyses, and much effort is therefore being made to optimize methods for quantitative metabarcoding (Krehenwinkel et al. 2017a; Piñol et al. 2019; Saitoh et al. 2016).

To overcome the amplification biases of metabarcoding, PCR-free approaches have been suggested (Jones et al. 2015). The simplest is shotgun sequencing of bulk samples, after which the generated reads are processed and compared against a reference database, an approach called metagenomics. However, this may not work well in spiders because very few spider genomes (nine, representing seven families) are currently available (Supplementary Table 2), making it difficult to identify the majority of sequences. A refinement of this method is "genome skimming," in which mitochondrial sequences are filtered from the recovered reads after shotgun sequencing (Papadopoulou et al. 2015). The filtered reads can then be assembled into longer contigs sometimes even spanning the whole mitochondrial genome, allowing community analysis with considerable phylogenetic support (Crampton-Platt et al. 2016). However, although mitochondria are abundant in cells, mitochondrial DNA sequences usually do not exceed 1% of the read population of genomic libraries (Zhou et al. 2013). Genome skimming thus requires a very high sequencing coverage. Capture assays are another option: DNA barcode probes

are used to capture barcode sequences from a community sample, allowing sequencing without PCR amplification (Shokralla et al. 2016). However, hybridization bias due to probe-target dissimilarities may also result in skewed abundance estimates; furthermore, a capture approach adds considerable cost and workload.

PCR-based metabarcoding is therefore still the most cost-effective and simple method for community analysis of spiders. Although PCR amplification bias can theoretically lead to skewed taxon abundances, this bias can be considerably mitigated using optimized protocols. Amplification with degenerate primers, or with primer binding in conserved DNA stretches, greatly improves taxon recovery while decreasing amplification bias (Krehenwinkel et al. 2017a). Furthermore, the response of individual taxa during a PCR is predictable, i.e., the relative abundance of a taxon in a community is linearly correlated with the recovered read abundance (Fig. 3). Only the slope of this correlation differs among taxa. If the slopes are known, then correction factors can be applied to estimate the relative abundance of taxa in a community (Thomas et al. 2016). The downside is that correction factors must be developed individually for different taxa.

Environmental DNA metabarcoding of spiders

A popular application of metabarcoding is the analysis of environmental DNA (eDNA). Every organism leaves traces of DNA in the environment, for example, from feces, skin fragments, or saliva. These traces can be enriched, amplified, and sequenced, allowing characterization of whole communities without needing to collect the organisms. Much work on eDNA has focused on aquatic ecosystems, using DNA extracts from filtered water (Valentini et al. 2016). However, terrestrial organisms can also be detected using eDNA, for example from

soil DNA extractions. Arthropods were recently shown to leave eDNA traces on wildflowers (Thomsen and Sigsgaard 2019). Thus, by washing eDNA off of plants, it may be possible to reconstruct associated arthropod communities.

Third generation sequencing-based barcoding and metabarcoding

DNA barcoding and metabarcoding applications are currently limited by the relatively short read length of second generation HTS applications, which cannot recover the whole 650-bp COI barcode region as a single sequence (Piper et al. 2019). A solution to this limitation is provided by the third generation sequencing technologies. Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) offer sequencing platforms that achieve read lengths superior to any previous sequencing technology, with reads of close to 2 megabases for ONT's MinION platform (Payne et al. 2018). Both technologies are well suited for amplicon sequencing, and dual indexes can be easily incorporated during PCR, allowing processing of thousands of DNA barcodes in a single sequencing run. The PacBio Sequel (Hebert et al. 2018) and ONT's MinION (Srivathsan et al. 2019) were recently suggested as cost-efficient alternatives to Sanger sequencing for the generation of DNA barcodes. ONT and PacBio platforms have also recently been used to sequence near complete nuclear ribosomal DNA clusters (Krehenwinkel et al. 2019a; Tedersoo et al. 2018). The advantage of rDNA barcoding is that conserved gene regions of the rDNA cluster can be used to design universal primers (anchored in the highly conserved 18S and 28S rDNA), which can resolve very old divergences (Hillis and Dixon 1991), while at the same time fast-evolving internal transcribed spacers (ITS) can be used to resolve relationships of closely related taxa (Schoch et al. 2012). Our work on

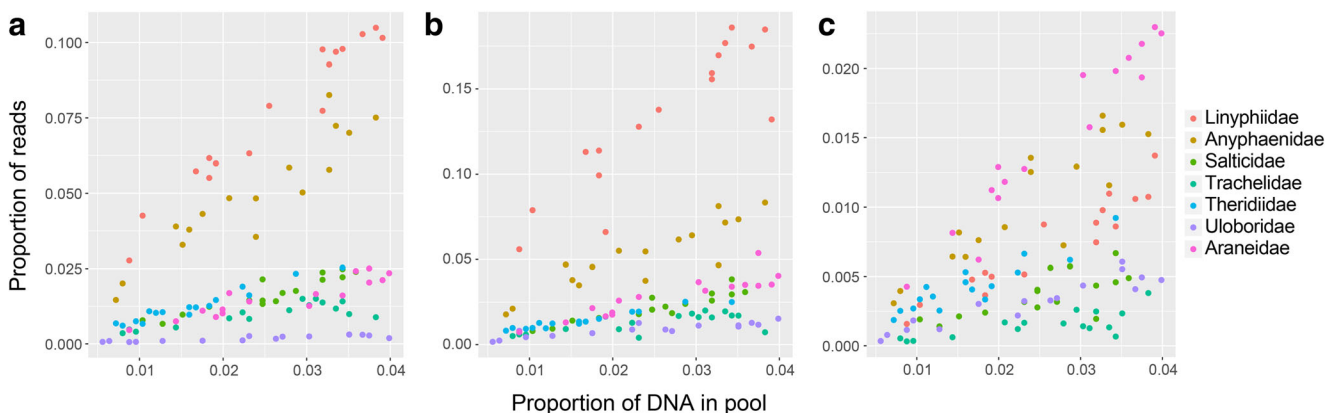


Fig. 3 Association of the relative abundance of spider species from seven different families in mock communities of 46 different arthropod species, with the relative read count recovered for the species after sequencing. The plots show the association for A) nuclear 18S rDNA, B) nuclear 28S rDNA and C) mitochondrial COI. The abundance of the species in

the respective communities is generally well correlated to the recovered read abundances. However, depending on marker and species, the slope of the association is very variable, such that accurate abundance estimates from read data would require careful calculation. Based on data from Krehenwinkel et al. 2017a

spiders suggests that such long rDNA amplicons are a well-suited complement to COI-based DNA barcoding (Supplementary Fig. 2; Krehenwinkel et al. 2019a). Long rDNA amplicons also offer very good phylogenetic support and thus may constitute a cost-effective alternative to the multiplexed Illumina amplicon sequencing for the community-wide phylogenetic analysis.

The third generation sequencing platforms offer an unprecedented read length, but their major downside is a high raw read error rate. At about 5–30% (Tedesoo et al. 2018; Wick et al. 2018), ONT and PacBio sequencers' raw read error rate is much higher than that of Illumina (error rate: 0.1–1%; Manley et al. 2016) and Sanger sequencing (error rate: 0.001–1%; Noguchi et al. 2006). However, highly accurate consensus sequences can be generated from ONT and PacBio data even at low coverage (Krehenwinkel et al. 2019a; Pomerantz et al. 2018). Recent advances show promise for further minimizing error. PacBio HiFi sequencing mode produces highly accurate reads using their circular consensus sequencing (CCS) technology, reducing raw read error to < 1% (Wenger et al. 2019). Similarly, rolling circle amplification can be used in metabarcoding applications to mitigate error rates of the nanopore-based sequencing platforms (Calus et al. 2018).

The possibility of long-read metabarcoding was also recently explored (Callahan et al. 2019; Krehenwinkel et al. 2019a; Tedesoo and Anslan 2019). Barcode sequences of several thousand base pairs for a whole community would greatly improve the phylogenetic resolution of metabarcoding and allow community-level phylogenetic analysis. However, the high raw read error rate poses a significant obstacle for accurate community characterization, as it is hard to distinguish whether a rare sequence variant belongs to a separate species in the community or is simply caused by sequencing error. Nonetheless, advances such as CCS and rolling circle amplification may soon solve this problem. One other issue is that community compositions of rDNA metabarcoding studies can be highly skewed, likely due to favorable PCR amplification of shorter rDNA fragments (Krehenwinkel et al. 2019a). Although the length of the nuclear rDNA region is relatively stable within spiders, biases can occur when additional taxa are included (Krehenwinkel et al. 2019a).

Mobile DNA barcoding by third generation sequencing

A particular strength of ONT's MinION is its portability. With the size of a USB stick, the device can be run outside of conventional laboratories (e.g., Menegon et al. 2017; Pomerantz et al. 2018). Using a mobile laboratory of miniaturized equipment, all steps from DNA extraction to PCR, library preparation and sequencing can be performed in the field using the MinION (Pomerantz et al. 2018; reviewed in

Krehenwinkel et al. 2019b). While field-based DNA barcoding is an exciting perspective, it is unlikely to become the method of choice for community barcoding. Researchers usually have access to molecular laboratories that allow for more standardized and higher throughput sample processing than field-based assays. Yet, a minimalistic and mobile DNA barcoding system can be of great advantage when field sites are remote or hard to access, or when time is of the essence for swift generation of biodiversity information (reviewed in Krehenwinkel et al. 2019b). Examples include monitoring of disease outbreaks (Quick et al. 2016; Walter et al. 2017) or documenting the immediate effects of ecological disasters such as forest fires or pipeline spillages. Another advantage of mobile barcoding is that it allows for in situ species monitoring without having to remove organisms from their habitat or send samples internationally (Pomerantz et al. 2018). This is especially relevant for endangered species. In the case of spiders, non-lethal sampling protocols could be applied for site-based monitoring without directly affecting the population (Longhorn et al. 2007; Petersen et al. 2007).

Trophic niche analysis by DNA barcoding

DNA barcoding for gut content analysis: toward community-level food webs

Surprisingly little is known about the dietary ecology of spiders. While most spiders have long been understood as generalist predators, recent work also highlights many examples of dietary specialists, like termite feeders (Petráková et al. 2015), araneophages (Benavides et al. 2017; Wood et al. 2012) and even herbivorous species (Meehan et al. 2009; Nyffeler et al. 2016). The compilation of dietary information for spiders from observational data is very time-consuming and often inaccurate. The sheer diversity of spiders additionally complicates the task. More detailed dietary information is available chiefly for species being considered as potential biocontrol agents (Schmidt et al. 2014; Roubinet et al. 2017).

Molecular gut content analysis has simplified the task of characterizing spider prey communities and associated strengths of predator-prey interactions, thereby allowing a more accurate reconstruction of the often-cryptic arthropod food web (Sint et al. 2019). In the simplest case, spiders can be tested for consumption of specific prey taxa by subjecting a spider's gut content to PCR assays using prey-specific primers (Schmidt et al. 2014; Whitney et al. 2018). This can be useful for determining whether a spider could serve as a biocontrol agent against a particular pest species. However, the limitation of this method is that the expected prey taxa must be known a priori, and specific PCR assays must be developed for every prey taxon. Multiplex PCR approaches (King et al. 2011;

Roubinet et al. 2017) can broaden taxonomic coverage of prey detection but are still limited in their taxonomic breadth.

A more complete prey spectrum can be recovered via metabarcoding of the gut contents (Deagle et al. 2009; Leray et al. 2013). In principle, the same DNA barcoding approaches used for community characterization can be applied to gut content analysis, i.e., by treating the prey DNA inside the gut as a “community.” However, there are some additional considerations for tailoring these approaches to the specific conditions affecting DNA extracted from the gut. PCR inhibitors may coprecipitate with the DNA, requiring additional purification of DNA from gut extractions. Also, prey DNA from the predator’s guts is often degraded and present at much lower concentrations than the DNA of a single specimen in a bulk community extract. Hence, short PCR amplicons are usually targeted to achieve a complete prey spectrum (Zeale et al. 2011; Kamenova et al. 2018). Gut content metabarcoding has become increasingly popular and has recently provided numerous novel insights into the trophic ecology of spiders. Examples include trophic niche differentiation within an adaptive radiation (Fig. 4; Kennedy et al. 2019), the effect of grazers on prey communities (Schmidt et al. 2018), ontogenetic shifts in diet (Verschut et al. 2019), and a stable diet despite differences in available prey communities along an elevational gradient (Eitzinger et al. 2019). Improved resolution is often achieved by combining HTS-based gut content screening with stable isotope analysis (Hambäck et al. 2016; Kennedy et al. 2019).

Enrichment of prey DNA

Dissecting the gut of a spider for prey recovery is time-consuming and laborious. A highly simplified approach was thus suggested by Piñol et al. (2014). In their study, the authors used DNA extractions from whole spider bodies and amplified DNA barcodes using universal primers. Predator barcodes, which co-amplified during PCR, were removed from the analysis, and the prey spectrum is reconstructed from the remaining sequences. As universal primers are needed in order to recover a full prey spectrum, a serious problem of this approach is that both predator and prey will be amplified. Consequently, the overabundant predator DNA can completely outcompete the prey DNA during PCR. Recent work suggests the use of predator-specific blocking primers, but due to the close relatedness of prey and predator, this approach has had only limited success in spiders (Michálek et al. 2017; Toju and Baba 2018). Another option is to enrich prey DNA from spider extracts. Prey DNA is quickly degraded in the spider’s digestive tract. By separating intact high molecular weight DNA from degraded DNA fragments, prey DNA can be significantly enriched (Fig. 5; Krehenwinkel et al. 2017b). However, this method will only work well if the predator DNA is not degraded; therefore, it is not suitable for old or poorly preserved specimens. An additional enrichment of prey DNA can be achieved by extracting DNA from the spider’s opisthosoma only, which contains the majority of the animal’s digestive tract (Krehenwinkel et al. 2017b; Macías-Hernández et al. 2018).

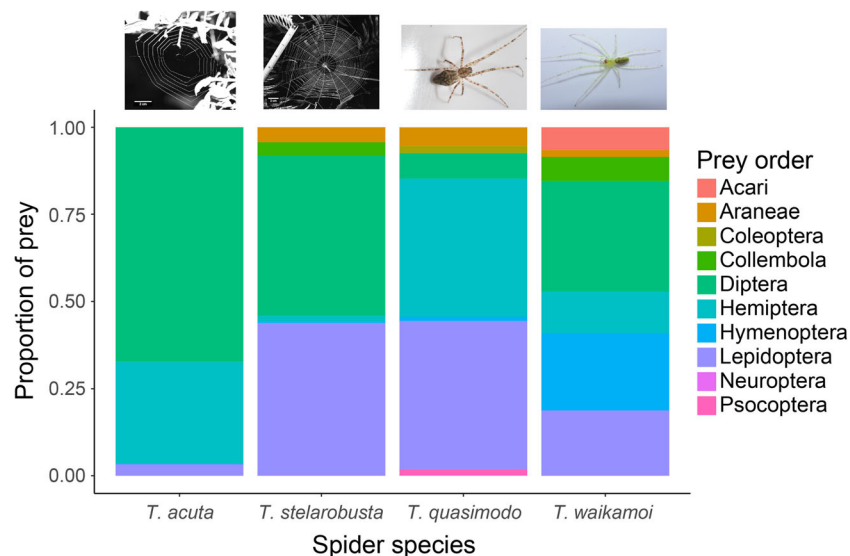


Fig. 4 Order-level prey compositions for four sympatric *Tetragnatha* species from the Hawaiian island of Maui, as recovered by molecular gut content analysis based on a short COI amplicon. The different lifestyles of the web builders (*T. acuta* and *T. stellarobusta*) and free-hunting species (*T. quasimodo* and *T. waikamoi*) are reflected in divergent prey spectra. However, prey community differences also become

apparent within web builders and free hunters. This effect may be due to interspecific differences in microhabitat and prey capture strategy. The green *T. waikamoi* resides and hunts on green leaves, while the brown *T. quasimodo* occurs on tree bark or dead leaves. The two web builders are distinguished by different web mesh widths, selectively catching different arthropod groups. Based on data from Kennedy et al. 2019

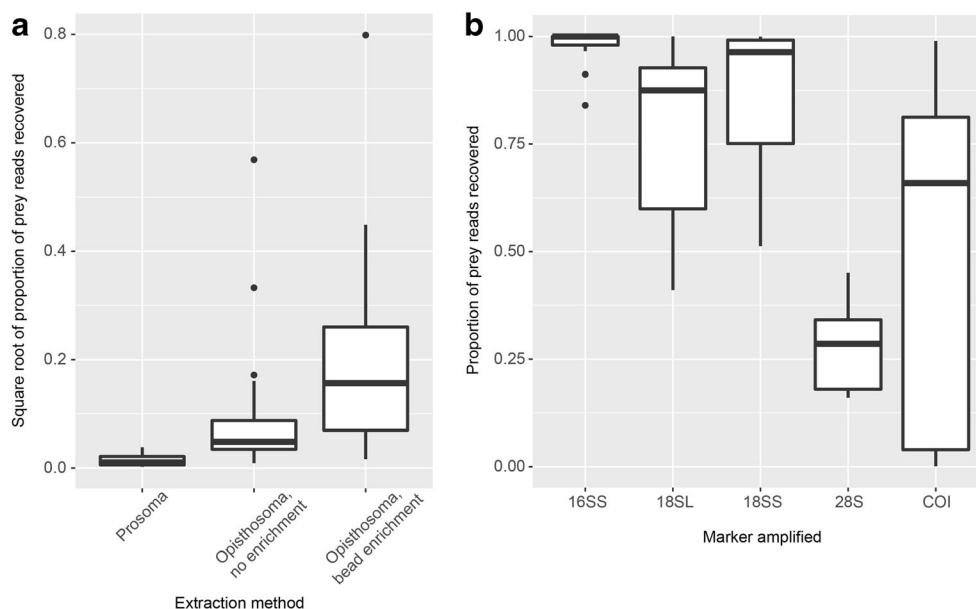


Fig. 5 Enrichment of prey DNA from extractions of spiders. A) The recovered relative amount of prey DNA of *Hololena adnexa* in relation to spider DNA increases significantly when DNA extractions are performed from the opisthosoma rather than the prosoma. The yield can be further increased by using a bead protocol to enrich the low molecular weight DNA from the DNA extract. This works because prey DNA

rapidly degrades in the spider's digestive tract. B) Enrichment of prey DNA in 12 spider species from seven families by using different lineage-specific primers and in comparison to a commonly used COI primer pair. Based on spider-specific 3'-primer mismatches, the amplification of spiders can be considerably reduced, enriching the prey DNA during PCR. Based on data from Krehenwinkel et al. 2017b, 2019c

To further optimize prey recovery and reduce the necessary sequencing depth, lineage-specific PCR was recently suggested (Fig. 5; Krehenwinkel et al. 2019c). Single mismatches at the 3'-end of a PCR primer can lead to a massive drop in amplification efficiency (Kwok et al. 1990). If primers are designed in very conserved regions but end at a lineage-diagnostic SNP, which distinguishes spiders from their insect prey, spiders will be mostly blocked from amplification. At the same time, the primer still amplifies a wide variety of arthropod prey. Blocking spiders from amplification also enables detection of prey for very long time periods, possibly up to a month after feeding (Krehenwinkel et al. 2019c). A downside of this approach is that spider-spider predation cannot be detected. While prey enrichment can now be routinely performed from spider gut content, further standardizations of the protocol may be necessary to integrate the resulting data into previously generated prey community data.

Non-lethal monitoring of spider prey communities

All methods mentioned above rely on DNA extraction from spiders or their body parts. Corse et al. (2019) suggest the use of DNA extracts from spider webs as an alternative source of prey DNA, without harming the spider. This method has several drawbacks. First, spider webs can collect airborne DNA from the environment, in addition to “bycatch” of insects that the spiders do not eat. This can lead to false positives if DNA from webs is used as a proxy for the spider's diet. Also, many

spiders rebuild their webs on a daily basis. Web DNA then only allows detection of the daily prey catch (as well as the bycatch described above), in contrast to several weeks recovered by gut content analysis. Many spiders do not spin capture webs but are active hunters, additionally limiting the broad applicability of this method. Another alternative was suggested by Sint et al. (2015), who used DNA extracted from spider feces as source of prey DNA. This is a promising approach but may be logistically challenging, as spiders must be kept in captivity until feces can be collected. Moreover, recent work in carabid beetles has shown that feces recover a less diverse, and therefore biased, prey spectrum compared to gut content extractions (Kamenova et al. 2018).

Pitfalls of HTS-based gut content analysis and how to avoid them

HTS-based gut content analysis has been shown to yield reliable and comprehensive prey spectra, allowing exploration of food web structure in whole communities of spiders. However, the method still has several issues. PCR-based amplification of prey DNA is very sensitive to contamination. Theoretically, a few molecules of insect DNA are sufficient to be amplified. This DNA can also derive from external sources, e.g., from insects caught or stored together with the spider specimen. This contamination can be minimized by bleach treatment of the collected specimens before gut content analysis (Greenstone et al. 2012). Another source of contamination is parasitoid larvae,

which may be located inside the spider but are not actual prey. Additional organisms inside the spider's tissue, such as nematodes, fungi, and bacteria, can also be coamplified along with the prey. Prey data must therefore be carefully analyzed to identify such potentially confounding factors.

Amplification bias can also strongly affect taxon recovery in HTS-based gut content analysis. This problem can be mitigated using multiplex PCR assays targeting several loci (Krehenwinkel et al. 2019c), which enables a good qualitative recovery of prey. Accurate quantitative assessments of prey communities, however, are not possible. This is partly due to biased amplification of different taxa, and partly to the timing of prey consumption. DNA of the recently ingested prey will outweigh that of earlier meals, which will already be mostly degraded. However, by completely omitting quantitative information, the prey spectrum is artificially biased toward very rare taxa; thus, the number of reads obtained for different prey taxa should be taken into account even if abundance per se cannot be reliably inferred (Deagle et al. 2019).

Another shortcoming of current HTS-based protocols is that they fail to detect cannibalism. The short amplified barcode sequences for gut content analysis are usually shared within a species, making it impossible to distinguish the DNA of cannibalized prey from that of the predator. Recent work (Michálek et al. 2017) suggests using intraspecific haplotypic variation to identify cannibalism events, but this method only works if the spider population has very high-haplotypic variation. Hence, in most spider species, the number of cannibalism events would be considerably underestimated.

Secondary predation is yet another issue, particularly for protocols with a long detection half-life of prey DNA. For example, gut DNA extracts of a spider that has fed on a ladybird (Coccinellidae) may also contain DNA of the ladybird's own aphid prey. However, secondary prey DNA is expected to be more degraded and less abundant than the DNA of an actual prey item, so a careful analysis of prey community data and the development of sequence coverage cutoffs may mitigate this issue. Such sequence coverage cutoffs will have to be derived experimentally in the future to enable accurate assignments of real prey taxa in gut content studies.

Future outlook: practical and theoretical developments in the field

Development of laboratory and field protocols

Recent technological developments have greatly contributed to the field of DNA barcoding. Whole communities can now be routinely characterized for manageable cost and effort. However, the field is still in its infancy, and further developments are warranted. One important focus is the completion of DNA barcode reference libraries. Only with complete

reference databases can DNA barcoding be used to its full potential. Museum collections are a promising source of specimens from which to generate barcode data for the world's spider biota. Also, future species descriptions should be coupled with the deposition of a DNA barcode sequence. Considering the limitations of the single-locus DNA barcoding, new unlinked barcode loci should additionally be developed. These should include information from the nuclear genome and be variable enough to distinguish species but also include conserved sequences which allow the design of universal primers. A set of multiple, unlinked DNA barcoding loci would greatly facilitate taxonomic discoveries in spiders and may aid in community phylogenetic analysis.

A focus of future research should also be on the optimization of quantitative taxon recovery from metabarcoding. Alternatively, further developments in amplification-free approaches may lead to a drop in cost, allowing this method to be applicable to the whole-community samples. Further optimizations should be performed on long read protocols so that they can be used for accurate metabarcoding analyses. This would considerably improve the phylogenetic resolution of metabarcoding data. With future simplifications to the protocol, portable barcoding could develop into a routine methodology for the exploration of remote ecosystems around the world. Gut content sequencing is currently revolutionizing our understanding of cryptic prey-predator interactions in arthropod communities. With further experimental developments, for example, into the avoidance of false positives and the enrichment of prey DNA, the methodology will enable an in-depth understanding of the arthropod food web structure, which is also critical for understanding the food web relationships at higher trophic levels.

Linking theoretical biology and DNA barcoding

High throughput sequencing-based DNA barcoding and metabarcoding have provided scientists with community-level datasets of unprecedented completeness and resolution. Nevertheless, the theoretical tools available for analyzing these data are still somewhat limited. Recent efforts, however, show great promise for improving the power and accuracy of DNA barcode data for the analysis of community species richness, abundance, phylogenetics, and interactions. Here we provide an overview of the current developments and future perspectives of integrating DNA barcoding data into theory.

Theoreticians are facing new opportunities for making inferences about past processes that have contributed to structuring communities using community-scale sequence data. Events at different timescales are recorded in different aspects of these data, with abundance distributions reflecting short, ecological timescales, population genetic variation reflecting medium timescales, and phylogenetic diversity reflecting long timescales. For example, if abundances can be estimated from bulk-

sampled sequence data using metabarcoding, then a variety of methods can be applied to differentiate neutral from non-neutral processes (Harpole and Tilman 2006; Tilman 2004), estimate assembly model parameters (Haegeman and Etienne 2017), or infer equilibrium state of the community using mechanistic theory (Jabot and Chave 2011) or tools from statistical mechanics (Harte and Newman 2014; Rominger and Merow 2017).

The distribution of genetic variation within a community provides another axis of information which is complementary to the abundance distribution (Vellend 2005; Vellend et al. 2014). Recently, Overcast et al. (2019) described a mechanistic model of community assembly that can generate linked patterns of abundance and genetic diversity under an assumption of joint ecological (Hubbell 2011) and evolutionary (Kimura 1983) neutrality to estimate community abundance structure using only intraspecific genetic variation. This method can serve as an alternative or a complement to estimates of abundance distributions from metabarcode data, which are confounded by PCR amplification bias as described above. As a proof of concept for this method, Overcast et al. (2019) analyzed the densely sampled abundances and community-scale population genetic data (COI sequences) from a community of spiders on La Réunion (Emerson et al. 2017) and demonstrated that the abundance structure of the community could be accurately estimated using only the intraspecific genetic variation.

Analysis of the community phylogenies provides a deep-time lens on community structure which can be used to estimate speciation and extinction rates (Manceau et al. 2015) and make inferences about diversification processes (Emerson and Gillespie 2008; Morlon 2014; Pearse et al. 2014). Recent methods have also been developed to simultaneously model trait evolution and species diversification (Weber et al. 2017) to investigate the importance of competition in shaping evolutionary radiations (Aristide and Morlon 2019), and the joint contribution of competition and environmental filtering in structuring ecological communities (Ruffley et al. 2019). Community-scale trait data can also be analyzed along with metabarcoding data in a hierarchical modeling framework to further account for feedbacks among processes happening at disparate timescales (Overcast et al. n.d.). Such theoretical developments enable increasingly reliable and detailed inferences on past processes in shaping present-day patterns, yielding many exciting new perspectives on community assembly.

Acknowledgments IO acknowledges the support of the Mina Rees Dissertation Fellowship in the Sciences. We also acknowledge the German Center for Integrative Biodiversity Research (sDiv program) for fostering discussions that gave rise to some of this work.

Funding information Open Access funding provided by Projekt DEAL. SK was funded by a kickstart grant from the Okinawa Institute of Science and Technology to Evan Economo. Part of this work was funded by the National Science Foundation's Dimensions of Biodiversity program (DEB 1241253) to RGG and a Deutsche Forschungsgemeinschaft post-doctoral fellowship to HK.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agnarsson I, Kuntner M (2007) Taxonomy in a changing world: seeking solutions for a science in crisis. *Syst Biol* 56:531–539
- Agustí N, Shayler SP, Harwood JD, Vaughan IP, Sunderland KD, Symondson WOC (2003) Collembola as alternative prey sustaining spiders in arable ecosystems: prey detection within predators using molecular markers. *Mol Ecol* 12:3467–3475
- Andersen JC, Mills NJ (2012) DNA extraction from museum specimens of parasitic hymenoptera. *PLoS One*. <https://doi.org/10.1371/journal.pone.0045549>
- Aristide L, Morlon H (2019) Understanding the effect of competition during evolutionary radiations: an integrated model of phenotypic and species diversification. *Ecol Lett*. <https://doi.org/10.1111/ele.13385>
- Astrin JJ, Höfer H, Spelda J, Holstein J, Bayer S, Hendrich L, Huber BA, Kielhorn KH, Krammer HJ, Lemke M, Monje JC (2016) Towards a DNA barcode reference database for spiders and harvestmen of Germany. *PLoS One*. <https://doi.org/10.1371/journal.pone.0162624>
- Barker GM (2002) Phylogenetic diversity: a quantitative framework for measurement of priority and achievement in biodiversity conservation. *Biol J Linn Soc* 76:165–194
- Barrett RD, Hebert PD (2005) Identifying spiders through DNA barcodes. *Can J Zool* 83:481–491
- Benavides LR, Giribet G, Hormiga G (2017) Molecular phylogenetic analysis of “pirate spiders” (Araneae, Mimetidae) with the description of a new African genus and the first report of maternal care in the family. *Cladistics* 33:375–405
- Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends Ecol Evol* 16:314–321
- Binford GJ (2001) Differences in venom composition between orb-weaving and wandering Hawaiian *Tetragnatha* (Araneae). *Biol J Linn Soc* 74:581–595
- Blagoev GA, deWaard JR, Ratnasingham S, deWaard SL, Lu L, Robertson J, Telfer AC, Hebert PD (2016) Untangling taxonomy: a DNA barcode reference library for Canadian spiders. *Mol Ecol Resour* 16:325–341
- Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, Douglas WY, De Bruyn M (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol Evol* 29:358–367
- Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, McGill SK, Dougherty MK (2019) High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res*. <https://doi.org/10.1093/nar/gkz569>
- Calus ST, Ijaz UZ, Pinto AJ (2018) NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the

- nanopore sequencing platform. GigaScience. <https://doi.org/10.1093/gigascience/giy140>
- Čandek K, Kuntner M (2015) DNA barcoding gap: reliable species identification over morphological and geographical scales. *Mol Ecol Resour* 15:268–277
- Cardoso P, Pekár S, Jocqué R, Coddington JA (2011) Global patterns of guild composition and functional diversity of spiders. *PLoS One*. <https://doi.org/10.1371/journal.pone.0021710>
- Carew ME, Coleman RA, Hoffmann AA (2018) Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding? *PeerJ*. <https://doi.org/10.7717/peerj.4980>
- Corse E, Tougaard C, Archambaud-Suard G, Agnès JF, Messu Mandeng FD, Bilong Bilong CF, Duneau D, Zinger L, Chappaz R, Xu CC, Megléc E (2019) One-locus-several-primers: a strategy to improve the taxonomic and haplotypic coverage in diet metabarcoding studies. *Ecol Evol* 9:4603–4620
- Cotoras D, Murray G, Kapp J, Gillespie R, Griswold C, Simison W, Green R, Shapiro B (2017) Ancient DNA resolves the history of *Tetragnatha* (Araneae, Tetragnathidae) spiders on Rapa Nui. *Genes*. <https://doi.org/10.3390/genes8120403>
- Crampton-Platt A, Douglas WY, Zhou X, Vogler AP (2016) Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience*. 5:1–11. <https://doi.org/10.1186/s13742-016-0120-y>
- Crespo LC, Domènech M, Enguñanos A, Malumbres-Olarte J, Cardoso P, Moya-Laraño J, Frías-López C, Macías-Hernández N, De Mas E, Mazzuca P, Mora E (2018) A DNA barcode-assisted annotated checklist of the spider (Arachnida, Araneae) communities associated to white oak woodlands in Spanish national parks. *Biodivers Data J* 6:1–273
- Cristescu ME (2014) From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends Ecol Evol* 29:566–571
- de Kerdrel G, Andersen JC, Kennedy SR, Gillespie R, Krehenwinkel H (2020) Rapid and cost-effective generation of single specimen multilocus barcoding data from whole arthropod communities by multiple levels of multiplexing. *Sci Rep UK* in press
- Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Mol Ecol* 18:2022–2038
- Deagle BE, Thomas AC, McInnes JC, Clarke LJ, Vesterinen EJ, Clare EL, Kartzinel TR, Eveson JP (2019) Counting with DNA in metabarcoding studies: how should we convert sequence reads to dietary data? *Mol Ecol* 28:391–406
- Dopheide A, Tooman LK, Grosser S, Agabiti B, Rhode B, Xie D, Stevens MI, Nelson N, Buckley TR, Drummond AJ, Newcomb RD (2019) Estimating the biodiversity of terrestrial invertebrates on a forested island using DNA barcodes and metabarcoding data. *Ecol Appl*. <https://doi.org/10.1002/eap.1877>
- Dupuis JR, Roe AD, Sperling FA (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Mol Ecol* 21:4422–4436
- Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10:996–998
- Edgar RC (2018) Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* 34:2371–2375
- Eitzinger B, Abrego N, Gravel D, Huotari T, Vesterinen EJ, Roslin T (2019) Assessing changes in arthropod predator–prey interactions through DNA-based gut content analysis—variable environment, stable diet. *Mol Ecol* 28:266–280
- Elbrecht V, Leese F (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLoS One*. <https://doi.org/10.1371/journal.pone.0130324>
- Elbrecht V, Vámos EE, Meissner K, Aroviita J, Leese F (2017) Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods Ecol Evol* 8:1265–1275
- Elbrecht V, Vámos EE, Steinke D, Leese F (2018) Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ*. <https://doi.org/10.7717/peerj.4644>
- Emerson BC, Gillespie RG (2008) Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol Evol* 23:619–630
- Emerson BC, Casquet J, López H, Cardoso P, Borges PA, Mollaret N, Oromí P, Strasberg D, Thébaud C (2017) A combined field survey and molecular identification protocol for comparing forest arthropod biodiversity across spatial scales. *Mol Ecol Resour* 17:694–707
- Fadrosh DW, Ma B, Gajer P, Sengamalai N, Ott S, Brotman RM, Ravel J (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*. <https://doi.org/10.1186/2049-2618-2-6>
- Foelix R (2011) *Biology of spiders*. Oxford University Press USA, New York
- Foley S, Lüddecke T, Cheng DQ, Krehenwinkel H, Künzel S, Longhorn SJ, Wendt I, von Wirth V, Tänzler R, Vences M, Piel WH (2019) Tarantula phylogenomics: a robust phylogeny of deep theraphosid clades inferred from transcriptome data sheds light on the prickly issue of urticating setae evolution. *Mol Phylogenet Evol*. <https://doi.org/10.1016/j.ympev.2019.106573>
- Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends Ecol Evol* 27:480–488
- Gibson J, Shokralla S, Porter TM, King I, van Konynenburg S, Janzen DH, Hallwachs W, Hajibabaei M (2014) Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *P Natl Acad Sci USA* 111:8007–8012
- Gillespie RG, Benjamin SP, Brewer MS, Rivera MAJ, Roderick GK (2018) Repeated diversification of ecomorphs in Hawaiian stick spiders. *Curr Biol* 28:941–947
- Greenstone MH, Shufan KA (2003) Spider predation: species-specific identification of gut contents by polymerase chain reaction. *J Arachnol* 31:131–135
- Greenstone MH, Weber DC, Coudron TA, Payton ME, Hu JS (2012) Removing external DNA contamination from arthropod predators destined for molecular gut-content analysis. *Mol Ecol Resour* 12:464–469
- Grey J, Thackeray SJ, Jones RI, Shine A (2002) Ferox trout (*Salmo trutta*) as ‘Russian dolls’: complementary gut content and stable isotope analyses of the loch ness foodweb. *Freshw Biol* 47:1235–1243
- Gruner DS (2004) Attenuation of top-down and bottom-up forces in a complex terrestrial community. *Ecology* 85:3010–3022
- Haegeman B, Etienne RS (2017) A general sampling formula for community structure data. *Methods Ecol Evol* 8:1506–1519
- Hajibabaei M, Singer GA, Hebert PD, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet* 23:167–172
- Hajibabaei M, Spall JL, Shokralla S, van Konynenburg S (2012) Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecol* 12:1–10. <https://doi.org/10.1186/1472-6785-12-28>
- Hambäck PA, 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. *Ecol Evol* 6:8431–8439
- Harpole WS, Tilman D (2006) Non-neutral patterns of species abundance in grassland communities. *Ecol Lett* 9:15–23
- Harte J, Newman EA (2014) Maximum information entropy: A foundation for ecological theory. *Trends Ecol Evol* 29:384–389

- Hebert PD, Gregory TR (2005) The promise of DNA barcoding for taxonomy. *Syst Biol* 54:852–859
- Hebert PD, Cywinski A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *P Roy Soc Lond B Bio* 270:313–321
- Hebert PD, Stoeckle MY, Zemplak TS, Francis CM (2004a) Identification of birds through DNA barcodes. *PLoS Biol*. <https://doi.org/10.1371/journal.pbio.0020312>
- Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004b) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *P Natl Acad Sci USA* 101:14812–14817
- Hebert PD, Braukmann TW, Prosser SW, Ratnasingham S, deWaard JR, Ivanova NV, Janzen DH, Hallwachs W, Naik S, Sones JE, Zakharov EV (2018) A sequel to sanger: amplicon sequencing that scales. *BMC Genomics* 19:1–14. <https://doi.org/10.1186/s12864-018-4611-3>
- Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Q Rev Biol* 66:411–453
- Hiruki LM, Schwartz MK, Boveng PL (1999) Hunting and social behaviour of leopard seals (*Hydrurga leptonyx*) at Seal Island, South Shetland Islands, Antarctica. *J Zool* 249:97–109
- Hubbell SP (2011) The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton
- Hurst GD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *P Roy Soc Lond B Bio* 272:1525–1534
- Irwin DE, Rubtsov AS, Panov EN (2009) Mitochondrial introgression and replacement between yellowhammers (*Emberiza citrinella*) and pine buntings (*Emberiza leucocephalos*) (Aves: Passeriformes). *Biol J Linn Soc* 98:422–438
- Jabot F, Chave J (2011) Analyzing tropical forest tree species abundance distributions using a nonneutral model and through approximate Bayesian inference. *Am Nat* 178:E37–E47
- Jones MB, Highlander SK, Anderson EL, Li W, Dayrit M, Klitgord N, Fabani MM, Seguritan V, Green J, Pride DT, Yooseph S (2015) Library preparation methodology can influence genomic and functional predictions in human microbiome research. *P Natl Acad Sci USA* 112:14024–14029
- Kamenova S, Mayer R, Rubbmark OR, Coissac E, Plantegenest M, Traugott M (2018) Comparing three types of dietary samples for prey DNA decay in an insect generalist predator. *Mol Ecol Resour* 18:966–973
- Kennedy S, Lim JY, Clavel J, Krehenwinkel H, Gillespie RG (2019) Spider webs, stable isotopes and molecular gut content analysis: multiple lines of evidence support trophic niche differentiation in a community of Hawaiian spiders. *Funct Ecol* 33:1722–1733
- Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge
- King RA, Moreno-Ripoll R, Agustí N, Shayler SP, Bell JR, Bohan DA, Symondson WO (2011) Multiplex reactions for the molecular detection of predation on pest and nonpest invertebrates in agroecosystems. *Mol Ecol Resour* 11:370–373
- Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79:5112–5120
- Krehenwinkel H, Pekár S (2015) An analysis of factors affecting genotyping success from museum specimens reveals an increase of genetic and morphological variation during a historical range expansion of a European spider. *PLoS One*. <https://doi.org/10.1371/journal.pone.0136337>
- Krehenwinkel H, Tautz D (2013) Northern range expansion of European populations of the wasp spider *Argiope bruennichi* is associated with global warming–correlated genetic admixture and population-specific temperature adaptations. *Mol Ecol* 22:2232–2248
- Krehenwinkel H, Graze M, Rödder D, Tanaka K, Baba YG, Muster C, Uhl G (2016) A phylogeographical survey of a highly dispersive spider reveals eastern Asia as a major glacial refugium for Palaearctic fauna. *J Biogeogr* 43:1583–1594
- Krehenwinkel H, Wolf M, Lim JY, Rominger AJ, Simison WB, Gillespie RG (2017a) Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Sci Rep UK*. 7:1–12. <https://doi.org/10.1038/s41598-017-17333-x>
- Krehenwinkel H, Kennedy S, Pekár S, Gillespie RG (2017b) A cost-efficient and simple protocol to enrich prey DNA from extractions of predatory arthropods for large-scale gut content analysis by Illumina sequencing. *Methods Ecol Evol* 8:126–134
- Krehenwinkel H, Kennedy SR, Rueda A, Lam A, Gillespie RG (2018) Scaling up DNA barcoding – primer sets for simple and cost efficient arthropod systematics by multiplex PCR and Illumina amplicon sequencing. *Methods Ecol Evol* 9:2181–2193
- Krehenwinkel H, Pomerantz A, Henderson JB, Kennedy SR, Lim JY, Swamy V, Shoobridge JD, Graham N, Patel NH, Gillespie RG, Prost S (2019a) Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *GigaScience*. <https://doi.org/10.1093/gigascience/giz006>
- Krehenwinkel H, Pomerantz A, Prost S (2019b) Genetic biomonitoring and biodiversity assessment using portable sequencing technologies: current uses and future directions. *Genes*. <https://doi.org/10.3390/genes10110858>
- Krehenwinkel H, Kennedy SR, Adams SA, Stephenson GT, Roy K, Gillespie RG (2019c) Multiplex PCR targeting lineage-specific SNPs: a highly efficient and simple approach to block out predator sequences in molecular gut content analysis. *Methods Ecol Evol* 10:982–993
- Krushelnicky PD, Loope LL, Gillespie RG (2007) Inventory of arthropods of the west slope shrubland and alpine ecosystems of Haleakala National Park. Honolulu (HI): Pacific cooperative studies unit, University of Hawaii at Manoa, Department of Botany. PCSU technical report, 148. 52 pages
- Kulkarni S, Wood H, Lloyd M, Hormiga G (2020) Spider-specific probe set for ultraconserved elements offers new perspectives on the evolutionary history of spiders (Arachnida, Araneae). *Mol Ecol Resour* 20:185–203
- Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsky JJ (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res* 18:999–1005
- Lafferty KD, Page CJ (1997) Predation on the endangered tidewater goby, *Eucyclogobius newberryi*, by the introduced African clawed frog, *Xenopus laevis*, with notes on the frog's parasites. *Copeia* 1997:589–592
- Lange V, Böhme I, Hofmann J, Lang K, Sauter J, Schöne B, Paul P, Albrecht V, Andreas JM, Baier DM, Nething J (2014) Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics*. <https://doi.org/10.1186/1471-2164-15-63>
- Leavitt DH, Starrett J, Westphal MF, Hedin M (2015) Multilocus sequence data reveal dozens of putative cryptic species in a radiation of endemic Californian mygalomorph spiders (Araneae, Mygalomorphae, Nemesiidae). *Mol Phylogenet Evol* 91:56–67
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front Zool*. <https://doi.org/10.1186/1742-9994-10-34>
- Longhorn SJ, Nicholas M, Chuter J, Vogler AP (2007) The utility of molecular markers from non-lethal DNA samples of the CITES II

- protected "tarantula" *Brachypelma vegans* (Araneae, Theraphosidae). *J Arachnol* 35:278–292
- Macías-Hernández N, Athey K, Tonzo V, Wangenstein OS, Amedo M, Harwood JD (2018) Molecular gut content analysis of different spider body parts. *PLoS One*. <https://doi.org/10.1371/journal.pone.0196589>
- Manceau M, Lambert A, Morlon H (2015) Phylogenies support out-of-equilibrium models of biodiversity. *Ecol Lett* 18:347–356
- Manley LJ, Ma D, Levine SS (2016) Monitoring error rates in Illumina sequencing. *J Biomol Tech* 27:125–128
- Marquina D, Esparza-Salas R, Roslin T, Ronquist F (2019) Establishing arthropod community composition using metabarcoding: surprising inconsistencies between soil samples and preservative ethanol and homogenate from malaise trap catches. *Mol Ecol Resour*. <https://doi.org/10.1111/1755-0998.13071>
- Martins FM, Galhardo M, Filipe AF, Teixeira A, Pinheiro P, Paupério J, Alves PC, Beja P (2019) Have the cake and eat it: optimizing non-destructive DNA metabarcoding of macroinvertebrate samples for freshwater biomonitoring. *Mol Ecol Resour*. <https://doi.org/10.1111/1755-0998.13012>
- Meehan CJ, Olson EJ, Reudink MW, Kyser TK, Curry RL (2009) Herbivory in a spider through exploitation of an ant-plant mutualism. *Curr Biol* 19:R892–R893
- Meier R, Wong W, Srivathsan A, Foo M (2016) \$1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples. *Cladistics* 32:100–110
- Meierotto S, Sharkey MJ, Janzen DH, Hallwachs W, Hebert PD, Chapman EG, Smith MA (2019) A revolutionary protocol to describe understudied hyperdiverse taxa and overcome the taxonomic impediment. *Deut Entomol Z* 66:119–145
- Menegon M, Cantaloni C, Rodriguez-Prieto A, Centomo C, Abdelfattah A, Rossato M, Bernardi M, Xumerle L, Loader S, Delledonne M (2017) On site DNA barcoding by nanopore sequencing. *PLoS One*. <https://doi.org/10.1371/journal.pone.0184741>
- Michálek O, Petráková L, Pekár S (2017) Capture efficiency and trophic adaptations of a specialist and generalist predator: a comparison. *Ecol Evol* 7:2756–2766
- Miller JA, Beentjes KK, van Helsdingen P, IJland S (2013) Which specimens from a museum collection will yield DNA barcodes? A time series study of spiders in alcohol. *ZooKeys* 365:245–261
- Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *PLoS Biol*. <https://doi.org/10.1371/journal.pbio.0020354>
- Morlon H (2014) Phylogenetic approaches for studying diversification. *Ecol Lett* 17:508–525
- Nichols RV, Vollmers C, Newsom LA, Wang Y, Heintzman PD, Leighton M, Green RE, Shapiro B (2018) Minimizing polymerase biases in metabarcoding. *Mol Ecol Resour* 18:927–939
- Noguchi H, Park J, Takagi T (2006) MetaGene: prokaryotic gene finding from environmental shotgun sequences. *Nucleic Acids Res* 34:5623–5630
- Nyffeler M, Birkhofer K (2017) An estimated 400–800 million tons of prey are annually killed by the global spider community. *Sci Nat Heidelberg*. <https://doi.org/10.1007/s00114-017-1440-1>
- Nyffeler M, Olson EJ, Symondson WO (2016) Plant-eating by spiders. *J Arachnol* 44:15–28
- Nyffeler M, Şekercioglu ÇH, Whelan CJ (2018) Insectivorous birds consume an estimated 400–500 million tons of prey annually. *Sci Nat Heidelberg* 105:1–13. <https://doi.org/10.1007/s00114-018-1571-z>
- Obertegger U, Cieplinski A, Fontaneto D, Papakostas S (2018) Mitonuclear discordance as a confounding factor in the DNA taxonomy of monogonont rotifers. *Zool Scr* 47:122–132
- Overcast I, Emerson BC, Hickerson MJ (2019) An integrated model of population genetics and community ecology. *J Biogeogr* 46:816–829
- Overcast I, Ruffley M, Rosindell J, Harmon L, Borges P, Chase J, Emerson B, Etienne RS, Gillespie R, Krehenwinkel H, Mahler L, Massol F, Parent C, Patiño J, Peter B, Week B, Wagner C, Hickerson MJ, Rominger AJ (2020) What a MESS!: On the distribution of abundance, genetic, and functional diversity in ecological communities. *bioRxiv*
- Papadopoulou A, Taberlet P, Zinger L (2015) Metagenome skimming for phylogenetic community ecology: a new era in biodiversity research. *Mol Ecol* 24:3515–3517
- Payne A, Holmes N, Rakyan V, Loose M (2018) BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* 35:2193–2198
- Pearse WD, Purvis A, Cavender-Bares J, Helmus MR (2014) Metrics and models of community phylogenetics. In: Garamszegi LZ (ed) *Modern phylogenetic comparative methods and their application in evolutionary biology*. Springer, Berlin, pp 451–464
- Petersen SD, Mason T, Akber S, West R, White B, Wilson P (2007) Species identification of tarantulas using exuviae for international wildlife law enforcement. *Conserv Genet* 8:497–502
- Petráková L, Líznařová E, Pekár S, Haddad CR, Sentenská L, Symondson WO (2015) Discovery of a monophagous true predator, a specialist termite-eating spider (Araneae: Ammoxenidae). *Sci Rep UK* 5:1–10. <https://doi.org/10.1038/srep14013>
- Piñol J, San Andrés 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. *Mol Ecol Resour* 14:18–26
- Piñol J, Mir G, Gomez-Polo P, Agustí N (2015) Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Mol Ecol Resour* 15:819–830
- Piñol J, Senar MA, Symondson WO (2019) The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Mol Ecol* 28:407–419
- Piper AM, Batovska J, Cogan NO, Weiss J, Cunningham JP, Rodoni BC, Blacket MJ (2019) Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *GigaScience*. <https://doi.org/10.1093/gigascience/giz092>
- Pomerantz A, Peñafiel N, Arteaga A, Bustamante L, Pichardo F, Coloma LA, Barrio-Amorós CL, Salazar-Valenzuela D, Prost S (2018) Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *GigaScience*. <https://doi.org/10.1093/gigascience/giy033>
- Porco D, Rougerie R, Deharveng L, Hebert P (2010) Coupling non-destructive DNA extraction and voucher retrieval for small soft-bodied arthropods in a high-throughput context: the example of Collembola. *Mol Ecol Resour* 10:942–945
- Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, automatic barcode gap discovery for primary species delimitation. *Mol Ecol* 21:1864–1877
- Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, Ouedraogo N (2016) Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530:228–232
- Raso L, Sint D, Mayer R, Plangg S, Recheis T, Brunner S, Kaufmann R, Traugott M (2014) Intraguild predation in pioneer predator communities of alpine glacier forelands. *Mol Ecol* 23:3744–3754
- Riechert SE, Lockley T (1984) Spiders as biological control agents. *Annu Rev Entomol* 29:299–320
- Robinson E, Blagojev G, Hebert P, Adamowicz S (2009) Prospects for using DNA barcoding to identify spiders in species-rich genera. *ZooKeys* 16:27–46
- Rominger AJ, Merow C (2017) meteR: an R package for testing the maximum entropy theory of ecology. *Methods Ecol Evol* 8:241–247
- Roubinet E, Birkhofer K, Malsher G, Staudacher K, Ekblom B, Traugott M, Jonsson M (2017) Diet of generalist predators reflects effects of

- cropping period and farming system on extra-and intraguild prey. *Ecol Appl* 27:1167–1177
- Ruffley M, Peterson K, Week B, Tank DC, Harmon LJ (2019) Identifying models of trait-mediated community assembly using random forests and approximate Bayesian computation. *Ecol Evol* in press
- Saitoh S, Aoyama H, Fujii S, Sunagawa H, Nagahama H, Akutsu M, Shinzato N, Kaneko N, Nakamori T (2016) A quantitative protocol for DNA metabarcoding of springtails (Collembola). *Genome* 59:705–723
- Satler JD, Carstens BC, Hedin M (2013) Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). *Syst Biol* 62:805–823
- Schmidt JM, Barney SK, Williams MA, Bessin RT, Coolong TW, Harwood JD (2014) Predator–prey trophic relationships in response to organic management practices. *Mol Ecol* 23:3777–3789
- Schmidt NM, Mosbacher JB, Eitzinger B, Vesterinen EJ, Roslin T (2018) High resistance towards herbivore-induced habitat change in a high Arctic arthropod community. *Biol Lett UK*. <https://doi.org/10.1098/rsbl.2018.0054>
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal Barcoding Consortium (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *P Natl Acad Sci USA* 109:6241–6246
- Shokralla S, Porter TM, Gibson JF, Dobosz R, Janzen DH, Hallwachs W, Golding GB, Hajibabaei M (2015) Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Sci Rep UK*. <https://doi.org/10.1038/srep09687>
- Shokralla S, Gibson J, King I, Baird D, Janzen D, Hallwachs W, Hajibabaei M (2016) Environmental DNA barcode sequence capture: Targeted, PCR-free sequence capture for biodiversity analysis from bulk environmental samples. *BioRxiv*. <https://doi.org/10.1101/087437>
- Sint D, Thurner I, Kaufmann R, Traugott M (2015) Sparing spiders: Faeces as a non-invasive source of DNA. *Front Zool* 12:1–5. <https://doi.org/10.1186/s12983-015-0096-y>
- Sint D, Kaufmann R, Mayer R, Traugott M (2019) Resolving the predator first paradox: arthropod predator food webs in pioneer sites of glacier forelands. *Mol Ecol* 28:336–347
- Srivathsan A, Hartop E, Puniamoorthy J, Lee WT, Kutty SN, Kurina O, Meier R (2019) Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biol* 17:1–20. <https://doi.org/10.1186/s12915-019-0706-9>
- Starrett J, Hedin M (2007) Multilocus genealogies reveal multiple cryptic species and biogeographical complexity in the California turret spider *Antrodiaetus riversi* (Mygalomorphae, Antrodiaetidae). *Mol Ecol* 16:583–604
- Sternes PR, Lee D, Kutyna DR, Borneman AR (2017) A combined metabarcoding and shotgun metagenomic analysis of spontaneous wine fermentation. *GigaScience*. <https://doi.org/10.1093/gigascience/gix040>
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol Ecol* 21:2045–2050
- Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends Ecol Evol* 18:70–74
- Tedersoo L, Anslan S (2019) Towards PacBio-based pan-eukaryote metabarcoding using full-length ITS sequences. *Environ Microbiol Rep*. <https://doi.org/10.1111/1758-2229.12776>
- Tedersoo L, Tooming-Klunderud A, Anslan S (2018) PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytol* 217:1370–1385
- Thomas AC, Deagle BE, Eveson JP, Harsch CH, Trites AW (2016) Quantitative DNA metabarcoding: improved estimates of species proportional biomass using correction factors derived from control material. *Mol Ecol Resour* 16:714–726
- Thomsen PF, Sigsgaard EE (2019) Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. *Ecol Evol* 9:1665–1679
- Thomson LJ, Hoffmann AA (2010) Natural enemy responses and pest control: importance of local vegetation. *Biol Control* 52:160–166
- Thongjue K, Chotigeat W, Bumrungsri S, Thanakiatkrai P, Kitpipit T (2019) A new cost-effective and fast direct PCR protocol for insects based on PBS buffer. *Mol Ecol Resour* 19:691–701
- Tilman D (2004) Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly. *P Natl Acad Sci USA* 101:10854–10861
- Toju H, Baba YG (2018) DNA metabarcoding of spiders, insects, and springtails for exploring potential linkage between above-and below-ground food webs. *Zool Lett* 4:1–12. <https://doi.org/10.1186/s40851-018-0088-9>
- Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol Ecol* 25:929–942
- Vellend M (2005) Species diversity and genetic diversity: parallel processes and correlated patterns. *Am Nat* 166:199–215
- Vellend M, Lajoie G, Bourret A, Múria C, Kembel SW, Garant D (2014) Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Mol Ecol* 23:2890–2901
- Verschut V, Strandmark A, Esparza-Salas R, Hambäck PA (2019) Seasonally varying marine influences on the coastal ecosystem detected through molecular gut analysis. *Mol Ecol* 28:307–317
- Walter MC, Zwirgmaier K, Vette P, Holowachuk SA, Stoecker K, Genzel GH, Antwerpen MH (2017) MinION as part of a biomedical rapidly deployable laboratory. *J Biotechnol* 250:16–22
- Wang WY, Srivathsan A, Foo M, Yamane SK, Meier R (2018) Sorting specimen-rich invertebrate samples with cost-effective NGS barcodes: validating a reverse workflow for specimen processing. *Mol Ecol Resour* 18:490–501
- Weber MG, Wagner CE, Best RJ, Harmon LJ, Matthews B (2017) Evolution in a community context: on integrating ecological interactions and macroevolution. *Trends Ecol Evol* 32:291–304
- Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A, Kolesnikov A, Olson ND, Töpfer A (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* 37:1155–1162
- Whitney TD, Sitvarin MI, Roualdes EA, Bonner SJ, Harwood JD (2018) Selectivity underlies the dissociation between seasonal prey availability and prey consumption in a generalist predator. *Mol Ecol* 27:1739–1748
- Wick RR, Judd LM, Holt KE (2018) Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. *PLoS Comput Biol*. <https://doi.org/10.1371/journal.pcbi.1006583>
- Wise DH, Moldenhauer DM, Halaj J (2006) Using stable isotopes to reveal shifts in prey consumption by generalist predators. *Ecol Appl* 16:865–876
- Wong WH, Tay YC, Puniamoorthy J, Balke M, Cranston PS, Meier R (2014) ‘Direct PCR’ optimization yields a rapid, cost-effective, non-destructive and efficient method for obtaining DNA barcodes without DNA extraction. *Mol Ecol Resour* 14:1271–1280
- Wood HM, Griswold CE, Gillespie RG (2012) Phylogenetic placement of pelican spiders (Archaeidae, Araneae), with insight into evolution of the “neck” and predatory behaviours of the superfamily Palpimanoidea. *Cladistics* 28:598–626
- Yeo D, Puniamoorthy J, Ngiam RWJ, Meier R (2018) Towards holomorphology in entomology: rapid and cost-effective adult–larva matching using NGS barcodes. *Syst Entomol* 43:678–691

- Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z (2012) Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol Evol* 3:613–623
- Zeale MR, Butlin RK, Barker GL, Lees DC, Jones G (2011) Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Mol Ecol Resour* 11:236–244
- Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29:2869–2876
- Zhang GK, Chain FJ, Abbott CL, Cristescu ME (2018) Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. *Evol Appl* 11:1901–1914
- Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*. <https://doi.org/10.1186/2047-217X-2-4>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.