GfÖ Ecological Society of Germany, Austria and Switzerland

Basic and Applied Ecology 50 (2021) 67–76

Basic and Applied Ecology

www.elsevier.com/locate/baae

Building food networks from molecular data: Bayesian or fixed-number thresholds for including links

Alyssa R. Cirtwill[a,b,*], Peter Hambäck[a]

[a]*Department of Ecology, Evolution, and Plant Sciences, Stockholm University, Stockholm, Sweden*
[b]*Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland*

Received 10 February 2020; accepted 26 November 2020
Available online 28 November 2020

Abstract

DNA metabarcoding of faeces or gut contents has greatly increased our ability to construct networks of predators and prey (food webs) by reducing the need to observe predation events directly. The possibility of both false positives and false negatives in DNA sequences, however, means that constructing food networks using DNA requires researchers to make many choices as to which DNA sequences indicate true prey for a particular predator. To date, DNA-based food networks are usually constructed by including any DNA sequence with more than a threshold number of reads. The logic used to select this threshold is often not explained, leading to somewhat arbitrary-seeming networks. As an alternative strategy, we demonstrate how to construct food networks using a simple Bayesian model to suggest which sequences correspond to true prey. The networks obtained using a well-chosen fixed cutoff and our Bayesian approach are very similar, especially when links are resolved to prey families rather than species. We therefore recommend that researchers reconstruct diet data using a Bayesian approach with well-specified assumptions rather than continuing with arbitrary fixed cutoffs. Explicitly stating assumptions within a Bayesian framework will lead to better-informed comparisons between networks constructed by different groups and facilitate drawing together individual case studies into more coherent ecological theory. Note that our approach can easily be extended to other types of ecological networks constructed by DNA metabarcoding of pollen loads, identification of parasite DNA in faeces, etc.

Keywords: DNA metabarcoding; Bayesian statistics; Error reduction; Species interactions

Introduction

Food webs and other ecological networks offer a valuable framework for understanding the structure and functioning of ecological communities. Constructing these networks requires robust information on species' interaction partners, but spatial and temporal variation in interactions, methodological limitations, and trade-offs in sampling effort mean that any observed network is only an approximation of the true community structure (Cirtwill et al., 2019). This is especially true for interactions involving generalist species (because of the greater potential for variation and measurement error) or rare or transient species (because they are usually observed very few times). The potential for errors in generalist species is worrying as these species are likely to be critical to community structure and functioning

*Corresponding author at: Department of Ecology, Evolution, and Plant Sciences, Stockholm University, Stockholm, Sweden.
E-mail address: alyssa.cirtwill@helsinki.fi (A.R. Cirtwill).

https://doi.org/10.1016/j.baae.2020.11.007

(Lai et al., 2012), while rare species are often conservation targets. To address the problem of approximation in empirical networks, new and improved methods for detecting interactions are continually being developed. One such method, DNA metabarcoding, can reveal many interactions which could not otherwise be observed.

Using DNA metabarcoding, prey DNA can be identified in predator gut contents, faeces, or regurgitates (Dalén et al., 2004; Pompanon et al., 2012; Waldner and Traugott, 2012). Parasite DNA can also be recovered from host faeces (Jirků et al., 2012; Pafčo et al., 2018), parasitoids can be identified when sampling hosts (Kitson et al., 2019; Wirta et al., 2014), and plant DNA can be identified from pollen loads recovered from individual pollinators (Bell et al., 2017; Jordano, 2016). Compared to traditional network construction based on direct observation of an interaction (either through field observations, rearing hosts until parasitoid emergence, or morphological identification of parasites, gut contents, or pollen grains), DNA metabarcoding has great potential to reveal interactions involving species which are difficult to observe interacting (e.g., nocturnally-feeding bats Arrizabalaga-Escudero et al., 2018, marine mammals Bowen and Iverson, 2013), do not leave morphologically-distinct traces of interaction (e.g., spiders with external digestion Roubinet et al., 2017; Verschut et al., 2019), or where the prey community is not known (Boyer et al., 2013). DNA metabarcoding can also be less labour-intensive than traditional methods (Bell et al., 2017; Kitson et al., 2019), allowing researchers to increase sample sizes.

Despite these benefits, DNA metabarcoding shares the problems of measurement error which plague other approaches to ecological network construction. There are two common strategies for network construction using molecular methods: restricting the analysis to broad groups of interaction partners using diagnostic PCR (e.g., Staudacher et al., 2016; Waldner and Traugott, 2012) or identifying species using metabarcoding (e.g., Arrizabalaga-Escudero et al., 2018; Deagle et al., 2019; Pompanon et al., 2012; Verschut et al., 2019), both of which are subject to errors and biases that can affect the empirical network produced. For example, DNA amplification may be biased towards a subset of species in the sample. This means that the abundance of sequences detected does not necessarily reflect the proportions of each prey in the diet (Bell et al., 2017; Deagle et al., 2019; Liu et al., 2019). There is also a risk that contamination or the formation of chimeric DNA during PCR amplification can result in the identification of species which are not truly present in the sample (Alberdi et al., 2018). Finally, sequence reads may be mislabelled during PCR amplification or sequencing (i.e., "tag-jumping") (Mathieu et al., 2020; Schnell et al., 2015). To reduce the risk of including such false positives, researchers commonly discard potential partners represented by few sequence reads from their data. This strategy creates a trade-off between eliminating potential false positives and retaining rare DNA sequences that represent prey which are truly eaten ("true prey"), plants which were truly visited, etc. This trade-off is analogous to the problem in graph theory of "pruning" a noisy observed network in order to approach the unknown, true "core" of the network (Dianati, 2016). Although there are many suggestions for suitable thresholds (e.g., 1% of sequences in Deagle et al., 2019, 1, 5, 10, or 100 reads in Alberdi et al., 2018, 5 reads in Verschut et al., 2019), in practice the choice of threshold is often somewhat arbitrary (or at least seldom justified in print). Why, for example, is a cutoff of five sequence reads superior to four or six reads?

As an alternative to an arbitrary cutoff value, we can use knowledge about the community as a whole to justify including or excluding uncertain links. Using food webs as an example, this changes the question from "Did we find enough reads of prey Y's DNA in predator X's gut to decide that X really ate Y?" to "Given what we know of the overall community structure and n observed reads of prey Y in the gut of predator X, what is the probability that predator X really ate prey Y?". Potential prey can be included or excluded based on these probabilities by treating the likelihood of a predator-prey interaction as a Bayesian process. Using a Bayesian approach prompts researchers to explicitly state their assumptions (Cirtwill et al., 2019; Spiegelhalter et al., 2000), which facilitates the evaluation and comparison of diets based on DNA metabarcoding across studies. While we cannot know the true structure of an empirical ecological network, we can strive to be as transparent as possible when estimating this structure.

In order to compare gut-content DNA analyses using a Bayesian framework with previous work using fixed cutoffs, we must understand which types of interactions are retained and rejected by the two approaches. We use gut-content DNA derived from *Pardosa* spiders, originally collected in order to test whether different species and life stages have overlapping diets (Verschut et al., 2019), to compare estimated spider diets between the Bayesian framework and approaches using a fixed cutoff and one based on the removal of tag jumping errors. We show that, while there are differences in the sets of links included using different methods of network construction, all three approaches identify a similar core of common, well-supported links. Given this similarity, we advocate for approaches to network construction that promote more explicit statements of the assumptions involved, such as our Bayesian framework.

Materials and methods

Data collection and fixed-cutoff network assembly

Data used in this study were collected and originally analysed in Verschut et al. (2019). Prey DNA was extracted from spider guts and sequenced using primers that do not amplify *Pardosa* DNA (see Verschut et al., 2019 for full details). Sequences were then compared to DNA barcodes

in the Barcode of Life Database (Ratnasingham and Hebert, 2007). As this database is continuously updated and expanded, we repeated the comparison of unique sequences identified during the Verschut et al. (2019) study to the database (see *Appendix A1: section S1*).

For each of eight sites, we assembled networks of interactions between prey taxa and spider individuals using all fixed cutoff values c between 1 and 101 (*Appendix A1: section S1*). Spiders were grouped by age and species: (i) adult *Pardosa agricola*, (ii) adult *P. prativaga*, (iii) subadults of both species, and (iv) juveniles of both species; subadults and juveniles could not be identified to species. The cutoff c=6 was used in Verschut et al. (2019) and was chosen based on the accumulation curve of unique links in order to strike a balance betweeen detection of diversity and error removal (Alberdi et al., 2018). This curve was steep when c<6 (many links and families were removed at each step) but shallower when c>6 (Fig. 1A-C). We therefore used the fixed-cutoff networks where c=6 as a baseline for a high-quality approximation of the true spider-prey network against which to compare the other networks.

Controlling for the risk of tag-jumping

Tag-jumping, where sequences are mis-labelled as coming from the wrong sample, is a potential issue in DNA metabarcoding studies using high-throughput sequencing and primer tags (Schnell et al., 2015). While rates of tag-jumping are generally low, these substitutions may nevertheless create substantial numbers of false positive links. To reduce the number of false positives included in downstream analyses, recent studies have removed a percentage of reads in each sample (Matesanz et al., 2019) or subtracted a correction factor based on the observed abundance of contaminants Makiola et al. (2019). Here, we apply a conceptually similar correction.

We assume that tag-jumping is a random process, meaning that the likelihood of a read being assigned to a sample by tag-jumping increases with the total number of reads for a prey. The overrall probability of tag jumping depends on the tagging structure and methods used during library building, and can be estimated from the distribution of reads across samples. In our dataset, we estimate this probability to be 0.56-3.99% (*Appendix A1: section S2*). As the top of this range assumes that all errors were due to tag jumping and therefore overestimates the problem, we assume that 1% of all reads for a prey taxon were erroneously assigned. We further assume that this 1% of tag-jumped reads is distributed evenly across spiders where this prey taxon was identified, and estimate a corrected number of reads κ_{ij} for each potential link between prey taxon i and spider j:

$$\kappa_{ij} = o_{ij} - 0.01 \times \frac{\tau_i}{\eta_{ji}}, \quad (1)$$

where o_{ij} is the original number of reads of prey taxon i recovered from spider j, τ_i is the total number of reads for prey i across all spiders, and η_i is the number of spiders from which prey i was recovered. If κ_{ij}<0, then we set $\kappa_{ij} = 0$. If all reads of prey i were recovered from a single spider (η_i=1), then we concluded that tag-jumping had not occurred and did not apply the correction.

We constructed networks for each site including all links where κ_{ij}>0. We used these "dynamic-cutoff" networks as the raw material when constructing Bayesian networks (described below). Within this framework, the tag-jumping correction can be viewed as part of the prior information used to estimate the true structure of the network.

Bayesian network assembly

Constructing a Bayesian network involves three steps: (1) identifying a suitable prior distribution, (2) modifying the prior with observed data, and (3) assembling networks based on posterior likelihoods of each interaction. The prior distribution describes the baseline probability of an interaction between an arbitrary prey and predator, and may be based upon data from similar published systems or from the focal system itself if a well-studied analogue is not available (Cirtwill et al., 2019). Posterior distributions are created by combining this prior distribution with the observed data for each prey and predator. These posterior distributions allow us to calculate the posterior probability of each interaction and select which interactions to include in the network (*Appendix A1: sections S3−S5*).

Identifying a prior

We based our prior on connectance: the overall probability that any given predator and prey will interact (Dunne et al., 2002). Connectance is calculated as the number of observed links divided by the product of the numbers of spiders and prey taxa. This prior assumes that, in general, spiders tend to consume similar numbers of prey. As the prey community (and numbers of taxa consumed per spider) likely varies between sites, we constructed separate priors for each site.

To obtain a distribution of interaction probabilities which does not depend upon a particular choice of cutoff, we calculated the connectance of the network for each site constructed using each cutoff value (101 networks per site; *Appendix A1: sections S3*). Connectance ranges between 0 (no species interacts with any other) and 1 (all species interact with all others). Our connectance-based prior distribution for the probability λ_{ij} that spider group i eats prey j at site k therefore follows a beta distribution:

$$\lambda_{ijk} \sim Beta(\alpha, \beta), \quad (2)$$

which is defined by two shape parameters, α and β. We determined maximum-likelihood estimates of these parameters using the R (R Core Team, 2016) function "fitdist" from the package *fitdistrplus* (Delignette-Muller and Dutang, 2015).

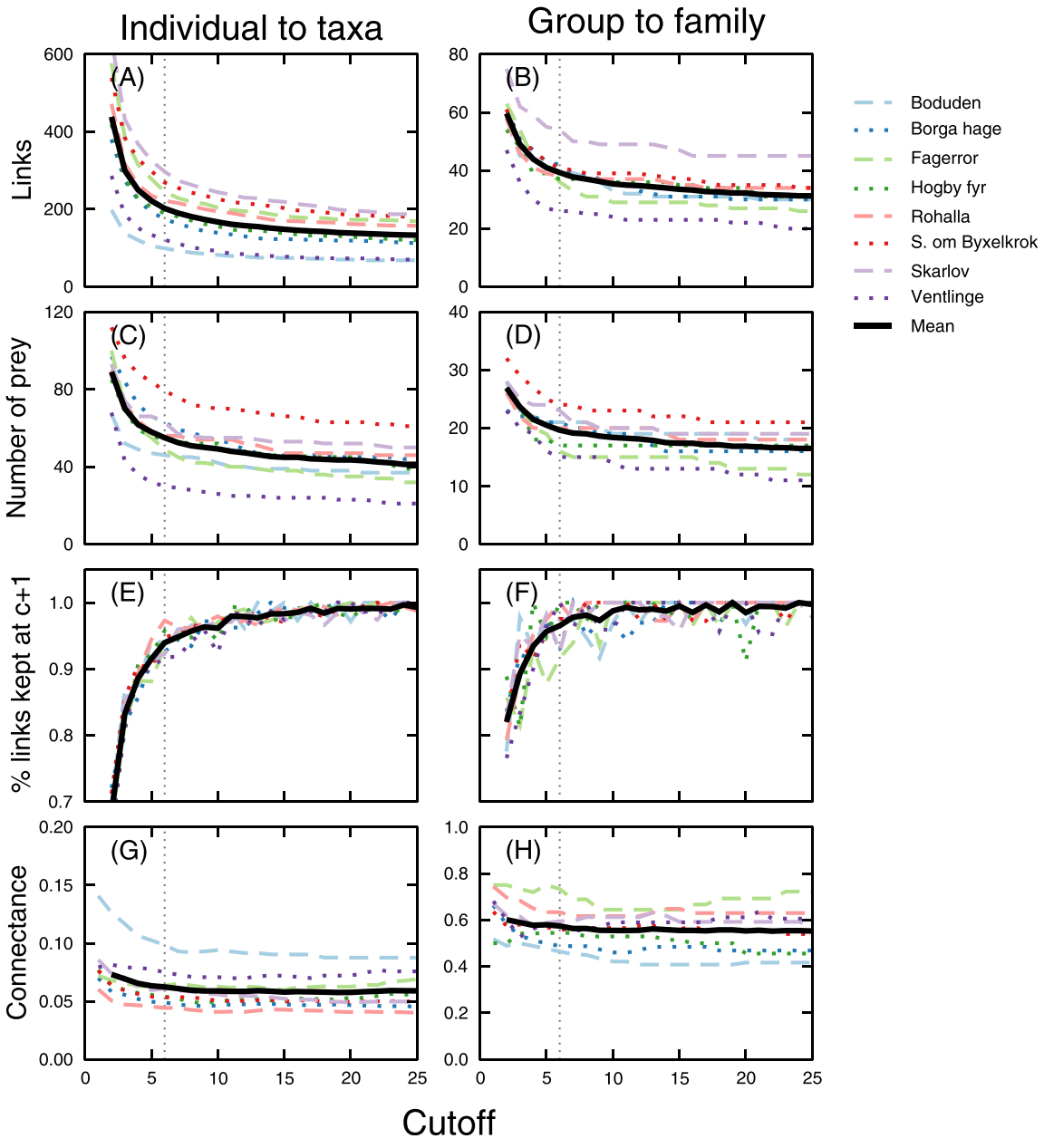

Fig. 1. Choice of cutoff affects the structure of fixed-cutoff networks by creating a trade-off between including true links and excluding false positives. This trade-off can be seen both in networks describing particular sites (broken, coloured lines) and in the mean across sites (thick, solid line). Although the cutoff of $c = 5$ in Verschut et al. (2019) (equivalent to c=6 in our framework; indicated by a dotted vertical line) was chosen by examining the properties of networks connecting spider age groups to prey families, networks connecting individual spiders to prey taxa respond similarly. As the cutoff threshold increases from one to six, **(A and B)** the number of links per site and **(C and D)** the number of prey families per site decrease rapidly while **(E and F)** the proportion of links included at cutoff c which are also included at cutoff $c + 1$ increases rapidly. As the cutoff threshold increases beyond six, the change in each of these properties is slower. **(G and H)** Connectance, a measure of the density of links within each site, changed very little over different cutoff thresholds as both links and nodes (spiders with no links supported by reads $>c$) can be removed with increasing cutoffs. Since connectance is robust to the choice of cutoff, we use the distribution of connectances across a broad range of cutoffs as the basis for our Bayesian prior distribution.

Incorporating observed data & constructing networks

The prior distribution gives the overall probability that an arbitrary spider from any age group has consumed an arbitrary prey. Our next step is to incorporate the observed data and obtain a posterior distribution of how likely it is that a given spider actually eats each observed prey, given the overall connectance of the network and the set of DNA sequences recovered from each spider. For an individual spider j and prey taxon i, consider the total number of identified sequence reads minus the tag-jumping correction (k_{ij}) as the

number of trials in a Bernoulli distribution (n_j). This is the maximum number of reads from spider j which could correspond to prey i after correcting for tag-jumping. These observations give the maximum likelihood estimates (MLEs) of the posterior parameters: $\alpha'_{ij} = \alpha + k_{ij}$ and $\beta'_{ij} = \beta + n_j - k_{ij}$. The posterior parameters in turn give the MLE for the probability of an interaction between i and j:

$$MLE = \frac{\alpha + k_{ij}}{\alpha + \beta + n_j}. \quad (3)$$

For spiders with many reads, the posterior MLE depends mostly on the observed data. For spiders with few reads, the MLE depends mainly on the connectance of the entire network. In effect, we use information from spiders with many reads to indicate the number of prey consumed by spiders with few reads. We constructed Bayesian networks for each site including all interactions with a threshold probability >0.01 (*Appendix A1: section S4*). In order to confirm that the chosen prior does not constrain the posterior networks to unreasonable structures, we

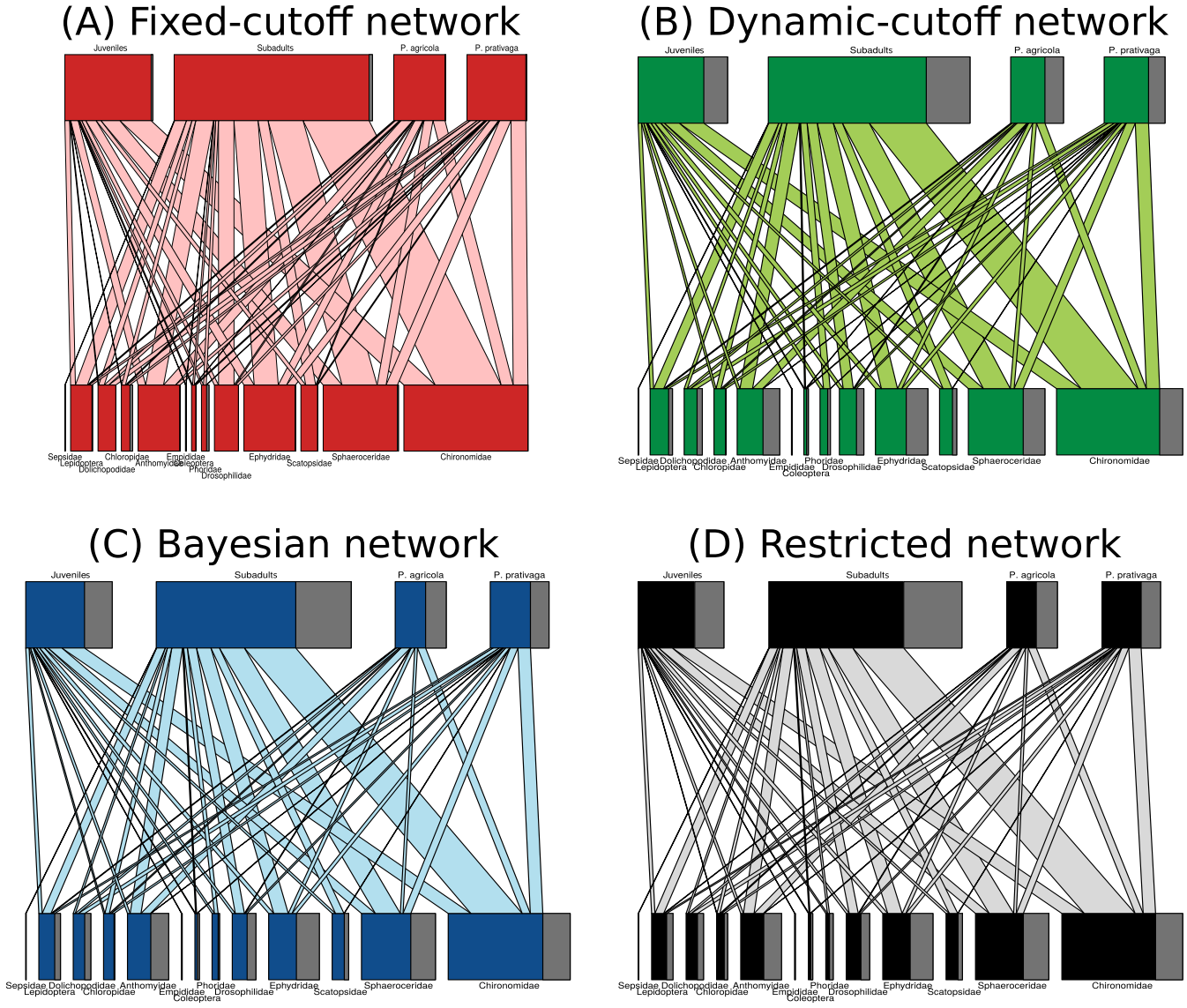

Fig. 2. When interactions are pooled to spider groups and prey families, both the fixed-cutoff, dynamic-cutoff (tag-jumping correction), and Bayesian approaches produce very similar networks. This leaves ecologists free to choose the approach which most clearly documents the assumptions used in network construction. Here we show food webs of *Pardosa* spiders assembled from **(A)** links included in the fixed-cutoff networks (red), **(B)** the dynamic-cutoff networks (green), **(C)** Bayesian networks (blue), and **(D)** networks including only links that were in both the fixed-cutoff and Bayesian networks ("restricted" networks; black). Note that all links in the Bayesian networks were also included in the dynamic-cutoff networks. Box widths reflect the total number of links observed each spider group or prey family; very rarely-consumed prey families have been removed for greater clarity. Light grey portions of boxed reflect links included in a network made by combining all network types but not included in the focal network. Networks have been scaled for constant widths of spider group boxes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

compared the posterior networks to networks simulated using the prior only (*Appendix A1: section S5*).

Comparison of network types

Having demonstrated the effect of including observed data on the Bayesian networks, we now compare the number of links included in different network types, as well as the distributions of these links between spider age groups. We were particularly interested in interactions included in only some network types (*Appendix A1: section S6*). As networks relating individual predators to prey taxa are the most closely-related to the original data and the most likely to highlight differences between the fixed-cutoff and Bayesian networks (because interactions are not pooled by predator group or prey family), we focus on these detailed networks rather than pooling predators or prey. See *Appendix A1: section S7* for a brief discussion of pooled networks.

Results

We compared networks using a fixed cutoff c=6 (Fig. 1) with networks which filter out links that are likely to be due to tag-jumping (dynamic-cutoff networks) and Bayesian networks which incorporate the tag-jumping filter, a connectance-based prior (with shape parameters for each site [see *Appendix A1: section S3, Table S3*]), and the observed data. The fixed-cutoff networks contained the most links and spiders, followed by the dynamic-cutoff networks (Fig. 2, *Appendix A1: section S8*). The mean number of links per spider also decreased from the fixed-cutoff networks to the dynamic-cutoff networks to the Bayesian networks (*Appendix A1: section S8*). The mean number of prey consumed per spider did not differ between age groups in the fixed-cutoff networks ($F_{3,471}$=2.34, p=0.072) but did differ between age groups in the dynamic-cutoff networks ($F_{3,472}$=3.32, p=0.020) and Bayesian networks ($F_{3,472}$=6.24, p<0.001). Higher-order network properties varied slightly between network types, with the fixed-cutoff networks generally being most distinct from the other network types (see *Appendix A1: section S8, Fig. S2*).

Links included in all network types tended to have higher numbers of reads than links included in only some networks types, although this was not significant for links included only in the dynamic-cutoff networks (Table 1A,B; Fig. 3A−C). Mean numbers of reads did not differ between links included in different subsets of networks (Table 1B). Links included in different network types also varied in both the total number of reads per spider and total number of reads per prey (Table 1, Fig. 3D−I). Links included in the Bayesian networks but not fixed-cutoff networks tended to involve prey with low total numbers of reads (and therefore low risk of tag-jumping) and spiders with low total numbers of reads. As the number of total reads per spider tended to be higher in adults of both species than in subadults and juveniles and total number of reads per prey varied across

Table 1. Mean numbers of reads, total reads per spider and total reads per prey differed for links included in the fixed-cutoff networks only, the dynamic-cutoff networks only, the dynamic-cutoff and Bayesian networks, and all networks. We give **(A)** mean values for each set of links and **(B-D)** p-values for Tukey's honest significant difference tests comparing means across groups. Note that all links included in the Bayesian networks were also included in the dynamic-cutoff networks.

(A) Mean values

Network type	Reads	Total reads per spider	Total reads per prey
Fixed-cutoff	11.9	21,589	6.00×10^5
Dynamic-cutoff	2.00	61,945	8.12×10^2
Bayesian & dynamic-cutoff	2.71	5547	6.44×10^2
All	6957	15,062	4.42×10^5

(B) Number of reads per link

Network type	Dynamic-cutoff only	Dynamic-cutoff & Bayesian	All
Fixed-cutoff only	<0.001	<0.001	<0.001
Dynamic-cutoff only	-	<0.001	0.151
Dynamic-cutoff & Bayesian	-	-	0.019

(C) Total number of reads per spider

Network type	Dynamic-cutoff only	Dynamic-cutoff & Bayesian	All
Fixed-cutoff only	<0.001	0.001	<0.001
Dynamic-cutoff only	-	<0.001	<0.001
Dynamic-cutoff & Bayesian	-	-	0.106

(D) Total number of reads per prey

Network type	Dynamic-cutoff only	Dynamic-cutoff & Bayesian	All
Fixed-cutoff only	<0.001	<0.001	<0.001
Dynamic-cutoff only	-	>0.999	<0.001
Dynamic-cutoff & Bayesian	-	-	<0.001

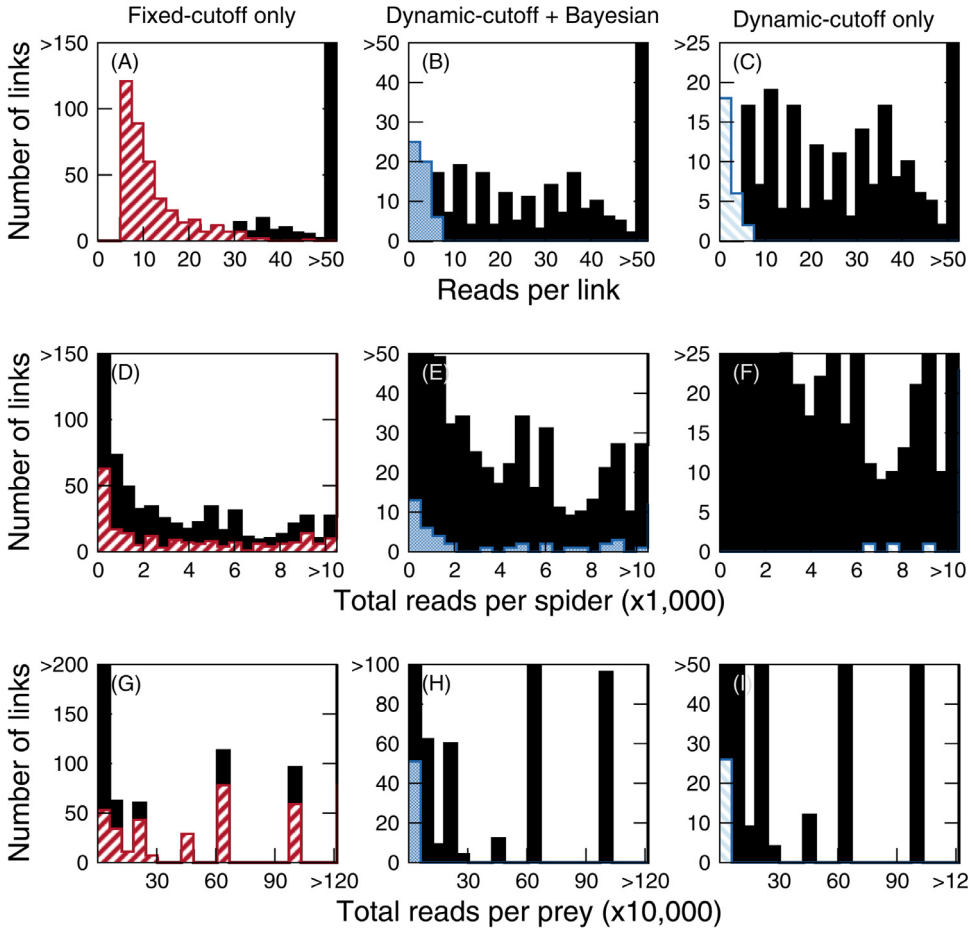

Fig. 3. Links included in only one type of network linking individual spiders to prey taxa tended to have different properties than links included in all networks. **(A-C)** Links included in all network types (solid black histograms) were usually represented by more reads than links included in the fixed networks only, dynamic-cutoff and Bayesian networks, or dynamic-cutoff networks only (striped red, spotted blue, or striped blue histogram, respectively). **(D–F)** Links included in the fixed-cutoff networks only came from spiders with a similar distribution of total reads as links included in all network types. Links included in the dynamic-cutoff and Bayesian networks were more likely to come from spiders with higher total numbers of reads, while links included in the dynamic-cutoff networks only came from spiders with very high numbers of total reads. **(G–I)** Links included in the fixed-cutoff networks only involved prey with a similar distribution to total reads as links involved in all network types. Links included in the dynamic-cutoff and Bayesian networks or dynamic-cutoff networks only tended to involve prey with especially low total numbers of reads. In all panels, the last bin collects all links or spiders with numbers of reads beyond the limit of the x-axis. We truncate these axes to improve clarity near the origin. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

families, these differences have consequences for the distribution of links across prey and predator groups. Links included only in one network type (e.g., fixed-cutoff networks only) were not distributed evenly across spider groups ($\chi^2_{df=9}$=36.4, p<0.001; p-value based on simulations) or prey families ($\chi^2_{df=108}$=621, p<0.001; p-value based on simulations).

Despite these minor differences, the three network types had very similar structures overall (Fig. 2). The majority of links were included in all network types, indicating that all approaches identified the same core network structure. The links not included in all network types tended to be supported by few reads, involve prey with few total reads, and/or come from spiders with few total reads. These links are therefore the most susceptible to various errors in the metabarcoding process and should be treated with caution when analysing higher-order network properties.

Discussion

DNA metabarcoding of gut contents or faeces has allowed researchers to obtain dietary information for groups such as spiders for which this information was previously difficult to access (Eitzinger et al., 2019; Waldner and Traugott, 2012). This new diet information, however, comes with the risk of false positives (mis-identified "prey" which are not actually eaten) (Alberdi et al., 2018; Roslin and Majaneva, 2016). False positives can arise, for example, through contamination during sampling or laboratory analysis, through

the formation of chimeric DNA during PCR replication, or due to tag jumping (Alberdi et al., 2018; Schnell et al., 2015). The typical strategy for limiting false positives is to eliminate prey supported by a small number of sequence reads, but the appropriate threshold for this filtering is not clear (Alberdi et al., 2018). Here, we re-frame the problem from identifying a number of reads which distinguishes between "true" and "false" prey to calculating the probability that each taxon identified in gut contents is a true prey. Any prey with a probability above a reasonable threshold is included in the network.

The Bayesian approach makes better use of the available data by using what information each spider provides, together with the prior, to estimate the most likely prey for each individual. Most of the links included in the Bayesian networks but not the fixed-cutoff networks involve spiders with few total reads. In these cases, few links meet the fixed threshold while the Bayesian networks are able to use prior information about connectance and the risk of tag jumping for different prey to provide a "best guess" for the diets of these individuals. The Bayesian framework's inclusion of predator and prey properties is analogous to graph-pruning algorithms which include or exclude links based on node properties as well as link strengths (Dianati, 2016; Radicchi et al., 2011; Zhou et al., 2012). These approaches acknowledge that filtering links only based on strength can eliminate low-strength structures of interest (Dianati, 2016). Given the similarities across network types, we recommend that researchers use the Bayesian approach because it obliges researchers to make their prior assumptions about the network explicit− and therefore open to validation or rejection) in a way that choosing fixed cutoffs does not (Spiegelhalter et al., 2000), and because it addresses the potential for errors that are not due to tag-jumping.

The Bayesian approach for network construction can be generalised beyond predator-prey networks to constructing host-parasite networks based on faecal DNA, plant-pollinator networks based on pollen DNA, etc. In systems where more information about the system is available, more complex priors can be constructed in order to generate more realistic posterior networks. For example, if varying amplification rates for different prey are known (Liu et al., 2019), this information could be used to build a prior which counteracts these biases. Prey abundances can also be used to inform a prior (Graham and Weinstein, 2018; Weinstein and Graham, 2017a), unless the aim of the study is to test predator selectivity (Cirtwill et al., 2019; Weinstein and Graham, 2017b). Other traits such as body mass can also be used to inform the prior, as long as these traits are known for the life stages that are most likely to interact (Brose, 2010; Brose et al., 2006; Gravel et al., 2016; Portalier et al., 2019). Finally, network-filtering algorithms could be borrowed from the graph theory literature (e.g., Dianati, 2016; Zhou et al., 2012), especially in systems where number of reads is strongly correlated with the strength of an interaction. All of these informative priors have the potential to give more realistic estimates of the true network structure. The strong assumptions used to build these priors also carry a risk of biasing the posterior networks against including 'unexpected' links which do not fit these assumptions. For example, a strong prior based on body mass ratios may exclude links involving predators with feeding strategies that allow them to consume unusually large or small prey, especially if the DNA of these prey does not amplify well with the chosen primer. The appropriate balance between making reasonable assumptions and avoiding imposing constraints on the final network will depend on the circumstances of each study and deserves careful consideration during experimental design.

In closing, it is important to remember that all methods of diet reconstruction or network construction are somewhat arbitrary. When using gut contents to establish interactions, researchers must choose which prey to focus on (and a corresponding suitable primer), assess the risk of including "prey" which were derived from the guts of predators consumed by the focal species, and determine which interactions are supported well enough to be included. In cutoff-based approaches, this is done by setting a threshold number or proportion of reads. In Bayesian approaches, this is done by setting a threshold probability of interaction. In either case, researchers must justify these decisions and support them as well as possible. By making the decision-making process more explicit, we can facilitate comparisons of results between networks constructed using slightly different methods and, as a field, develop best practices for building networks using gut content DNA.

Declaration of Competing Interest

This research was funded by a Stockholm University Department of Ecology, Evolution, and Plant Sciences departmental grant to Alyssa R. Cirtwill. Neither the department nor Stockholm University had any role in designing the study.

Acknowledgments

Funding: This work was supported by a Stockholm University departmental grant. Stockholm University had no involvement in the study.

Supplementary material

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.baae.2020.11.007

References

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable

metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*, 134–147. doi:10.1111/2041-210X.12849.

Arrizabalaga-Escudero, A., Clare, E. L., Salsamendi, E., Alberdi, A., Garin, I., Aihartza, J., & Goiti, U. (2018). Assessing niche partitioning of co-occurring sibling bat species by DNA metabarcoding. *Molecular Ecology*, *27*, 1273–1283. doi:10.1111/mec.14508.

Bell, K. L., Fowler, J., Burgess, K. S., Dobbs, E. K., Gruenewald, D., Lawley, B., Morozumi, C., & Brosi, B. J. (2017). Applying pollen DNA metabarcoding to the study of plant-pollinator interactions. *Applications in Plant Sciences*, *5*, 1600124. doi:10.3732/apps.1600124.

Bowen, W. D., & Iverson, S. J. (2013). Methods of estimating marine mammal diets: A review of validation experiments and sources of bias and uncertainty. *Marine Mammal Science*, *29*, 719–754. doi:10.1111/j.1748-7692.2012.00604.x.

Boyer, S., Wratten, S. D., Holyoake, A., Abdelkrim, J., & Cruickshank, R. H. (2013). Using next-generation sequencing to analyse the diet of a highly endangered land snail (*powelliphanta augusta*) feeding on endemic earthworms. *PLoS ONE*, *8*, e75962. doi:10.1371/journal.pone.0075962.

Brose, U. (2010). Body-mass constraints on foraging behaviour determine population and food-web dynamics. *Functional Ecology*, *24*, 28–34. doi:10.1111/j.1365-2435.2009.01618.x.

Brose, U., Jonsson, T., Berlow, E. L., Warren, P., Banasek-Richter, C., Bersier, L. F., Blanchard, J. L., Brey, T., Carpenter, S. R., Blandenier, M. F. C., Cushing, L., Dawah, H.a., Dell, T., Edwards, F., Harper-Smith, S., Jacob, U., Ledger, M. E., Martinez, N. D., Memmott, J., Mintenbeck, K., Pinnegar, J. K., Rall, B. C., Rayner, T. S., Reuman, D. C., Ruess, L., Ulrich, W., Williams, R. J., Woodward, G., & Cohen, J. E. (2006). Consumer-resource body-size relationships in natural food webs. *Ecology*, *87*, 2411–2417. doi:10.1890/0012-9658(2006)87[2411:CBRINF]2.0.CO;2.

Cirtwill, A. R., Eklöf, A., Roslin, T., Wootton, K., & Gravel, D. (2019). A quantitative framework for investigating the reliability of empirical network construction. *Methods in Ecology and Evolution*, *10*, 902–911. doi:10.1111/2041-210X.13180.

Dalén, L., Götherström, A., & Angerbjörn, A. (2004). Identifying species from pieces of faeces. *Conservation Genetics*, *5*, 109–111. doi:10.1023/B:COGE.0000014060.54070.45.

Deagle, B. E., Thomas, A. C., McInnes, J. C., Clarke, L. J., Vesterinen, E. J., Clare, E. L., Kartzinel, T. R., & Eveson, J. P. (2019). Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular Ecology*, *28*, 391–406. doi:10.1111/mec.14734.

Delignette-Muller, M. L., & Dutang, C. (2015). Fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, *64*, 1–34. doi:10.18637/jss.v064.i04.

Dianati, N. (2016). Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Physical Review E*, *93*. doi:10.1103/PhysRevE.93.012304.

Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002). Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters*, *5*, 558–567. doi:10.1046/j.1461-0248.2002.00354.x.

Eitzinger, B., Abrego, N., Gravel, D., Huotari, T., Vesterinen, E. J., & Roslin, T. (2019). Assessing changes in arthropod predator-prey interactions through DNA-based gut content analysis-variable environment, stable diet. *Molecular Ecology*, *28*, 266–280. doi:10.1111/mec.14872.

Graham, C. H., & Weinstein, B. G. (2018). Towards a predictive model of species interaction beta diversity. *Ecology letters*, *21*, 1299–1310. doi:10.1111/ele.13084.

Gravel, D., Albouy, C., & Thuiller, W. (2016). The meaning of functional trait composition of food webs for ecosystem functioning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*, 20150268. doi:10.1098/rstb.2015.0268.

Jirků, M., Pomajbíková, K., Petrželková, K. J., Hůzová, Z., Modrý, D., & Lukeš, J. (2012). Detection of *plasmodium* spp. in human feces. *Emerging Infectious Diseases*, *18*, 634–636. doi:10.3201/eid1804.110984.

Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, *30*, 1883–1893. doi:10.1111/1365-2435.12763.

Kitson, J. J. N., Hahn, C., Sands, R. J., Straw, N. A., Evans, D. M., & Lunt, D. H. (2019). Detecting host-parasitoid interactions in an invasive lepidopteran using nested tagging dna metabarcoding. *Molecular Ecology*, *28*, 471–483. doi:10.1111/mec.14518.

Lai, S. M., Liu, W. C., & Jordán, F. (2012). On the centrality and uniqueness of species from the network perspective. *Biology Letters*, *8*, 570–573. doi:10.1098/rsbl.2011.1167.

Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J., & Burridge, C. P. (2019). A practical guide to DNA metabarcoding for. *Ecological Entomology*. doi:10.1111/een.12831.

Makiola, A., Dickie, I. A., Holdaway, R. J., Wood, J. R., Orwin, K. H., Lee, C. K., & Glare, T. R. (2019). Biases in the metabarcoding of plant pathogens using rust fungi as a model system. *MicrobiologyOpen*, *8*, e780. doi:10.1002/mbo3.780.

Matesanz, S., Pescador, D. S., Pías, B., Sánchez, A. M., Chacón-Labella, J., Illuminati, A., de la Cruz, M., López-Angulo, J., Marí-Mena, N., Vizcaíno, A., & Escudero, A. (2019). Estimating belowground plant abundance with DNA metabarcoding. *Molecular Ecology*, *19*. doi:10.1111/1755-0998.13049.

Mathieu, C., Hermans, S. M., Lear, G., Buckley, T. R., Lee, K. C., & Buckley, H. L. (2020). A systematic review of sources of variability and uncertainty in eDNA data for environmental monitoring. *Frontiers in Ecology and Evolution*, *8*, 135. doi:10.3389/fevo.2020.00135.

Pafčo, B., Čížková, D., Kreisinger, J., Hasegawa, H., Vallo, P., Shutt, K., Todd, A., Petrželková, K. J., & Modrý, D. (2018). Metabarcoding analysis of strongylid nematode diversity in two sympatric primate species. *Scientific Reports*, *8*, 5933. doi:10.1038/s41598-018-24126-3.

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

Portalier, S. M., Fussmann, G. F., Loreau, M., & Cherif, M. (2019). The mechanics of predator-prey interactions: First principles of physics predict predator-prey size ratios. *Functional Ecology*, *33*, 323–334. doi:10.1111/1365-2435.13254.

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. doi:10.1017/CBO9781107415324.004.

Radicchi, F., Ramasco, J. J., & Fortunato, S. (2011). Information filtering in complex weighted networks. *Physical Review E*, *83*, 046101. doi:10.1103/PhysRevE.83.046101.

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The barcode of life data system. *Molecular Ecology Notes*, *7*, 355–364. doi:10.1111/j.1471-8286.2006.01678.x.

Roslin, T., & Majaneva, S. (2016). The use of DNA barcodes in food web construction-terrestrial and aquatic ecologists unite!. *Genome*, *59*, 603–628. doi:10.1139/gen-2015-0229.

Roubinet, E., Birkhofer, K., Gerard, M., Karin, S., Ekboom, B., Michael, T., & Mattias, J. (2017). Diet of generalist predators reflects effects of cropping period and farming system on extra- and intraguild prey. *Ecological Applications*, *27*, 1167–1177. doi:10.1002/eap.1510.

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, *15*, 1289–1303. doi:10.1111/1755-0998.12402.

Spiegelhalter, D. J., Myles, J. P., Jones, D. R., & Abrams, K. R. (2000). Bayesian methods in health technology assessment: A review. *Health Technology Assessment*, *4*, 1–130. doi:10.3310/hta4380.

Staudacher, K., Jonsson, M., & Traugott, M. (2016). Diagnostic PCR assays to unravel food web interactions in cereal crops with focus on biological control of aphids. *Journal of Pest Science*, *89*, 281–293. doi:10.1007/s10340-015-0685-8.

Verschut, V., Strandmark, A., Esparza-Salas, R., & Hambäck, P. A. (2019). Seasonally varying marine influences on the coastal ecosystem detected through molecular gut analysis. *Molecular Ecology*, *28*, 307–317. doi:10.1111/mec.14830.

Waldner, T., & Traugott, M. (2012). DNA-Based analysis of regurgitates: A noninvasive approach to examine the diet of invertebrate consumers. *Molecular Ecology Resources*, *12*, 669–675. doi:10.1111/j.1755-0998.2012.03135.x.

Weinstein, B. G., & Graham, C. H. (2017). On comparing traits and abundance for predicting species interactions with imperfect detection. *Food Webs*, *11*, 17–25. doi:10.1016/j.fooweb.2017.05.002.

Weinstein, B. G., & Graham, C. H. (2017). Persistent bill and corolla matching despite shifting temporal resources in tropical hummingbird-plant interactions. *Ecology Letters*, *20*, 326–335. doi:10.1111/ele.12730.

Wirta, H. K., Hebert, P. D. N., Kaartinen, R., Prosser, S. W., Várkonyi, G., & Roslin, T. (2014). Complementary molecular information changes our perception of food web structure. *Proceedings of the National Academy of Sciences*, *111*, 1885–1890. doi:10.1073/pnas.1316990111.

Zhou, F., Mahler, S., & Toivonen, H. (2012). Simplification of networks by edge pruning. *Bisociative knowledge discovery* (pp. 179–198). Berlin: Springer.