OIKOS

Research

Methodological overview and data-merging approaches in the study of plant–frugivore interactions

Elena Quintero, Jorge Isla and Pedro Jordano

E. Quintero (https://orcid.org/0000-0003-4979-6874) ✉ (equintero@ebd.csic.es), J. Isla (https://orcid.org/0000-0002-2307-9730) and P. Jordano (https://orcid.org/0000-0003-2142-9116), Estación Biológica de Doñana, CSIC, Sevilla, Spain. PJ also at: Dept Biología Vegetal y Ecología, Univ. de Sevilla, Sevilla, Spain.

Oikos
00: 1–18, 2021
doi: 10.1111/oik.08379

Subject Editor: Eliana Cazetta
Editor-in-Chief: Pedro Peres-Neto
Accepted 10 June 2021

Recording species interactions is one of the main challenges in ecological studies. Frugivory has received much attention for decades as a model for mutualisms among free-living species, and a variety of methods have been designed and developed for sampling and monitoring plant–frugivore interactions. The diversity of techniques poses an important challenge when comparing, combining or replicating results from different sources with different methodologies. With the emergence of modern techniques, such as molecular analysis or multimedia remote recorders, issues when combining data from different sources have become especially relevant. We provide an overview of all the techniques used for monitoring endozoochorous primary seed dispersal, focusing on a critical appraisal of the advantages and limitations, as well as the context-dependency nature, of the different methods. We propose five data merging approaches potentially useful to combine frugivory interactions data from different methodologies. Additionally, we provide two case studies where we combine empirical data from plant–animal interactions in Mediterranean shrublands using different methodologies. Data merging resulted in a net increase in the number of distinct pairwise interactions recorded and compensated biases inherent to different methods, resulting in a more robust estimation of network topological descriptors. These case studies clarify the context-dependent character of the merging approaches, highlighting the value of collecting detailed information on the sampling effort in terms of reliable results and reproducibility. Finally, we discuss the trends with different methodological approaches used in the last decades and future perspectives in this field.

Keywords: ecological networks, endozoochory, frugivory, methods, plant–animal interactions, seed dispersal

Introduction

One of the greatest challenges that ecologists face is to properly determine the biodiversity present in their study systems, i.e. the presence and relative abundance of species (Magurran 1988). An important facet, yet a frequently dismissed one, in biodiversity analysis is to document how species interact with one another, and what are the outcomes of these interactions (Valiente-Banuet et al. 2014). Scholar accounts

NORDIC SOCIETY OIKOS

www.oikosjournal.org

of the myriad connections among species date back at least to al-Jāhiz in the 9th century or even earlier to Aristotle in the 4th century BCE (Egerton, 2007). Yet, the more formal onset of the ecology of interactions took place later, fostered by late 18th-century naturalists. Pioneer studies of ecological interactions were focused on trophic cascades within food webs (Cohen 1978, Polis and Strong 1996), and later unfolded into the analysis of complex networks of ecological interactions in the late 1990s (Bascompte and Jordano 2014). Effectively incorporating the quantification and analysis of ecological interactions is essential to recent efforts to preserve the value of Biodiversity (IPBES 2019) yet we are still far from achieving this goal, not only by assessing the actual richness and diversity of interactions in nature, but also by assessing the ecological services associated to them.

Frugivory has received much attention for decades and a variety of methods have been designed and developed to track how encounters between animal frugivores and plants result in seed dispersal events for the plants and food resource provisioning for the animals (Estrada and Fleming 1986, Fleming and Estrada 1993, Levey et al. 2002, Dennis et al. 2007). This reciprocal service is the basis of coevolved plant–frugivore mutualistic interactions and implies enormous consequences for forest regeneration and ecosystem functioning (Howe and Smallwood 1982, Schupp et al. 2002). A crucial aspect of interaction sampling, besides recording the mere presence of an interaction, is also measuring its relative frequency and its impact, i.e. the outcome of interactions in terms of fitness effects for the interacting partners.

The study and monitoring of seed dispersal events became increasingly apparent from late eighties with the publication of the first volume of FSD (Estrada and Fleming 1986), and direct observation and census at focal plants became a standard method to inventory plant–frugivore interactions with multiple objectives. Yet, new methods have emerged in the last decades allowing indirect, delayed recording of these interactions and opening new possibilities for research on frugivory and seed dispersal (Forget and Wenny 2005, Carlo et al. 2009, González-Varo et al. 2014). The diversity of techniques available to monitor species interactions pose the important challenge of comparing results obtained with different methodologies, replicating the results or incorporating interaction data from different sources. With the emergence of modern techniques, such as DNA-based molecular analysis (Carreon-Martinez and Heath 2009, Valentini et al. 2009, González-Varo et al. 2017, Mata et al. 2019), this has become especially relevant. Merging data from different sources allows us to maximize information and improve research potential for any kind of frugivory and seed dispersal study. Combining the distinct data types and information yielded by such a diversity of methods can become a difficulty and even a limitation if there are no well-established guidelines.

Given the wide spectrum of seed dispersal interactions that exist, in the first part of this manuscript we provide a methodological overview where we primarily focus on endozoochorous seed dispersal. Our goal in this part is not a comprehensive review of methods, rather we aim to offer a critical appraisal of the advantages and limitations as well as the context-dependent nature of the major sampling methods, focusing on methods complementarity, reproducibility and sampling effort. In the second part of this manuscript we propose and illustrate five different merging approaches to combine datasets originated with different sampling methodologies. The merging data approaches we describe here may be also applicable to other interaction forms aside endozoochorous seed dispersal, such as synzoochory, epizoochory or secondary seed dispersal (Costa et al. 2014, Gómez et al. 2019) or even other types of interactions like pollination, host–parasite or plant–plant facilitation. To exemplify and validate the described merging methods we provide two case studies, using empirical data where we compare and combine different methodologies using an interaction network approach. Finally, we discuss the trends in the use of different approaches over the last decades and the future perspectives in this field. We hope that this overview and the combination strategies proposed here can serve as a useful reference for researchers when approaching future frugivory studies and may complement other papers in this Special Issue dealing with plant–frugivore interactions and thorough field-sampling approaches.

Study focus, scale and resolution

Depending on the study's objective, the term 'interaction' and its measurement can vary greatly. The strength of an interaction (i.e. the effect magnitude of its outcome, in addition to its frequency of occurrence) can change depending on the focus of study and how its outcome for the partners is measured. Therefore, the study question will determine when and how interactions are monitored (Niquil et al. 2020).

Focus may be directed towards the plant partner (i.e. phytocentric), the animal partner (i.e. zoocentric) or both (Jordano 2016). These approaches impose different sampling challenges and information, varying in their characteristics, accessibility, visibility, potential biases, logistic limitations, sampling effort demand, etc. In many cases the goals themselves can clearly establish the characteristics of the study focus (e.g. understand the role in seed dispersal of juvenile versus adult individual animal frugivores, González-Varo et al. (2019). However, sometimes the focus of study may be more complex to define (e.g. select a phytocentric or zoocentric approach when comparing seed dispersal networks between sites).

Other important aspects include the resolution and scale of the sampling, being the intra-individual level the most refined, and scaling up to the aggregation of species in groups, eventually including higher taxonomic levels, morphological or functional groups (Moran et al. 2004). Clearly defining the spatial and temporal scale of the study is key. Plant–frugivore interaction patterns at different spatial scales are not necessarily consistent (Jordano 1993, García et al. 2004), furthermore it becomes extremely difficult to extend analyses of e.g. dispersal kernels, beyond the local scales

(García and Borda-de-Água 2017). Likewise, temporal variations driven by the phenology of the species or the availability of resources will largely determine the interactions detected (Carnicer et al. 2009, Costa et al. 2020).

When combining studies or methodologies, it is advisable to look at the study focus and at the scale at which each data source has been gathered. At the end of the combining methodologies section we propose a way of correcting the divergence that may exist between scales. Also, one of the case studies illustrates an example of data merging for two methodologies with different focus; observations of foraging animals, as plant-focused and mist-netting, as animal-focused.

A general overview of sampling methods

In order to illustrate current methodologies, we will divide sampling techniques into three intuitive categories based on the main stages of the seed dispersal process (Schupp et al. 2017). Depending on when we are collecting information for animal–plant interactions, the sampling will be directed towards one of these three stages (Fig. 1): 'visitation', 'transport' or 'deposition'. While some methods may be directed to more than one stage, we have classified them in the most representative one. The first, or early stage ('visitation'), refers to the actual interaction on the plant, when the animal is manipulating, removing or ingesting the fruits on the plant. The subsequent two stages refer to the dissemination process, where the seeds are first transported (i.e. moved some distance away from the source plant, 'transport') and then deposited (i.e. disseminated), which may involve actual burying of the seed or just dropping, e.g. by spitting, regurgitation or defecation ('dissemination').

Methods targeting 'visitation'

Methods used to monitor the initial 'visitation' stage (Fig. 1, Table 1) are typically directed towards seed sources (i.e. maternal) plant individuals with standing fruit crops, where it is possible to observe the interaction occurring. When considering the resource-harvest type of mutualism characteristic to most plant–frugivore interactions (Janzen 1983, Ollerton 2006), this typically refers to the feeding phase, when the source plant and individual animal actually interact. The methods used in this stage can be classified into direct methods that allow us to observe the interaction directly, and indirect methods based on the detectable signals of the interaction. Traditional methods are field focal observations at fruiting plants (Howe and Vande Kerckhove 1980, Herrera and Jordano 1981, Snow and Snow 1988, Jordano and Schupp 2000, Stevenson et al 2015), transects (Galetti and Pizo 1996), animal visual trackings (Gestich et al. 2019) and spot censuses (Howe and Vande Kerckhove 1981, Rother et al. 2015). The implementation of technological advances such as camera traps or other multimedia recording systems (e.g. action cameras) also allows us to observe the interaction taking place (Campos-Arceiz et al. 2012, Miguel et al. 2018). These non-invasive multimedia techniques avoid the interference of

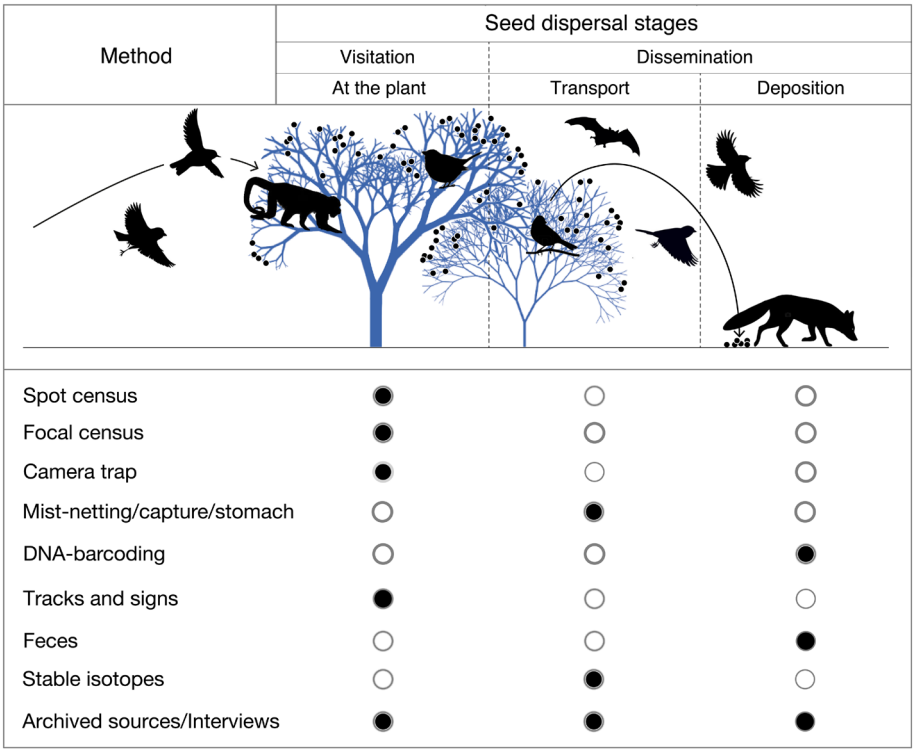

Figure 1. A non-exhaustive overview of the most frequently used methods for recording plant–frugivore interactions. Sampling methods are classified according to the seed dispersal stage in which they are most frequently applied: 'visitation', 'transport' and 'deposition'.

Table 1. A summary of sampling methods used in plant–frugivore interaction studies. Including information about: seed dispersal stage, sampling effort, costs, detectability, replicability, additional information that can be obtained (apart from the interaction per se), units of measurement and reference studies illustrating the usage of each method. See the Supporting information for the full reference list of the example studies.

Method	Stage	Sampling effort	Cost	Detectability	Replicability	Additional information	Units	Example studies
Direct observations	Visitation	Requires long time in the field.	Low. High cost in sampling time.	Medium–low (human presence can affect significantly).	Limited – affected by the expertise of the observer.	Foraging behaviour, fruit handling behaviour.	Frequency (e.g. visits/h and/or feeding rate) Total count	1, 2, 3
Spot censuses	Visitation	Requires long time in the field.	Low. High cost in sampling time.	Medium-low (human presence can affect significantly).	Limited – affected by the expertise of the observer.	Foraging behaviour.	Frequency (e.g. interactions recorded/area/h) Total count	4, 5 ,6
Animal followings	Visitation	Requires long time in the field.	Low. High cost in sampling time.	Medium-low (human presence can affect significantly).	Limited – affected by the expertise of the observer.	Foraging behaviour, fruit handling behaviour.	Frequency (e.g. interactions/ time, fruits/visit) Total count	27
Multimedia (e.g. phototrapping)	Visitation	Low in the field but remarkable processing video efforts.	Expensive at first (big inversion, then low - only batteries).	High for large-bodied animals. Low for small or fast animals.	Medium-high – can be challenging for high canopy plants.	Behaviour, n of fruits, feeding time, etc.	Frequency (e.g. visits/h). Total count.	7, 8
Footprint traps. Bill and teeth marks	Visitation	Low.	Low in material. High cost in sampling time.	Medium, useful for interactions with well-known signals.	Medium. Highly dependent on substrate type or in species signals.	Animal body size, traits of discarded fruits.	Frequency: visits/ time. Total counts: Number of contacted fruits.	9, 10
Fruit removal in the plant	Visitation	Low-mid: problematic if huge crop sizes.	Low in material. High cost in sampling time.	High with medium-low crop sizes. Subsampling could be useful.	Very system-dependent (e.g. unique seed disperser or species with different temporal activity patterns).	Temporal activity patterns.	Total counts: Fruit removal metrics.	11, 12
Offerings	Visitation	Low-mid: problematic if low crop size.	Low in material. High cost in sampling time.	Low specificity.	Medium-High.	Feeding preferences.	Frequency: visits/ time. Total counts: Fruit removal metrics.	13, 7
Stomach content (after dead)	Transport	High, needs large samples. Bio-ethical considerations.	Low in material. Medium-high cost in sampling time and processing.	High.	Low – depends on sample availability.	Animal phenotypic traits. Fruit preferences.	Frequency: seeds/ sample. Total counts: Number of fruits/seeds consumed.	14, 15

Method	Stage	Sampling effort	Cost	Detectability	Replicability	Additional information	Units	Example studies
Mist netting, traps	Transport	High-may need a long time to capture sufficiently-representative sample.	Mid-high.	Low – very selective method, elusive animals are hard to capture.	Mid-low.	Animal phenotypic traits. Supplementary diet resources.	Frequency: seeds/bird Number of seeds per stomach content or fruits consumed.	16, 17
Fecal samples or regurgitated/spat seeds (in the field)	Deposition	High, requires field work in addition to sample treatment and identification.	Low in material. High cost in sampling time and processing.	Medium. It is difficult to detect and identify samples of rare species.	High.	Micro-habitat deposition. Other food resources. Precise spatial reference. Fruit preferences and handling.	Seed shadow metrics. Frequency of seeds per faecal sample. Faeces or seeds per transect. No. fruits consumed.	18, 19
DNA-based molecular techniques (e.g. barcoding)	Deposition	Mid-Relatively short time in the field.	High both in materials and processing time.	High for rare species Low for species not adequately sampled in seed traps.	High-based on protocol.	Micro-habitat deposition. Other food resources. Precise spatial reference. Fruit preferences.	Interaction frequency. Seed shadow metrics. Frequency of seeds per faecal sample. No. dispersed seeds.	20, 21
Stable isotopes	Transport–deposition	High – may need a long time to capture sufficiently-representative sample.	High.	Limited, usually not species-specific detection.	High.	Possibilities for large-scale analysis. Dispersal distances.	Interaction frequency. Presence of plant material in diet.	22, 23
Interviews	Visitation–transport–deposition	Low.	Low cost in terms of material. High cost in time for interviews.	Medium–low (human presence can affect significantly-sometimes difficult to identify at species level).	Limited-Low. (High area-dependent effect).	Behaviour, no. of fruits, feeding time, long-term trends.	Number of positive responses for total interviews (total count).	24
Bibliographic review	Visitation–transport–deposition	Low, no field. Hard review and researchers contact work.	Limited material. High cost in time.	High, M&M dependent.	High.	Behaviour, no. of fruits, feeding time, frugivore community and plant resources availability.	Specific units of each study. Number of studies. Total interactions for all sources.	25, 26

the observer with the animal in the field, allow continuous sampling over day and night and extended periods of time, and enable simultaneous monitoring over large study areas, thus increasing the probability of detection of rare interactions and improving the description of interactions distribution. Other indirect methods such as footprint traps allow to identify the species of animal that visits the plant (Jácome-Flores et al. 2020), or offerings (Garrote et al. 2018) that allow to quantify frugivory rates, can also be very useful, since they do not require a continued presence nor entail a high economic cost. Bill and teeth marks are signals that can be used to infer interactions too (Alves-Costa and Lopes 2001). An alternative indirect method in this phase is the estimation of the fruit removal caused by frugivores by counting the plant crop size over consecutive periods of time. This method becomes useful for plants with one exclusive frugivore (i.e. exclusive frugivory on islands, Hansen and Traveset 2012; or cases of double mutualism, Hansen and Mueller 2009, Gomes et al. 2013) as well as to discern between guilds of daytime and night-time frugivores (Palmeirim et al. 1989, Korine et al. 2000).

Methods targeting 'transport'

Methodologies used during the 'transport' stage are typically those where the animal is intercepted by means of capture, before any kind of fruit or seed release or deposition has naturally taken place (Fig. 1, Table 1). The capture methods depend on the target species, for example the mist nets are the most commonly used for medium-small size birds and bats (Herrera 1984a, Costa et al. 2020). Live traps are used for mammals (Genrich et al. 2017), and for fish there are also capture methods that allow obtaining the stomach content of the captured animals (Weiss et al 2016). Other sources of information include the stomach contents of animals after death from directed hunting (Remsen et al 1993), fishing (Galetti et al. 2008, Correa et al. 2015) or roadkills (Vaz et al. 2012). Interactions can be quantified on the basis of the number of seeds of different species found in the feces or in their stomach contents, properly accounting for pulp and/or seed remains due to potential biases generated by differential gut treatment (Oliveira et al. 2002; Supporting information).

Methods targeting 'deposition'

Methods targeting the 'deposition' stage (Fig. 1, Table 1) are used when the seed has reached its final destination by means of defecation, regurgitation, spitting, scatter-hoarding, discarding or unfortunate drop. In this case, sampling is directed towards the final seed destination (except when secondary or subsequent dispersal events are involved, i.e. re-caching), typically in scats or droppings. Collection of fecal samples or regurgitated seeds can be carried out in different ways. For example, using transects or established areas to collect samples in the field has been widely used to study seed dispersal by mammals (González-Varo et al. 2013, Perea et al. 2013). In the case of primates, continuous monitoring of individuals can prove to be useful (Gestich et al. 2019). For bird dispersal, the use of seed traps is more common, since it greatly facilitates sample detection and can limit bias effects such as post-dispersal predation or secondary seed dispersal (Jordano et al. 2007).

Direct identification of frugivores by the shape and size of the feces is possible for some carnivore species (Guitián and Munilla 2010). Individual tracking or identification of other frugivore droppings can be challenging, such as for reptiles, birds or bats; fortunately, new molecular techniques such as DNA-barcoding or metabarcoding offer a great potential to solve this problem. DNA-barcoding methods, allow the identification of the frugivorous species from the genetic material (animal origin) present on the seeds after their dispersal, matching the sequences obtained with reference sequences deposited in the BarCode of Life databank (Hebert et al. 2004, Kress et al. 2005, González-Varo et al. 2017). More and more studies are using this method, which promotes the expansion of species with available reference sequences and the optimization and adjustment of the protocol. Conversely, if we are interested in identifying the plant species consumed by a frugivore over a period of time, DNA metabarcoding techniques would be the ideal option, however see Tercel et al. (2021) for a discussion on potential limitations of this method. This molecular approach allows the simultaneous identification of multiple taxa from a single frugivore scat containing a mixture of DNAs by means of high-throughput sequencing of a carefully selected parts of the genome, a technique widely used in plant–herbivore interaction studies (Evans et al. 2016, Kartzinel et al. 2019).

Lastly, stable-isotopes analyses can also be useful in frugivory studies, although with less resolution than with other techniques (Galetti et al. 2016). This approach is based on the premise that there is a relationship between stable isotopic compositions of consumer tissues and the stable isotopic compositions of the diet (DeNiro and Epstein 1978, 1981). The stable isotope technique is only useful in situations where two isotopically distinct dietary sources are available for frugivorous species (i.e. relative contributions of C3, and C4, plant-based proteins to avian diets, see Hobson and Clark 1992), although it may not be useful for describing interaction patterns across multiple partners.

Alternative methods

There are other approaches to compile interaction data, such as bibliographic searches, image repositories, interviews or word-of-mouth (Koike and Masaki 2008). These interaction records do not normally come with specific information on the moment of the seed dispersal process, therefore it is not possible to assess a specific interaction detection moment. For example, for some specific studies like seed dispersal systems in remote areas, conducting interviews with native inhabitants can be a precious source of information (Cámara-Leret et al. 2019), due to their close relationship with the natural environment or even their use of fleshy fruits. Another example of these compilation methods could be citizen science studies

(Bath-Rosenfeld 2019) or image repositories (Gonçalves dos Santos et al. 2019). Likewise, thorough bibliographic review in search of interaction data can provide a useful method for more general reviews or greater scope studies (Bufalo et al. 2016, Bello et al. 2017). These data gathering strategies can be very powerful, yet they come with some limitations, such as the information obtained will come from different methodologies and the sampling effort or precise georeferencing may be difficult or impossible to establish.

Complementary information obtained with different methods

Different stages of the plant–frugivore interaction process (Fig. 1) will provide varied and valuable information. The first part of the dispersal process ('visitation') is the only stage where we are able to observe both partners together. The source plant will be exclusively present during this phase, leaving a progeny in the form of a seed, to be present in later stages. This enables us to get data on feeding rates (e.g. fruits per visit, visit length), handling damage to fruits and seeds and fruit foraging behavior (Herrera and Jordano 1981, Moermond and Denslow 1985, Levey 1987, Snow and Snow 1988, Jordano and Schupp 2000). In addition, we may collect valuable information about intrinsic and extrinsic attributes of the mother plant (e.g. crop size, fruit traits, conspecific neighbourhood densities), that would not be possible otherwise (Sallabanks 1993, Miguel et al. 2018).

Methods targeting the second stage ('transport') can be very useful for zoocentric studies, since they provide valuable information on dispersing animals. During this phase individual identification and marking is possible, as well as, we can gather additional data on animal body condition, morphological traits or even measurements of gut passage time (Herrera 1984b, Remsen et al. 1993). Radio trackers can also be settled in captured animals to study dispersal distances (Uriarte et al. 2011). This type of complementary data related to animal vectors and their behavior, allows us to better understand how and why the interactions we detect are taking place, as well as to be able to project and model the consequences of their dispersal (Nathan et al. 2012).

Sampling carried out during the last seed dispersal stage ('deposition') can be suitable for plant demographic studies (Howe 1990), or animal habitat use, occupation or home range studies (Gestich et al. 2019). Maternal genetic correlates, such as relatedness between seeds, can be obtained through molecular techniques (García et al. 2009) and can help disentangling spatial genetic patterns of plant growth. Methods targeting seed deposition also provide important evidence on dispersal distance and can help identify long-distance dispersal (LDD) events (Nathan et al. 2012), with recent extensions based on extreme events theory allowing the exploration of very long-distance events (García and Borda-de-Água 2017).

Combining sampling techniques can often increase the complementary information available (Schlautmann et al. 2021). The combination of methods with different focus of study can be useful to acquire more in-depth knowledge about interactions outcome (i.e. combining phytocentric and zoocentric methods).

Sampling methods: constraints, potential limitations and sampling effort

By definition, no sample is complete; a key aspect is to evaluate how far from completeness we are when analyzing a specific system with a specific sampling method. Different methods (Fig. 1, Table 1) are subjected to different constraints (e.g. logistic, temporal, accessibility, economic cost or technical difficulty) and these may differentially affect sampling completeness. Having a robust sampling design is important. When monitoring comprises several individuals, species or areas, the sampling effort should be adjusted and its adequacy explicitly evaluated (e.g. with accumulation methods). Otherwise, a posteriori corrections need to be incorporated to account for unequal sampling effort (Jordano 2016, Vizentin-Bugoni et al. 2016). Another aspect to consider is the potential bias derived from each sampling method, mostly arising from detectability biases (Schlautmann et al. 2021).

Costs can be evaluated in terms of time, necessary expert workforce, economic expenses, material or logistics. Once the samples are collected, variation exists in terms of processing costs. A tradeoff between collection and processing costs emerges for different monitoring interaction techniques (Supporting information). Genetic or high-tech methods such as DNA-barcoding or camera traps are economically costly but they can reduce the laborious time spent in the field. While these methods can save time in the field, they frequently impose longer processing times for robust identification of animal visitors or during laboratory work; however, recent advances in automatic detection may contribute to alleviate this issue (Norouzzadeh et al. 2018).

Given the above constraints and limitations, sampling effort eventually becomes limiting for obtaining an adequate completeness of the data. Interaction accumulation curves (IAC) (Fig. 2) provide an excellent tool to estimate the sampling completeness of a study and its robustness (Jordano 2016, Macgregor et al. 2017, Mata et al 2019). This method is a simple reformulation of the species accumulation curves (SAC) method (Gotelli and Colwell 2001, Chao et al. 2014) that plots the cumulative number of unique pairwise interactions recorded as a function of sampling effort (Jordano 2016). Completeness can be estimated as the percentage of interaction richness detected with our sampling, where the observed interactions are divided by the total number of estimated interactions with Chao2 and multiplied by 100 (Chacoff et al. 2012).

Sampling effort can be measured from different perspectives: it may represent the time spent recording or identifying interactions (Fig. 2A– B), as well as number of samples collected (Fig. 2C–D) or the number of sites sampled. This approach provides an estimation of how many distinct

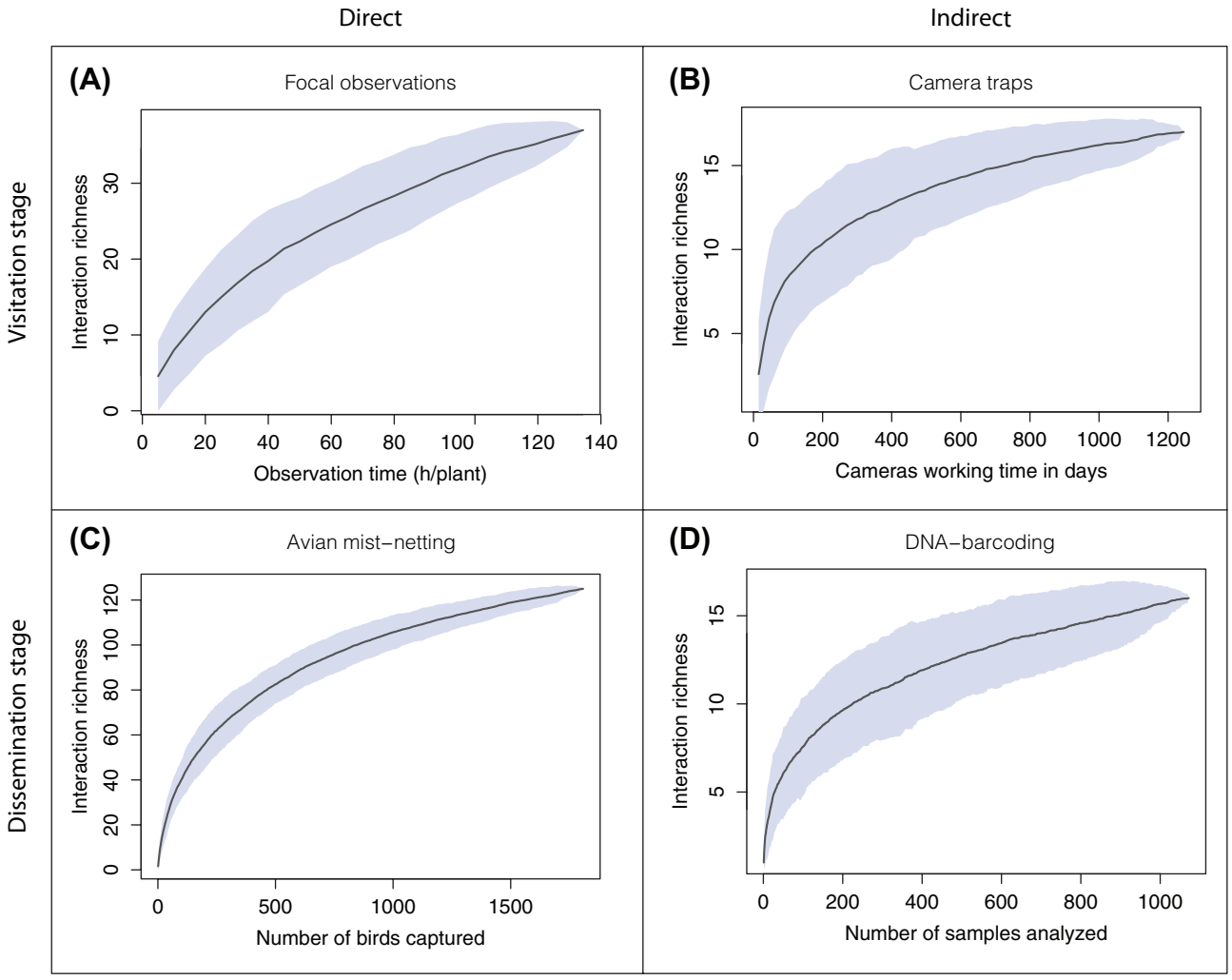

Figure 2. Examples of interaction accumulation curves (IAC), where the number of unique pairwise interactions (y-axis) accumulates as the sampling effort increases (x-axis). Each plot represents a different field sampling methodology with different sampling effort associated (x-axis). (A) Focal Observations: frugivore visits to *Cecropia glaziovii* individual plants; where the sampling effort is represented by the number of individual trees observed. (B) Camera traps: animal interactions with *Juniperus phoenicea* individual plants, where the sampling effort is represented by the number of camera-days. (C) Mist-netting: plant–frugivore interactions at community level in a Mediterranean shrubland, where the sampling effort is represented by the number of samples analyzed from captured birds. (D) DNA-barcoding: frugivore interactions with *Pistacia lentiscus* individual plants, where the sampling effort is represented by the number of fecal samples analyzed.

pairwise interactions, among the possible ones that can be recorded in the study area, are actually recorded. Different sampling methods will saturate their accumulation curve faster than others, approaching asymptotic sample completeness to variable degrees. Most recent studies of plant–frugivore interactions report sampling completeness in some way (Olesen et al. 2011, Bello et al. 2017, Acosta-Rojas et al. 2019, Costa et al. 2020).

Combining data obtained with different methodological approaches

Combining data allows overcoming the limitations of each method and obtaining a more accurate and complete representation of the interaction network (Bosch et al. 2009). The problem of data combination is central in frugivory studies, for example in analyses of complex networks aiming to get the maximum information from diverse sources to obtain a robust estimation of the interactions present. In this section we describe five different approaches to merge interaction data coming from different methodologies. To illustrate the data merging options we will consider, as an example, two matrices of pairwise interactions between a set of frugivore species and their food plants, assumed to result from different sampling approaches (Fig. 3). Interactions are tallied and summarized as adjacency matrices, with rows representing animal species and columns indicating plant species, so that matrix elements a_{ij} can represent estimates of the presence/absence (i.e. qualitative) or interaction strength (i.e.

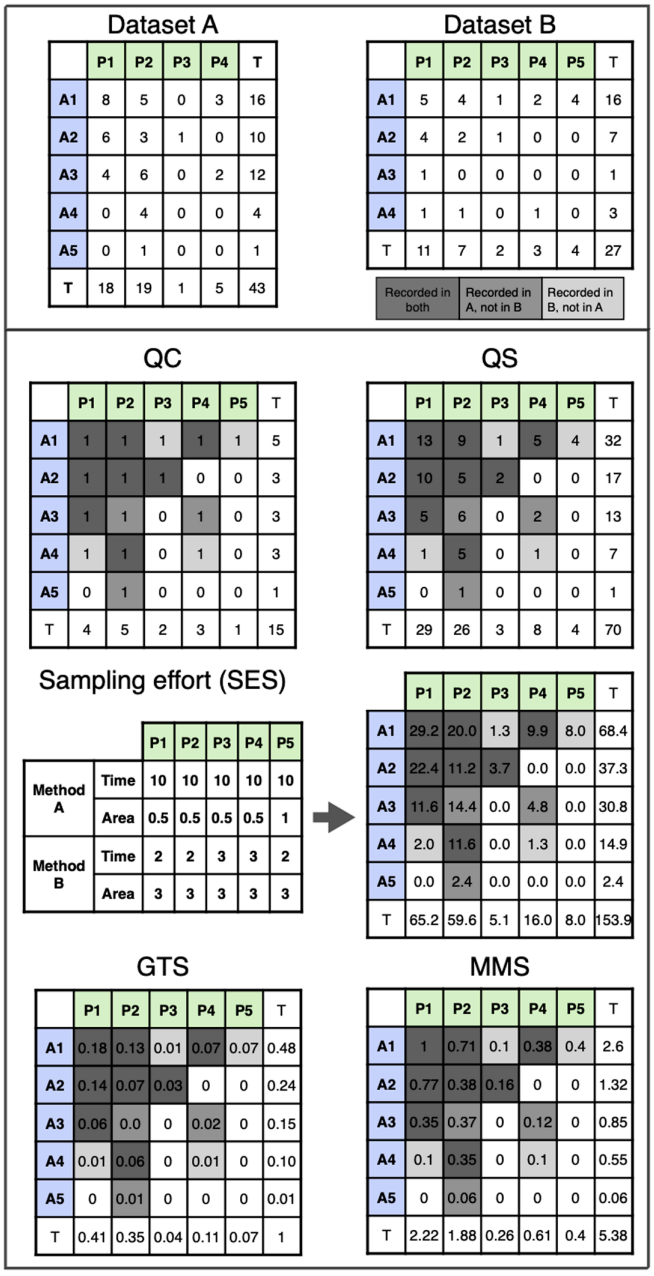

Figure 3. Illustrative example showing five merging methods for interaction data matrices considered in this study: qualitative combination (QC), quantitative sum (QS), sampling effort standardization (SES), grand total standardization (GTS) and min–max scaling (MMS). Matrices show the result of merging the simulated datasets A and B. For the SES approach we include sampling effort information for a simulated phytocentric study: area of the plant sampled and observation time on each plant.

quantitative) between animal species *i* and plant species *j* (Bascompte and Jordano 2014).

Qualitative combination (QC)

For all those cases where the characteristics of the datasets are hardly comparable, or if they just refer to presence/absence of the interaction (0–1), a qualitative combination of matrices (QC, Fig. 3) may be the most conservative option. This straightforward approach maximizes the number of pairwise interactions recorded, taking advantage of the full detectability potential of both sampling methods. Although qualitative matrices can be useful when describing frugivory assemblages (Bascompte and Jordano 2014, Almeida et al. 2018), quantitative information better describes the complexity of the structure of natural systems (Banasek-Richter et al. 2004, Dormann et al. 2009).

Quantitative sum (QS)

The simplest way to merge two weighted matrices without losing information, is to make a direct sum (QS, Fig. 3) of both datasets (Timóteo et al. 2018). This merging approach can be useful to combine data with equivalent sampling efforts whenever an absolute sum of records can be achieved without sacrificing biological interpretation. Despite incorporating more detailed information than the previous approach, it also has important limitations. Merging datasets that differ greatly in their measurement units, associated sampling efforts or spatio-temporal scales may yield unreliable results (Miranda et al. 2019) with this merging method.

Sampling effort standardization (SES)

Having detailed knowledge of the sampling effort associated with a given interaction survey (e.g. time, area, number of individuals sampled) allows using a more realistic and reliable standardization method, a sampling effort standardization (SES, Fig. 3). In order to conduct the data combination, both datasets need to be referred to the same 'currency' of interaction or unit, controlling for the sampling effort (e.g. visitation frequency in phytocentric studies, see Simmons et al. 2018, or ingestion rates in zoocentric studies). Once both matrices are standardized to a common ground, one could merge values by using the mean (option shown in Fig. 3) or the highest value recorded for each pairwise interaction. Averaging values can be problematic if the detectability of specific interactions differs significantly among techniques, as it may downplay or overestimate the weight of some interactions. On the other hand, selecting the maximum value for each interaction tries to harness the highest number of interactions observed but may also produce upward-biased values with some methods. This approach is appropriate for methodologies that share the same focus of study (i.e. only zoocentric or only phytocentric), as it can be challenging to find a common reference unit between a plant-focused study and an animal-focused one.

Grand total standardization (GTS)

When sampling techniques are very different and sampling effort correction cannot be applied, either because it is unreliable or not available, an option to collapse information is standardizing by the total number of interactions recorded. This approach, which we refer to as grand total standardization

(GTS, Fig. 3), is solely based on information from the adjacency matrix, and is recommended when sampling efforts are unknown or difficult to compare. Using a GTS approach, all the values in the adjacency matrix are weighted by the total number of interactions recorded under each specific method (i.e. the sum of all the matrix element values):

$$a'_{ij} = \frac{a_{ij}}{\sum_{i=1}^{A}\sum_{j=1}^{P} a_{ij}}$$

where a_{ij} is the interaction value for animal species i and plant species j, divided by the total sum of interactions in the adjacency matrix across all the A animal and P plant species.

Once both matrices are weighted by their respective total interactions, the final combined matrix can be calculated with a mean. This type of standardization has an immediate biological interpretation: the final matrix element value for a specific pairwise interaction, a_{ij}, indicates the probability that a randomly-chosen interaction in that community corresponds to that specific pair of partners. Merging two matrices with very different grand totals can also yield biased results, because of the strong influence of the matrix with the lowest sampling effort. Once we divide by a grand total and calculate the relative frequency of each interaction related to that grand total, we lose any information about the sample size/effort (i.e. 1/10 will weigh equal to 100/1000). Thus, even small deviations in the least sampled matrix can bias the final matrix.

Min–max scaling (MMS)

Min–max scaling is a mathematical alternative to GTS if we want to scale interactions instead of using frequentist measures (MMS, Fig. 3). This method normalizes all unique pairwise interactions into a range of values from 0 to 1. Unique pairwise interactions are scaled by subtracting the minimum value and dividing the result by the difference between the maximum and minimum values for all pairwise interactions (a_{ij}) present in the adjacency matrix (A_{ij}):

$$a'_{ij} = \frac{a_{ij} - \min(A_{ij})}{\max(A_{ij}) - \min(A_{ij})}$$

By rescaling both matrices we give a weight for each interaction on a scale of 0–1, and so allow comparison of the datasets, and their combination through a mean. This mathematical approach maintains the relative distance between the interaction weights, and the results should be interpreted in terms of interaction scoring, not probability of pairwise interaction.

Other normalization alternatives, such as those based on z-score or mean-value normalization can be problematic for two reasons. First, the biological interpretation of the resulting merged matrix can be challenging; for example, given that the distribution of interaction frequencies is highly skewed, a z-score deviation from a 'interaction frequency' mean can be misleading. Second, they produce negative values, which may preclude some types of network statistical analysis.

Preliminary considerations

We must emphasize that all the quantitative merging methods implicitly assume two comparable datasets of animal–plant interactions without major biases between them. But, it is worthy to draw attention to the eventual data differences that may hinder a successful quantitative merging, such as study scale and sampling completeness.

Often datasets differ in the temporal, spatial or taxonomic scale of resolution. Several studies may refer to incomplete phenological periods, different spatial scales or to a grouping of interactions taking place (i.e. referring to a higher taxonomic level instead of single species). Substantial differences in completeness between datasets can introduce sizable biases because a subset of the records may become overrepresented in the merged dataset (e.g. common species with more frequent interactions). If we are able to calculate the relative weight that a certain group of species, area or phenologic period has in the study datasets, we will be able to refer all interactions weights to a common ground. For example, by considering the differences in length of the study, the weight of those interactions belonging to the less complete dataset can be corrected. Standardizing our data based on the spatial, temporal or taxonomic scale would allow reliable combination between datasets.

A further issue we may encounter is a substantial difference in sampling completeness. A possibility to overcome this issue is weighting each matrix by its degree of completeness (determined from its IAC analysis) so as to have each method valued by their sampling effort coverage. Another possibility is calculating standard errors for each observed interaction probability (p) in the matrix to estimate an 'uncertainty' associated with their occurrence:

$$SE(p) = \sqrt{\frac{p(1-p)}{N}}$$

where p is the probability of the pairwise interaction occurring and N is the total number of interactions recorded.

Case-dependent weighting or adjustment of the databases prior to generating an analysis matrix is recommended to generate truthful and interpretable information.

Case studies

To illustrate the advantages and shortcomings of merging data collected through different sampling methods we use two empirical case studies, with two different organization levels. Both case studies are focused on plant–frugivore interactions taking place in the Mediterranean shrubland of Doñana National Park, Huelva, Spain. In each case study two sampling methods were used to maximise animal–plant interactions detected. The first case is an individual-based study on the avian frugivore assemblage of *Pistacia lentiscus* (Anacardiaceae) in El Puntal area, where monitoring cameras and DNA-barcoding were used to record interactions

(Quintero et al. 2021a). The second case is a community-based study aiming to document species-specific plant–frugivore interactions in Hato Ratón, where analysis of fecal samples obtained with mist-netting and focal observations were used to detect interactions (Jordano 1984, 1987, 1989, Olesen et al. 2011). All detailed information on sampling methods and protocols for each study can be found in the Supporting information.

We used the data merging approaches described above to combine sampling methodologies within each case study: qualitative combination (QC), quantitative sum (QS), sampling effort standardization (SES), grand total standardization (GTS) and min–max scaling (MMS), but the SES method was only applied in El Puntal case study. Note that for the Hato Ratón dataset, the fact that data come from a phytocentric approach (spot-censuses at plants along transects), on one hand, and from a zoocentric approach (mist-netting avian frugivores and faecal analysis), on the other, precludes the standardization to comparable units needed for a SES approach.

To standardize interaction data according to sampling effort (SES merge) for El Puntal, all interactions were referred to the number of visits per hour received by each individual plant (visits h^{-1} $plant^{-1}$). In order to do this conversion, we referred all DNA-barcoding data to the time in hours that seed traps were settled under individual plants, as well as to the plant cover area sampled by the seed traps. The same transformation for time and space was conducted with the monitoring cameras data. Bird visitation detected with the cameras was referred to hours and corrected by the percentage of canopy area observed in the videos.

We built bipartite interaction networks for each study, following the different merging methods for both initial adjacency matrices and the merged ones. We evaluated the resulting networks structure with basic metrics representing complementary aspects of the structure of mutualistic networks (Table 2, Supporting information).

Results for case studies

Interactions and species gain

The different sampling methods yielded different numbers of species, links and unique pairwise interactions in both case studies (Table 2, Fig. 4). This was expected, since some methods have unavoidable biases in sampling, e.g. mist-netting failing to capture canopy-dwelling, large frugivorous birds, limited sampling time of GoPro cameras, etc.

DNA-barcoding was the most productive method for the El Puntal case study, identifying up to 16 frugivorous bird species, compared to only seven avian species detected by the monitoring cameras. DNA-barcoding also rendered most unique pairwise interactions between individual plants and bird species (166), compared to 91 from the monitoring cameras. Yet, cameras detected 19 new distinct pairwise interactions, so combining both methods improved the completeness of the final interaction matrices.

For the Hato Ratón case study both sampling methods provided a similar number of detected species. Mist-netting aimed and was more effective in detecting plant species consumed, while the focal observations aimed to detect foraging birds, and so was more effective in detecting animals. Mist netting noticeably recorded more unique pairwise interactions than visual censuses, although focal observations yielded an increase of 30 unique pairwise interactions when combining both methods (mostly corresponding to avian frugivore species rarely or never captured in mist nets). Regarding the total number of interactions, mist netting yielded more interactions than censuses. The remarkable number of bird species detected by exclusively either one of the methods ($n = 20$) in the Hato Ratón case study, and the exclusive number of pairwise links ($n = 90$), highlights the great potential of these methods combination and data merging approaches.

Consistency and complementarity of merging methods

Pearson's product-moment and Kendall's rank correlations were used to explore how the merging methods resembled each other and how consistent they were to the initial adjacency matrices in terms of both quantitative and rank correspondence (Supporting information). Rather than focusing on the significance of these correlations we were interested in showing how variable these correlations are and whether they tend to be high or low for specific combinations of methods. All the final merged matrices showed high and significant Kendall's and Pearson's correlation between them, revealing consistent proportional weights and concordant rankings for all the unique pairwise interactions (Supporting information). However the two initial adjacency matrices in both case studies showed lower correlation between them when compared to the correlations between either the initial and merged matrices or between merged matrices resulting from different methods of data combination (Supporting information). This is expected from the substantial differences in species detectability intrinsic to each sampling method and the resulting different weights assigned to specific interactions.

For the El Puntal case study, the Kendall's correlations between initial matrices and merged ones were higher for DNA-barcoding method, indicating that ranking was better preserved for this specific methodology than for the cameras (probably since barcoding rendered much more interactions than the cameras, i.e. 1162 versus 397 records, respectively). Yet when regarding Pearson's correlation, the matrices resulting from grand total standardization (GTS) and sampling effort standardization (SES) merging methods were more correlated to the cameras than to the barcoding, indicating higher quantitatively consistency with the camera interactions records. The SES merged matrix differed the most from the other merged matrices in terms of Pearson's correlation, being most similar to GTS, but still significantly correlated to all.

Regarding Hato Ratón datasets, the merged matrices were all highly correlated both value- and rank-wise. Yet, the ranking (i.e. Kendall's correlation) of the mist-netting methodology was better preserved than the ranking of focal observations. In the case of specific interaction weights (i.e. Pearson's correlation), those of mist-netting were better preserved for quantitative sum (QS) and grand total

Table 2. Summary of species, interaction richness and network statistics recorded with different sampling methods in two study areas, El Puntal (DNA barcoding of collected samples, and monitoring cameras on *P. lentiscus* individual plants) and Hato Ratón (faecal sample analysis from mist-netting bird captures and direct focal observations during censuses), within the general area of Doñana National Park (SW Spain). The table indicates the number of species (bird species in El Puntal; bird and plant species in Hato Ratón), number of distinct pairwise links, and total number of interactions recorded in the samplings. Numbers in parentheses indicate the number of shared species, links or interactions; for modularity, number of distinct modules. Number of interactions for Hato Ratón are rounded to the nearest integer, as faecal sample analysis yields fractional fruit consumption data. Network metrics were calculated for the two initial matrices in each case study, and for the resulting matrices from the different merging approaches used: QC, QS, GTS and MMS for both case studies and additionally SES for El Puntal case study. Observed values of weighted connectance (wC), weighted nestedness (wNODF) and modularity are reported, bracketed values indicate a bootstrap-estimated confidence interval.

		Species	Pairwise links	Interactions	Weighted connectance	wNODF	Modularity
El Puntal	DNA-barcoding	16 birds 40 plants	166	1162	0.308 [0.305–0.312]	34.87 [34.08–35.66]	0.171 (4) [0.168–0.173]
	Monitoring cameras	7 birds 40 plants	91	397	0.241 [0.239–0.245]	40.75 [39.48–42.03]	0.226 (4) [0.219–0.232]
	Qualitative combination (QC)				0.287 [0.284–0.290]	72.72* [72.17–73.27]	0.321 (8) [0.316–0.327]
	Quantitative sum (QS)				0.308 [0.306–0.311]	39.69 [39.08–40.30]	0.148 (4) [0.145–0.152]
	Grand total standardization (GTS)	16 birds (7) 40 plants (40)	185 (72)	1559 (634)	0.288 [0.285–0.291]	43.36 [42.63–44.08]	0.157 (4) [0.153–0.160]
	Min–max scaling (MMS)				0.305 [0.303–0.308]	42.58 [41.77–43.39]	0.148 (4) [0.145–0.151]
	Sampling effort standardization (SES)				0.240 [0.237–0.243]	47.39 [46.65–48.15]	0.192 (4) [0.187–0.197]
Hato Ratón	Mist-netting	24 birds 15 plants	114	3541	0.095 [0.091–0.099]	65.77 [64.62–66.93]	0.120 (2) [0.111–0.129]
	Focal observations	30 birds 14 plants	82	2031	0.134 [0.131–0.138]	44.14 [42.86–45.42]	0.201 (4) [0.134–0.209]
	Qualitative combination (QC)				0.217 [0.213–0.220]	63.94* [62.60–65.27]	0.348 (5) [0.340–0.356]
	Quantitative sum (QS)	37 birds (17)	143 (53)	5572 (2042)	0.096 [0.092–0.100]	49.11 [47.92–50.30]	0.151 (4) [0.142–0.161]
	Grand total standardization (GTS)	15 plants (14)			0.103 [0.099–0.107]	46.92 [45.69–48.16]	0.162 (4) [0.151–0.172]
	Min–max scaling (MMS)				0.111 [0.107–0.114]	44.93 [43.50–46.36]	0.174 (4) [0.164–0.184]

* Note that qualitative merged matrices (QC merging method) report unweighted nestedness (NODF) and their modularity was calculated using Beckett's algorithm. See the Supporting information for analysis details.

standardization (GTS), while focal observations had a higher influence for min–max scaling (MMS) merging.

Network properties

Regarding the network properties, the largest differences in assemblage patterns and resulting indexes were found between initial adjacency matrices. This result indicates that network metrics differ more between sampling methods than between merging approaches (Table 2).

Raw connectance for El Puntal increased when obtaining the merged adjacency matrix (merged = 0.289, DNA-barcoding = 0.259, cameras = 0.142), due to matrix filling with new interactions (Bosch et al. 2009). In contrast, the Hato Ratón merged matrix connectance slightly decreased relative to the mist-netting dataset due to an increase in matrix size when considering the species recorded in mist-netting and direct observations together (merged = 0.257, mist-netting = 0.316, focal observations = 0.195). When considering weighted connectance for GTS and SES matrices, El Puntal showed lower values since both merging methods gave more weight to the cameras dataset (see Pearson's correlation Supporting information), thus more closely resembling camara weighted connectance. The same happened for the Hato Ratón dataset; the weighted connectance of the merged matrices was more similar to the specific sampling methods with which they have higher Pearson correlation (i.e. QS and GTS to mist-netting and MMS to focal observations).

The merged networks in El Puntal showed higher weighted nestedness (wNODF) values than the individual source matrices separately (Table 2), except for QS methods which were similar to the camera-derived network. Note that QC matrices are qualitative (i.e. 0–1 values), consequently unweighted NODF was computed, making its comparison with the other wNODF values unreliable weighted nestedness values for Hato Ratón merged matrices were intermediate between both methods. wNODF for the mist-netting derived adjacency

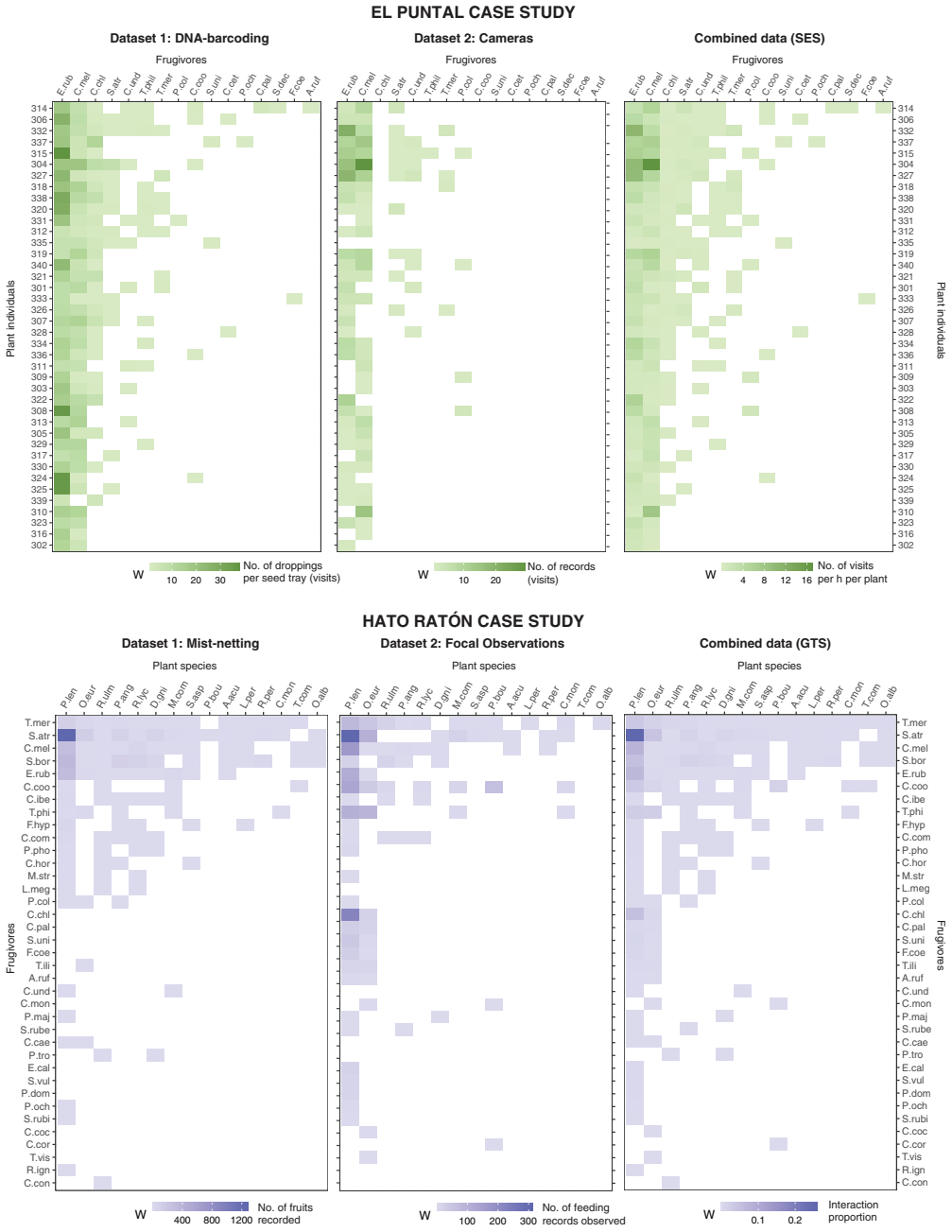

Figure 4. Empirical adjacency matrices for the two case studies, El Puntal (DNA barcoding of dispersed seeds and camera-trap monitoring of individual *P. lentiscus* plants, upper panels) and Hato Ratón (faecal sample analysis from mist-netting bird captures and direct focal observations during censuses, lower panels). The matrices on the right correspond to merged datasets, using the SES and GTM methods, respectively. Note that for El Puntal case study rows indicate plant individuals and columns indicate frugivore species (phytocentric study), while for Hato Ratón rows indicate frugivore species and columns indicate plant species (zoocentric study). Colour shade intensities indicate relative values of interaction strength (W). Animal species codes in alphabetical order: A.ruf = *Alectoris rufa*, C.cae = *Cyanistes caeruleus*, C.cet = *Cettia cetti*, C.chl = *Chloris chloris*, C.coc = *Coccothraustes coccothraustes*, C.com = *Curruca communis*, C.con = *Curruca conspicillata*, C.coo= *Cyanopica cooki*, C.cor = *Corvus corax*, C.hor = *Curruca hortensis*, C.ibe = *Curruca iberiae*, C.mel = *Curruca melanocephala*, C.mon = *Corvus monedula*, C.pal = *Columba palumbus*, C.und = *Curruca undata*, E.cal = *Emberiza calandra*, E.rub = *Erithacus rubecula*, F.coe = *Fringilla coelebs*, F.hyp = *Ficedula hypoleuca*, L.meg = *Luscinia megarhynchos*, M.str = *Muscicapa striata*, P.col= *Phylloscopus collybita*, P.dom = *Passer domesticus*, P.maj = *Parus major*, P.och = *Phoenicurus ochruros*, P.pho = *Phoenicurus phoenicurus*, P.tro = *Phylloscopus trochilus*, R.ign = *Regulus ignicapilla*, S.atr = *Sylvia atricapilla*, S.bor = *Sylvia borin*, S.dec = *Streptopelia decaocto*, S.rube = *Saxicola rubetra*, S.rubi = *Saxicola rubicola*, S.uni = *Sturnus unicolor*, S.vul = *Sturnus vulgaris*, T.ili = *Turdus iliacus*, T.mer = *Turdus merula*, T.phil = *Turdus philomelos*, T.vis = *Turdus viscivorus*. Plant species codes in alphabetical order: A.acu = *Asparagus acutifolius*, C.mon = *Crataegus monogyna*, D.gni = *Daphne gnidium*, L.per = *Lonicera periclymenum*, M.com = *Myrtus communis*, O.eur = *Olea europaea* var. *sylvestris*, O.alb = *Osyris alba*, P.ang = *Phillyrea angustifolia*, P.len = *Pistacia lentiscus*, P.bou = *Pyrus bourgaeana*, R.lyc = *Rhamnus lycioides*, R.per = *Rubia peregrina*, R.ulm = *Rubus ulmifolius*, S.asp = *Smilax aspera*, T.com = *Tamus communis*.

matrix was considerably higher than the value of the focal observations censuses matrix. This is likely attributable to the limited detectability of the mist-net captures, which selectively sample a subset of all the birds present in the area.

Modularity was similar for all matrices, being highest for the qualitative merging (QC) in both case studies. It is unreal to compare modularity results produced by QC method with the rest of merging approaches since different algorithms are used for qualitative (Beckett's algorithm) and weighted (Newman's algorithm) adjacency matrices (Dormann et al. 2009). El Puntal network derived from monitoring cameras showed higher modularity compared to theDNA-barcoding network (Table 2), probably corresponding to an increase in DNA-barcoding species detectability. In Hato Ratón the modularities of weighted merged matrices were intermediate between those of the source datasets.

In general, when both sampling methods were efficient and complementary, as in Hato Ratón study, the resulting merged matrices had intermediate values for the different network descriptors. However, in El Puntal case, where sampling methods were more redundant, network descriptors for the merged matrices resembled more to either one of the initial matrices, depending on the sampling methodology with which they had a higher Pearson's correlation (GTS and SES resembling more to cameras and QS and MMS to DNA-barcoding; Supporting information).

Discussion

Most plant–frugivore interaction studies involve some type of sampling to gain insight into the interaction partners: their diversity, numbers, spatial and temporal trends, etc. Our results provided an overview of different alternatives for data-merging, linked to the specific stage of the animal-mediated seed dispersal process being studied. The methodological approaches used with animal frugivores and fleshy-fruited plants have diversified enormously since the pioneer, observation-based methods (Howe and Vande Kerckhove 1980, Snow and Snow 1988), now including a plethora of active, passive, automated, direct, indirect and big-data oriented methods. Rather than aiming at an exhaustive review, or even a complete comparative analysis, we focused on analyzing the potential to combine multiple data sources in a biologically-insightful way.

Methodological advances in frugivory studies

With the arrival of new molecular and multimedia methods, the field of plant–frugivore interactions has expanded a great deal its exploration potential. Passive sampling methods (i.e. not requiring the active presence of the researcher during the interaction) have allowed us longer sampling extensions, leading to less work time in the field but higher post-processing efforts. Both the scale of sampling (ability to record interactions over broader spatial scales) and its precision (ability to detect rare interactions) have increased immensely. Confronted with such a variety of methods an under-researched aspect has been the development of merging strategies capable of combining data coming from a variety of sources and approaches.

Some obvious biases seem, however, unavoidable; for example, geographic and habitat-type generated biases. Focal and camera-trap observations are probably better suited for tropical areas, where the spatial scale of samplings necessarily has to be more extensive than in temperate areas, just to be able to sample rare species and interactions. On the other hand, indirect methods like those based on DNA-barcoding analyses may become more limited in tropical areas because of sample processing, collection and preservation. Furthermore, the lack of DNA sequence data for many species, some not even known, limits the use of these molecular techniques in megadiverse areas. Studies in insular habitats may require a broader combination of methodological approaches, given that their frugivore assemblages tend to include a more diverse array of frugivore higher taxa.

Combining approaches

Our analysis reveals that any combination of methods yielded better results in terms, among other things, of completeness and representability, than resorting to a single sampling method and simply ignoring potential biases inherent to it.

The high and significant correlations between different merging approaches in the two case studies analyzed shows that they all produce consistent results. Provided that the sampling has been robust and sufficient, merging simply yields a more complete and thorough dataset and may compensate for sampling biases inherent to the initial methods. Accordingly, the selection of the merging method should depend mainly on the characteristics of the available data and the interpretable output needed (e.g. in terms of probability, ranking, frequencies, etc.). Note that SES approach appears more limiting when facing the merging of data obtained with different study focus, such as when combining samplings of animal feces and observations at focal plants. When the sampling methods to be combined have both the same approach, either phyto- or zoo-centric, the SES combination appears more straightforward, given that it involves similar currencies to quantify interaction strengths. The Hato Ratón case study (involving both phyto- and zoo-centric methods) suggests that merging approaches such as GTS or MMS can be a suitable tool to increase data availability in a reliable way, allowing the merging of datasets sampled with rather different approaches. Specific consideration should then be given to the biological interpretation of the merged results, e.g. probabilistic estimates of interspecific interaction or pairwise interaction scoring.

While both methods in El Puntal were indirect (i.e. with no disturbance because of human presence), the DNA-barcoding allows recording interactions for longer time (a passive method, sampling the seed rain), yet for a smaller plant area (i.e. a limited percentage of the plant canopy surveyed). In contrast, while the monitoring cameras worked for substantially less time, they provided coverage for monitoring

frugivore activity and visitation over most of the plant. This resulted in a tradeoff between area and time. It is worthy to draw attention to the difference in area and time scales between methods. While the area correction scale ranged from 0 to 100% of the plant cover sampled, the time correction scale was much ample (from hours to months). This resulted in a significant detriment for the DNA-barcoding method (the longest sampled method in time), whose interactions lost weight when equated to camera data. It is therefore important to consider the imbalance that may emerge between methods, whenever these scales are very different (Jordano 2016). Techniques allowing a correction by sampling effort will help in those cases (e.g. those based on cumulative sampling effort).

Our results highlight the relevance of achieving adequate standardization of data, ability to evaluate data completeness, ensure reproducibility and provide details of the data merging approaches used. The qualitative combination may be applicable to rapid interactions surveys (analogous to a biodiversity survey) for large areas or regions, where only qualitative records of the interactions being present is available.

Future perspectives

We advocate for further research within mainstream ecological studies to explore data-merging strategies, an undeveloped study line in comparison to other knowledge areas with analogous problems related to data merging from diverse experimental sources (Huttenhower et al. 2006, Steele and Tucker 2008, Lagani et al. 2016). This is timely, now that data gathering in plant–frugivore interactions is greatly increasing and that we have resources to provide open data access or data papers (Bello et al. 2017).

More and more researchers are starting to share their databases in public and open repositories (Bello et al 2017). The composition and the structure in which these databases are provided is a key aspect. Data is usually shared as an interaction adjacency matrix or an edge list (Bascompte and Jordano 2014), however, such a dataset contains summarised information, losing the variation sources. Sharing extended databases that contain information for the recorded individual pairwise interactions would allow answering more questions and would help data combination through more sophisticated methods. Providing high quality metadata associated with the datasets is also essential, and this can be readily accomplished using specific R or python packages implementing standard open science grammars for metadata specification (Boettiger and Salmon 2021). Metadata should contain not just the basic information (author, site, dates, etc.) but also information on sampling effort and both temporal and spatial scope as much detailed as possible to ensure reproducibility (e.g. number of hours of observation per individual or square meters of mist-net per time). This is fundamental for reliable dataset combination and comparison. Furthermore, providing quantitative and complementary information of the study sites and species (e.g. independently-estimated species abundance or vegetation cover) can be useful to address broader questions.

Given the diversity of methods (and their combinations) developed to study plant–frugivore interactions, one of the challenges will be to select the one or those that can best help us answer our questions. Our analysis reveals that data combination approaches open new ways towards more robust sampling of plant–frugivore interactions. No specific method is probably perfect for all situations; yet when adequately combined, even disparate methods outperform single-methods in estimating interaction richness. It seems more difficult to find an interaction that cannot be sampled than to find a method to sample it.

Acknowledgements – We thank Juan Miguel Arroyo who provided essential help in the laboratory for the DNA-barcoding analysis, and Iago Ferreiro assisted with the camera-trap data preparation. We thank the logistic and facilities support form ICTS-RBD Doñana and the Doñana National Park administrators for onsite access authorizations. Discussions and comments with Francisco Rodríguez-Sánchez and Blanca Arroyo were most helpful during the preparation of the manuscript. The 7th Frugivores and Seed Dispersal Symposium, in Corbett, India, provided the excellent atmosphere to discuss and share ideas that motivated this study.
Funding – Our work was funded by project by CGL2017-82847-P from the Agencia Estatal de Investigación, Spain. During their respective PhD projects, EQ is being supported by 'La Caixa' Foundation grant ID LCF/BQ/DE18/11670007, and JI by FPI grant PRE2018-085916 from the Spanish Ministry of Science and Innovation.

Author contribtuions

Elena Quintero: Conceptualization (equal); Data curation (lead); Formal analysis (equal); Funding acquisition (supporting); Writing – original draft (lead); Writing – review and editing (equal). **Jorge Isla**: Conceptualization (equal); Data curation (supporting); Formal analysis (equal); Funding acquisition (supporting); Visualization (equal); Writing – original draft (lead); Writing – review and editing (equal). **Pedro Jordano**: Conceptualization (equal); Data curation (lead); Formal analysis (equal); Funding acquisition (lead); Writing – original draft (supporting); Writing – review and editing (equal).

Data availability statement

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.jm63xsjb8> (Quintero et al. 2021b).

References

Acosta-Rojas, D. C. et al. 2019. An integrative approach to discern the seed dispersal role of frugivorous guilds in a Mediterranean semiarid priority habitat. – PeerJ 7: e7609.

Almeida, A. de and Mikich, S. B. 2018. Combining plant–frugivore networks for describing the structure of neotropical communities. – Oikos 127: 184–197.

Alves-Costa, C. P. and Lopes, A. V. 2001. Using artificial fruits to evaluate fruit selection by birds in the field. – Biotropica 33: 713–717.

Banasek-Richter, C. et al. 2004. Sampling effects and the robustness of quantitative and qualitative food-web descriptors. – J. Theor. Biol. 226: 23–32

Bascompte, J. and Jordano, P. 2014. Mutualistic networks. – Princeton Univ. Press.

Bath-Rosenfeld, R. J. 2019. Engaging citizen scientists to determine the importance of non-agricultural fruiting trees for seed-dispersing frugivores on working landscapes in southern Costa Rica. – PhD thesis, Univ. of Northern Arizona, ProQuest Dissertations Publishing, 13883577.

Bello, C. et al. 2017. Atlantic frugivory: a plant–frugivore interaction data set for the Atlantic Forest. – Ecology 98: 1729–1729.

Boettiger, C. and Salmon, M. 2021. Codemetar: generate 'CodeMeta' metadata for R packages. ver. 0.9.1. – <https://github.com/ropensci/codemetar/>.

Bosch, J. et al. 2009. Plant–pollinator networks: adding the pollinator's perspective. – Ecol. Lett. 12: 409–419.

Bufalo, F. S. et al. 2016. Seed dispersal by primates and implications for the conservation of a biodiversity hotspot, the Atlantic forest of South America. – Int. J. Primatol. 37: 333–349.

Cámara-Leret, R. et al. 2019. Indigenous knowledge networks in the face of global change. – Proc. Natl Acad. Sci. USA 116: 9913–9918.

Campos-Arceiz, A. et al. 2012. Asian tapirs are no elephants when it comes to seed dispersal. – Biotropica 44: 220–227.

Carlo, T. A. et al. 2009. A new method to track seed dispersal and recruitment using N-15 isotope enrichment. – Ecology 90: 3516–3525.

Carnicer, J. et al. 2009. The temporal dynamics of resource use by frugivorous birds: a network approach. – Ecology 90: 1958–1970.

Carreon-Martinez, L. and Heath, D. D. 2009. Revolution in food web analysis and trophic ecology: diet analysis by DNA and stable isotope analysis. – Mol. Ecol. 19: 25–27.

Chacoff, N. P., et al. 2012. Evaluating sampling completeness in a desert plant–pollinator network: sampling a plant–pollinator network. – J. Anim. Ecol. 81: 190–200.

Chao, A. et al. 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. – Ecol. Monogr. 84: 45–67.

Cohen, J. E. 1978. Food webs and niche space. – Princeton Univ. Press.

Correa, S. B. et al. 2015. Overfishing disrupts an ancient mutualism between frugivorous fishes and plants in Neotropical wetlands. – Biol. Conserv. 191: 159–167.

Costa, J. M. et al. 2014. Endozoochory largely outweighs epizoochory in migrating passerines. – J. Avian Biol. 45: 59–64.

Costa, J. M. et al. 2020. Species temporal persistence promotes the stability of fruit–frugivore interactions across a 5-year multilayer network. – J. Ecol. 108: 1888–1898.

Deniro, M. J. and Epstein, S. 1978. Influence of diet on the distribution of carbon isotopes in animals. – Geochim. Cosmochim. Acta 42: 495–506.

Deniro, M. J. and Epstein, S. 1981. Influence of diet on the distribution of nitrogen isotopes in animals. – Geochim. Cosmochim. Acta 45: 341–351.

Dennis, A. J. et al. (eds) 2007. Seed dispersal: theory and its applications in a changing world. – CAB International.

Dormann, C. F. et al. 2009. Indices, graphs and null models: analyzing bipartite ecological networks. – Open Ecol. J. 2: 7–24.

Egerton, F. N. 2007. Understanding food chains and food webs, 1700–1970. – Bull. Ecol. Soc. Am. 88: 50–69.

Estrada, A. and Fleming, T. H. (eds) 1986. Frugivores and seed dispersal. – Dr. W. Junk Publishers, Dordrecht.

Evans, D. M. et al. 2016. Merging DNA metabarcoding and ecological network analysis to understand and build resilient terrestrial ecosystems. – Funct. Ecol. 30: 1904–1916.

Fleming, T. and Estrada, A. (eds) 1993. Frugivory and seed dispersal: ecological and evolutionary aspects. – Kluwer Academic Publishers.

Forget, P.-M. and Wenny, D. 2005. How to elucidate seed fate? A review of methods used to study seed removal and secondary seed dispersal. – In: Forget, P.-M. et al. (eds), Seed fate. Predation, dispersal and seedling establishment. CAB International, pp. 393.

Galetti, M. and Pizo, M. A. 1996. Fruit eating birds in a forest fragment in southeastern Brazil. – Ararajuba Rev. Bras. Ornitol. 4: 71–79.

Galetti, M. et al. 2008. Big fish are the best: seed dispersal of *Bactris glaucescens* by the pacu fish *Piaractus mesopotamicus* in the Pantanal, Brazil. – Biotropica 40: 386–389.

Galetti, M. et al. 2016. Trophic niche differentiation in rodents and marsupials revealed by stable isotopes. – PLoS One 11: e0152494.

García, C. and Borda-de-Água, L. 2017. Extended dispersal kernels in a changing world: insights from statistics of extremes. – J. Ecol. 105: 63–74.

García, C. et al. 2009. Maternal genetic correlations in the seed rain: effects of frugivore activity in heterogeneous landscapes. – J. Ecol. 97: 1424–1435.

García, D. et al. 2004. Patterns of resource tracking by avian frugivores at multiple spatial scales: two case studies on discordance among scales. – Ecography 27: 187–196.

Garrote, P. J. et al. 2018. Strong among population variation in frugivory strength by functional diverse frugivores: a 'reciprocal translocation' experiment. – Oecologia 187: 143–154.

Genrich, C. M. et al. 2017. Duality of interaction outcomes in a plant–frugivore multilayer network. – Oikos 126: 361–368.

Gestich, C. C. et al. 2019. From dropping to dropping: the contribution of a small primate to seed dispersal in Atlantic Forest. – Acta Oecol. 100: 103464.

Gomes, V. G. N. et al. 2013. Pollination and seed dispersal of *Melocactus ernestii* Vaupel subsp. *ernestii* (Cactaceae) by lizards: an example of double mutualism. – Plant Biol. 16: 315–322.

Gómez, J. M. et al. 2019. Synzoochory: the ecological and evolutionary relevance of a dual interaction. – Biol. Rev. 94: 874–902.

Gonçalves dos Santos, L. C. et al. 2019. A interação frugívora ave-planta no distrito federal segundo a base de dados 'WikiAves.' – In: Poster in XIV Congresso Ecologia do Brasil, São Lourenço, Brasil.

González-Varo, J. P. et al. 2013. Functional diversity among seed dispersal kernels generated by carnivorous mammals. – J. Anim. Ecol. 82: 562–571.

González-Varo, J. P. et al. 2017. Unravelling seed dispersal through fragmented landscapes: frugivore species operate unevenly as mobile links. – Mol. Ecol. 26: 4309–4321.

González-Varo, J. P. et al. 2014. Who dispersed the seeds? The use of DNA barcoding in frugivory and seed dispersal studies. – Methods Ecol. Evol. 5: 806–814.

González-Varo, J. P. et al. 2019. Seed dispersal by dispersing juvenile animals: a source of functional connectivity in fragmented landscapes. – Biol. Lett. 15: 20190264.

Gotelli, N. J. and Colwell, R. K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. – Ecol. Lett. 4: 379–391.

Guitián, J. and Munilla, I. 2010. Responses of mammal dispersers to fruit availability: rowan (*Sorbus aucuparia*) and carnivores in mountain habitats of northern Spain. – Acta Oecol. 36: 242–247.
Hansen, D. M. and Mueller, C. B. 2009. Reproductive ecology of the endangered enigmatic Mauritian endemic *Roussea simplex* (Rousseaceae). – Int. J. Plant Sci. 170: 42–52.
Hansen, D. M. and Traveset, A. 2012. An overview and introduction to the special issue on seed dispersal on islands. – J. Biogeogr. 39: 1935–1937.
Hebert, P. D. N. et al. 2004. Identification of birds through DNA barcodes. – PLoS Biol. 2: 1657–1663.
Herrera, C. M. 1984a. Adaptation to frugivory of Mediterranean avian seed dispersers. – Ecology 65: 609–617.
Herrera, C. M. 1984b. A study of avian frugivores, bird-dispersed plants and their interaction in Mediterranean scrublands. – Ecol. Monogr. 54: 1–23.
Herrera, C. M. and Jordano, P. 1981. *Prunus mahaleb* and birds: the high-efficiency seed dispersal system of a temperate fruiting tree. – Ecol. Monogr. 51: 203–218.
Hobson, K. A. and Clark, R. G. 1992. Assessing avian diets using stable isotopes I: turnover of C-13 in tissues. – Condor 94: 181–188.
Howe, H. F. 1990. Seed dispersal by birds and mammals: implications for seedling demography. – In: Bawa, K. S. and Hadley, M. (eds), Reproductive ecology of tropical forest plants. Man and the biosphere series, Vol. 7. UNESCO and Parthenon Publ. Group, pp. 191–218.
Howe, H. F. and Vande Kerckhove, G. A. 1980. Nutmeg dispersal by tropical birds. – Science 210: 925–927.
Howe, H. F. and Vande Kerckhove, G. A. 1981. Removal of wild nutmeg *Virola surinamensis* crops by birds. – Ecology 62: 1093–1106
Howe, H. F. and Smallwood, J. 1982. Ecology of seed dispersal. – Annu. Rev. Ecol. Syst. 13: 201–228.
Huttenhower, C. et al. 2006. A scalable method for integration and functional analysis of multiple microarray datasets. – Bioinformatics 22: 2890–2897.
IPBES 2019. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Brondizio, E. S. et al. (eds). – IPBES Secretariat, Bonn, Germany, <https://doi.org/10.5281/zenodo.3831673>.
Jácome-Flores, M. E. et al. 2020. Interaction motifs variability in a Mediterranean palm under environmental disturbances: the mutualism–antagonism continuum. – Oikos 129: 367–379.
Janzen, D. H. 1983. Seed and pollen dispersal by animals – convergence in the ecology of contamination and sloppy harvest. – Biol. J. Linn. Soc. 20: 103–113.
Jordano, P. 1984. Relaciones entre plantas y aves frugívoras en el matorral Mediterráneo del área de Doñana. – PhD thesis, Univ. de Sevilla, <https://hdl.handle.net/11441/70423>.
Jordano, P. 1987. Avian fruit removal: effects of fruit variation, crop size and insect damage. – Ecology 68: 1711–1723.
Jordano, P. 1989. Pre-dispersal biology of *Pistacia lentiscus* (Anacardiaceae): cumulative effects on seed removal by birds. – Oikos 55: 375–386.
Jordano, P. 1993. Geographical ecology and variation of plant–seed disperser interactions: southern Spanish junipers and frugivorous thrushes. – In: Fleming, T. H. and Estrada, A. (eds), Frugivory and seed dispersal: ecological and evolutionary aspects. Kluwer Academic Publishers, pp. 85–104.
Jordano, P. 2016. Sampling networks of ecological interactions. – Funct. Ecol. 30: 1883–1893.
Jordano, P. and Schupp, E. W. 2000. Seed disperser effectiveness: the quantity component and patterns of seed rain for *Prunus mahaleb*. – Ecol. Monogr. 70: 591–615.
Jordano, P. et al. 2007. Differential contribution of frugivores to complex seed dispersal patterns. – Proc. Natl Acad. Sci. USA 104: 3278–3282.
Kartzinel, T. R. et al. 2019. Covariation of diet and gut microbiome in African megafauna. – Proc. Natl Acad. Sci. USA 116: 23588–23593.
Koike, S. and Masaki, T. 2008. Frugivory of Carnivora in central and southern parts of Japan analyzed by literature search. – J. Jpn For. Sci. 90: 27–36.
Korine, C. et al. 2000. Fruit characteristics and factors affecting fruit removal in a Panamanian community of strangler figs. – Oecologia 123: 560–568.
Kress, W. J. et al. 2005. Use of DNA barcodes to identify flowering plants. – Proc. Natl Acad. Sci. USA 102: 8369–8374.
Lagani, V. et al. 2016. A comparative evaluation of data-merging and meta-analysis methods for reconstructing gene–gene interactions. – BMC Bioinfo. 17: S194.
Levey, D. J. 1987. Seed size and fruit-handling techniques of avian frugivores. – Am. Nat. 129: 471–485.
Levey, D. J. et al. (eds) 2002. Seed dispersal and frugivory: ecology, evolution and conservation. – CAB International.
Macgregor, C. J. et al. 2017. Estimating sampling completeness of interactions in quantitative bipartite ecological networks: incorporating variation in species specialisation. – bioRxiv: 195917 <https://doi.org/10.1101/195917>.
Magurran, A. E. 1988. Ecological diversity and its measurement. – Princeton Univ. Press.
Mata, V. A. et al. 2019. How much is enough? Effects of technical and biological replication on metabarcoding dietary analysis. – Mol. Ecol. 28: 165–175.
Miguel, M. F. et al. 2018. Context-dependency and anthropogenic effects on individual plant–frugivore networks. – Oikos 127: 1045–1059.
Miranda, P. N. et al. 2019. The dilemma of binary or weighted data in interaction networks. – Ecol. Complex. 38: 1–10.
Moermond, T. C. and Denslow, J. S. 1985. Neotropical avian frugivores: patterns of behavior, morphology and nutrition, with consequences for fruit selection. – Ornithol. Monogr. 36: 865–897.
Moran, C. P. et al. 2004. Functional variation among frugivorous birds: implications for rainforest seed dispersal in a fragmented subtropical landscape. – Oecologia 141: 584–595.
Nathan, R. E. et al. 2012. Dispersal kernels: review. – In: Clobert, J. et al. (eds), Dispersal ecology and evolution. Oxford Univ. Press, pp. 186–210
Niquil, N. et al. 2020. Shifting levels of ecological network's analysis reveals different system properties. – Phil. Trans. R. Soc. B 375: 20190326.
Norouzzadeh, M. S. et al. 2018. Automatically identifying, counting and describing wild animals in camera-trap images with deep learning. – Proc. Natl Acad. Sci. USA 115: E5716–E5725.
Olesen, J. M. et al. 2011. Missing and forbidden links in mutualistic networks. – Proc. R. Soc. B 278: 725–732.
Oliveira P, et al. 2002. Diet of the endemic Madeira laurel pigeon and fruit resource availability: a study using microhistological analyses. – Condor 104: 811–822.
Ollerton, J. 2006. 'Biological barter': patterns of specialization compared across different mutualisms. – In: Waser, N. M. and

Ollerton, J. (eds), Plant–pollinator interactions: from specialization to generalization. Chicago Univ. Press, pp. 411–435.
Palmeirim, J. M. et al. 1989. Trophic structure of a neotropical frugivore community: is there competition between birds and bats? – Oecologia 79: 403–411.
Perea, R. et al. 2013. Context-dependent fruit–frugivore interactions: partner identities and spatio-temporal variations. – Oikos 122: 943–951.
Polis, G. A. and Strong, D. R. 1996. Food web complexity and community dynamics. – Am. Nat. 147: 813–846.
Quintero, E. et al. 2021a. Super-generalist species in complex ecological networks: forms of interactions and implications. – Unpubl. PhD thesis, Univ. of Sevilla. <https://github.com/PJordano-Lab/MS_Oikos_FSD_Monitoring_interactions>.
Quintero, E. et al. 2021b. Data from: Methodological overview and data-merging approaches in the study of plant–frugivore interactions. – Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.jm63xsjb8>.
Remsen, J. V. et al. 1993. The diets of neotropical trogons, motmots, barbets and toucans. – Condor 95: 178–192.
Rother, D. C. et al. 2015. Community-wide spatial and temporal discordances of seed-seedling shadows in a tropical rainforest. – PLoS One 10: e0123346.
Sallabanks, R. 1993. Hierarchical mechanisms of fruit selection by an avian frugivore. – Ecology 74: 1326–1336.
Schlautmann, J. et al. 2021. Observing frugivores or collecting scats: a method comparison to construct quantitative seed dispersal networks. – Oikos 10.1111/oik.08175.
Schupp, E. W. et al. 2002. Dissemination limitation and the origin and maintenance of species-rich tropical forests. – In: Levey, D. et al. (eds), Seed dispersal and frugivory ecology, evolution and conservation. CAB International, pp. 19–33.
Schupp, E. W. et al. 2017. A general framework for effectiveness concepts in mutualisms. – Ecol. Lett. 20: 577–590.
Simmons, B. I. et al. 2018. Moving from frugivory to seed dispersal: incorporating the functional outcomes of interactions in plant–frugivore networks. – J. Anim. Ecol. 87: 995–1007.
Snow, B. K. and Snow, D. W. 1988. Birds and berries. – T. and A.D. Poyser, Calton, UK.
Steele, E. and Tucker, A. 2008. Consensus and meta-analysis regulatory networks for combining multiple microarray gene expression datasets. – J. Biomed. Inform. 41: 914–926.
Stevenson, P. R. et al. 2015. Frugivory in canopy plants in a western Amazonian forest: dispersal systems, phylogenetic ensembles and keystone plants. – PLoS One 10: e0140751.
Tercel, P. T. G. et al. 2021. The problem of omnivory: a synthesis on omnivory and DNA metabarcoding. – Mol. Ecol. 30: 2199–2206.
Timóteo, S. et al. 2018. Multilayer networks reveal the spatial structure of seed-dispersal interactions across the Great Rift landscapes. – Nat. Commun. 9: 140–140.
Uriarte, M. et al. 2011. Disentangling the drivers of reduced long-distance seed dispersal by birds in an experimentally fragmented landscape. – Ecology 92: 924–937.
Valentini, A. et al. 2009. DNA barcoding for ecologists. – Trends Ecol. Evol. 24: 110–117.
Valiente-Banuet, A. et al. 2014. Beyond species loss: the extinction of ecological interactions in a changing world. – Funct. Ecol. 29: 299–307.
Vaz, V. C. et al. 2012. Notes on food habits of armadillos (Cingulata, Dasypodidae) and anteaters (Pilosa, Myrmecophagidae) at Serra da Capivara National Park (Piauí State, Brazil). – Edentata 13: 84–89.
Vizentin-Bugoni, J. et al. 2016. Influences of sampling effort on detected patterns and structuring processes of a Neotropical plant–hummingbird network. – J. Anim. Ecol. 85: 262–272.
Weiss, B. et al. 2016. Viability of seeds consumed by fishes in a lowland forest in the Brazilian central Amazon. – Trop. Conserv. Sci. 9: 1–10.

OIKOS SPECIAL ISSUE

Online Supplementary Material

Methodological overview and data-merging approaches in the study of plant-frugivore interactions

A. Trade offs between collection and processing costs for different monitoring interaction techniques

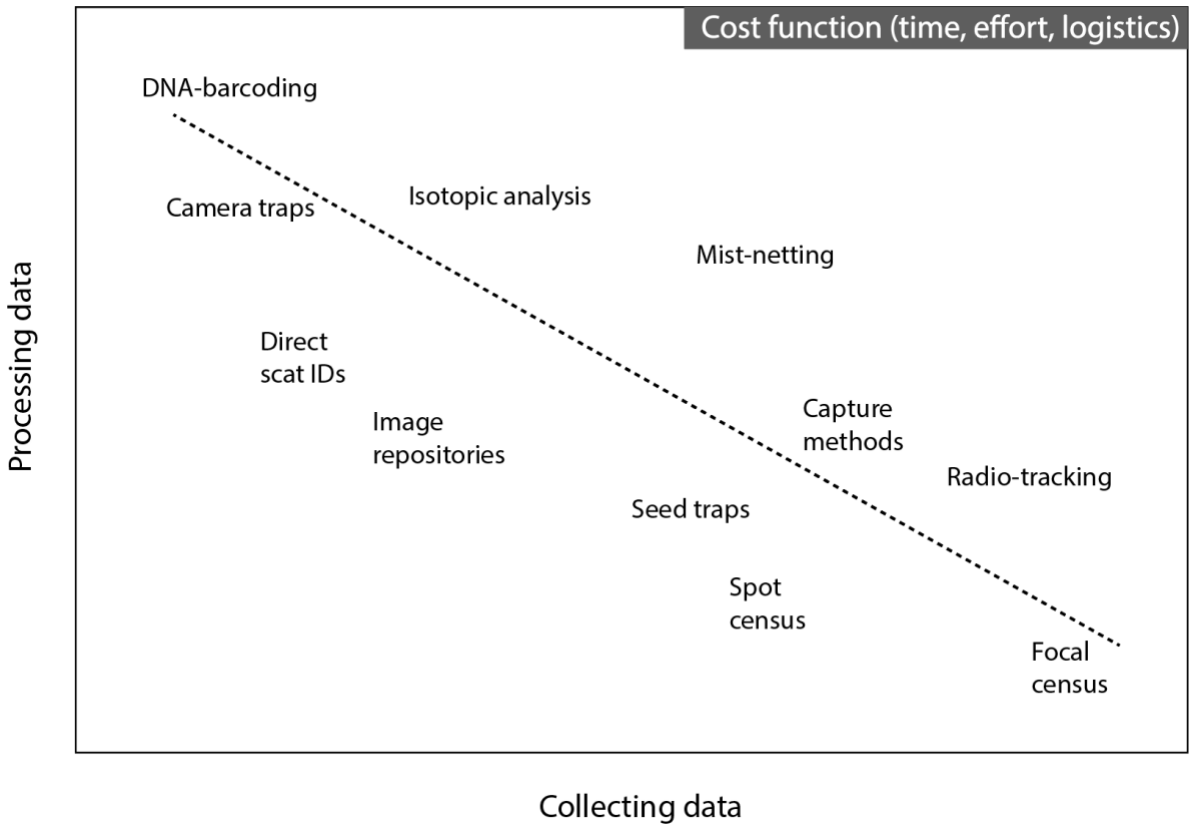

Figure SM-1. Approximate cost function trade-offs for methods used to assess plant-frugivore interactions. Different methods are placed approximately at locations that combine the costs (time, effort, logistic) of collecting the data compared to the costs (e.g., laboratory work, sample analysis, video image processing) involved in data processing.

B. Case Studies: materials and methods

We used two empirical datasets to illustrate data merging approaches. The first study locality is a lowland Mediterranean shrubland covering ca. 15 ha in Mancha del Rabicano in the area of El Puntal, Doñana National Park, Huelva, SW Spain. A total of 40 *P. lentiscus* plants were marked and surveyed during the fruiting season of 2018-2019. Data collected for this study is partially complete, since it includes plant-frugivore interactions taking place just during the winter season, right in the middle of the fruiting peak, not depicting the complete frugivore assemblage in the area.

In order to capture all avian visitors interacting with *P. lentiscus* individual plants, we used two indirect sampling methods: one focused at the 'Visitation' stage and the other at the 'Deposition' stage. The 'Visitation' method involved placing continuous-monitoring cameras (GoPro Hero® 7 model) facing individual plants (Fig. SM-2). Forty individual plants were filmed for approximately 2 hours in several runs in different days (total of 84.5h), and any avian visitation was recorded as an interaction, yielding a total of 397 visitation records. Cameras were operative from sunrise for 2h recording set at maximum resolution. Data resulting from this sampling can be given as total number of records, or standardized by sampling time (no. records h^{-1}).

Figure SM-2. Interaction record using camera traps at *Pistacia lentiscus* individual plants in order to obtain estimates of individual-based plant-animal interaction networks. Camera records allow in many instances obtaining data on fruit handling, feeding rates, etc., in addition to just the visitation record. A male Sardinan warbler *Curruca melanocephala* just after picking a ripe *P. lentiscus* fruit.

The 'Deposition' method was based on DNA-barcoding identification of faecal samples collected in seed traps (plastic trays covered with 1 cm mesh wire) under the same forty individual trees (González-Varo et al. 2014). All samples were retrieved from seed trays located under individual plants, working for 102.7±8.9 days (mean ±SD) per plant. A total of 1371 faecal samples were

analyzed (mean no. per plant: 33.8±15.2). Samples were collected regardless of whether or not they had seeds, as an indicator of a visitation event. Eventually, all samples containing *P. lentiscus* seeds indicate the role of those frugivore species as legitimate dispersers. Yet, since effective dispersal is not our scope, and for the sake of comparison with the monitoring camera data, any visitation event is considered. Faecal samples were stored at -20ºC and later processed following protocols described in detail by González-Varo et al. (2014). Avian DNA was extracted from the surface of defecated or regurgitated seeds or the surface of the scat (samples without seeds) (see Marrero et al. 2009), allowing the identification of the frugivore species that contributed each dispersal event or potential visit to the plant. Frugivore species identification was based on a 464-bp mitochondrial DNA region (COI: cytochrome c oxidase subunit I), employing the 'Barcode Of Life Data' identification system (BOLD: http://www. boldsystems.org; Hebert et al. 2004). BOLD accepts sequences from the COI gene and returns species-level identification whenever possible and assigns a percentage of similarity to matched sequences. All samples were amplificated by PCR using the COI-fsdF and COI-fsdR primers (see González-Varo et al. 2014). This product was latter sequenced and verified for its matching with COI sequences from BOLD databases. Data resulting from this sampling can be given as a total number of records with positive identification of a given frugivore species, or standardized by the sampling time with seed traps actively operating in the field (no. records/trap/day or similar).

The second case is a community-based study aiming to document species-specific plant-frugivore interactions in Hato Ratón, an area with similar landscape physiognomy to El Puntal (N Doñana Natl. Park, S Spain; Jordano, 1984, 1987a, 1989; Olesen et al. 2011). Data collected for this study completely spans two fruiting seasons (1981-1982 and 1982-1983) and also focuses on avian frugivores.

Two sampling techniques were used: the first, focused at the 'Transport stage', using bird mist-netting to collect avian fecal samples that were subsequently examined under microscope, quantifying the presence and relative contribution of different fruit species, either by seed or exocarp remains (see Jordano 1984, 1988 for details). Estimation of dietary diversity of frugivore species by relying just on seed identification in scats invariably underestimates the actual diversity of fruits consumed (Jordano 1988). To avoid this bias we used a microhistological technique to identify fruit species present with no seeds by examining under microscope (40X, 100X) the shape, size, and structures (trichomes, glands) of exocarp tissue cells. Similar techniques are routinely used to study the diet composition of herbivorous animals (see e.g. Marrero and Nogales 2005). This allowed not only the identification of fruit species when no seeds are present but also the relative volume occupied in the sample, so that an estimate of the corresponding number of fruits ingested can be derived (Jordano 1988). For example, a given sample of Blackcap may contain seeds from just two species (e.g., *P. lentiscus* and *Phillyrea angustifolia*, yet remains of up to 7 different species may be present and identified under microscope (Jordano 1984, 1988). Between 6-10 mist nets were operated weekly for 1-2 days (for a total of 84

sampling days and 4080.5 mist-net hours), totalling 3541 fecal samples analyzed (Jordano, 1984; Olesen et al. 2011).

The second method focused at the 'Visitation' stage through the use of focal observations. Feeding records of frugivores visiting fruiting plants were obtained during 1.0 km-length walk censuses in the area, with 2-5 censuses carried out per month (123 sampling days), totalling 89.5 km and 2031 records. These are not focal observations spanning a given time period focusing at fruiting plants (Snow and Snow 1988), but spot censuses where interactions are recorded during short stops as the observer advances along a fixed transect. A feeding record involves a frugivore seen handling a fruit (Snow and Snow 1988); in some cases (<15 % of the records) where no handling was observed but just the visit to the plant, the number of fruits was approximated from data on feeding rate (no. fruits/visit). Data resulting from this sampling can be given as total number of records, or standardized by sampling time (no. records km^{-1} census or no. records h^{-1} or day^{-1}, or similar).

C. Analyses of interaction network statistics and indexes for adjacency matrices estimated with different sampling methods in two study areas, Doñana National Park (SW Spain).

Data and code corresponding to these analyses is available at the manuscript repository: https://github.com/PJordano-Lab/MS_Oikos_FSD_Monitoring_interactions. DOI: 10.5281/zenodo.4751889.

Adjacency matrices were obtained for each sampling method. Connectance (C) is the proportion of observed links divided by the number of total potential links (Jordano 1987b). Since these are weighted networks, we also analyzed weighted connectance (wC), which is a similar connectivity metric but based on linkage density (Bersier et al. 2002). To evaluate to what extent link distribution is not structured randomly, we calculated the weighted nestedness (wNODF) and the modularity (M) and number of modules (nM) of the networks. Nestedness represents the degree to which the interactions of less-connected species are a subset of those of more connected species (Ulrich et al. 2009). Modularity is the tendency of a network to be organized in clusters, where highly inter-connected subsets of nodes are less connected to nodes in other subsets (Olesen et al. 2007). Network metrics were calculated using R package 'bipartite' (version 2.15, Dorman et al., 2009) in R statistical software version 4.0.3 (R Development Core team 2008).

Weighted connectance, wC, measures the fraction of interactions actually occurring, out of all the potential, in which each link is weighed on the basis of its frequency. Weighted connectance was computed by the weighting the number of pairwise interactions in the network with the observed frequency of each pairwise interaction (Dormann et al. 2009), *i.e.*, the linkage density divided by number of species in the network (Bersier et al. 2002). Raw connectance was calculated for the qualitative matrices (QC merging method) that report unweighted interaction values.

Modularity (M) and number of modules (nM) were estimated using the function ComputeModules in the R package bipartite (Dormann and Strauss 2014; Beckett 2016). For unweighted networks (QC) the algorithm developed by Beckett was used while for the weighted networks the Dormann algorithm was computed. The number of distinct modules was obtained for each run of the modularity algorithm and is reported as the average number of modules found in repeated runs (N= 5). These parameters quantify the tendency of a network to be organized into distinct clusters, i.e. modular networks showing distinct subsets of taxa interacting more frequently among each other than with taxa in other modules. Given that the estimation for the number of modules can vary between runs, the number of modules was calculated as the average (±SD) for 5 runs.

D. Analyses of consistency and complementarity between sampling methods aimed to obtain merged datasets.

We tested the consistency in interaction value estimates by means of both quantitative (Pearson's correlation) and non-parametric, rank-based tests (Kendall's correlation). Tests were carried out to compare the adjacency matrices estimated with different sampling methods, as well as the correlation between the merged matrices and each of the original matrices being merged. Pearson's correlation is a parametric test that indicates the consistency and correlation of the interactions weight values for two methods being compared, while Kendall's correlation is a non-parametric test that indicates the correspondence in the ranking of pairwise interactions for two different methods being compared.

Pearson's and Kendall's correlation significance was tested using permutation tests (n=9999 permutations) using function perm.relation in the R package wPerm (Weiss 2015), and resulted highly significant ($p < 0.001$) for all correlations obtained.

Figure SM-3 summarizes the correlation values obtained, overall suggesting a sizable degree of consistency both in the quantitative values and rank estimates for the pairwise interactions.

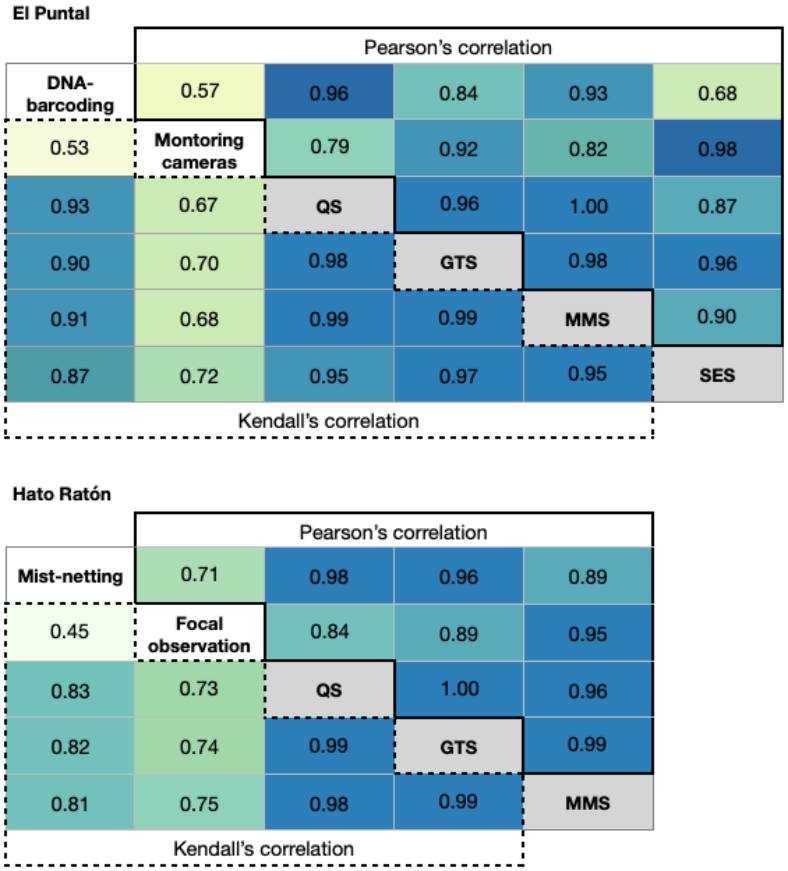

Figure SM-3. Summary of Pearson's (above diagonal) and Kendall's rank correlation (below diagonal) coefficients to assess consistency between methods used to compile plant-frugivore interaction data at two study sites: El Puntal and Hato Ratón, within the general area of Doñana National Park (SW Spain). Correlations were estimated on the raw interaction data derived from the two compared methods for all the pairwise interactions. Pearson's correlation coefficients suggest high consistency of the quantitative values recorded, while Kendall rank correlations indicate high consistency in the interaction ranks according to their value. Colors indicate different interval levels for correlations from high (dark blue) to lower correlation (light green).

E. List of bibliographic references included in Table 1, as examples of study methods for plant-frugivore interactions.

1	Howe, H.F. and Vande Kerckhove, G.A. 1980. Nutmeg dispersal by tropical birds. - Science 210: 925–927.
2	Herrera, C.M. and Jordano, P. 1981. *Prunus mahaleb* and birds : the high-efficiency seed dispersal system of a temperate fruiting tree. - Ecological Monographs 51: 203–218.
3	Snow, B.K. and Snow, D.W. 1988. Birds and berries. - T. and A.D. Poyser. Calton, UK
4	Howe, H.F. and Vande Kerckhove, G.A. 1981. Removal of Wild Nutemg (*Virola surinamensis*) Crops by Birds. - Ecology 62 (4), 1093-1106
5	Rother, D.C., et al. 2015. Variation in seed dispersal effectiveness: the redundancy of consequences in diversified tropical frugivore assemblages. Oikos, 125, 336–342.
6	Galetti, M. and Pizo M.A, 1996. Fruit eating by birds in a forest fragment in sourheastern Brazil. - Ararajuba 4 (2). 103 - 105.
7	Miguel, M.F. et al. 2018. Context-dependency and anthropogenic effects on individual plant-frugivore networks. - Oikos 127: 1045–1059.
8	Campos, R.T., et al. 2012 Bird and mammal frugivores of *Euterpe edulis* at Santa Catarina island monitored by camera traps. - Studies on Neotropical Fauna and Environment, 47:2, 105-110
9	Jácome-Flores, M.E. et al. 2020. Interaction motifs variability in a Mediterranean palm under environmental disturbances: the mutualism–antagonism continuum. - Oikos 129: 367–379.
10	Alves-Costa, C.P. and Lopes, A.V. 2001. Using artificial fruits to evaluate fruit selection by birds in the field. - Biotropica 33: 713-717.
11	Korine, C. et al. 2000. Fruit characteristics and factors affecting fruit removal in a Panamanian community of strangler figs. - Oecologia 123: 560–568.

12	Hansen, D.M. and Traveset, A. 2012. Seed dispersal on islands. - Journal of Biogeography 39: 1933-2088.
13	Garrote, P.J. et al. 2018. Strong among population variation in frugivory strength by functional diverse frugivores: a 'reciprocal translocation' experiment. - Oecologia 187: 143–154.
14	Remsen J.V. et al. 1993. The diets of neotropical trogons, motmots, barbets and toucans. - The Condor 95:178-192
15	Vaz, V.C., et al. 2012. Notes on food habits of armadillos (Cingulata, Dasypodidae) and anteaters (Pilosa, Myrmecophagidae) at Serra da Capivara National Park (Piauí State, Brazil). - Edentata, 13(1), 84-89.
16	Herrera, C.M. 1984b. A study of avian frugivores, bird-dispersed plants, and their interaction in Mediterranean scrublands. - Ecological Monographs 54: 1–23.
17	Costa, J.M. et al. 2020. Species temporal persistence promotes the stability of fruit–frugivore interactions across a 5-year multilayer network. - Journal of Ecology 00: 000-000.
18	Jordano, P. et al. 2007. Differential contribution of frugivores to complex seed dispersal patterns. - Proceedings of the National Academy of Sciences USA 104: 3278–3282.
19	Perea, R. et al. 2013. Context-dependent fruit-frugivore interactions: Partner identities and spatio-temporal variations. - Oikos 122: 943–951.
20	González-Varo, J. P., Arroyo, J. M. & Jordano, P. 2014. Who dispersed the seeds? The use of DNA barcoding in frugivory and seed dispersal studies. - Methods in Ecology and Evolution 5: 806–814.
21	Mata, V.A. et al. 2019. How much is enough? Effects of technical and biological replication on metabarcoding dietary analysis. - Molecular Ecology 28: 165–175.
22	Carlo, T.A. et al. 2009. A new method to track seed dispersal and recruitment using N-15 isotope enrichment. - Ecology 90: 3516–3525.

23	Carreon-Martínez, L. and Heath, D.D. 2009. Revolution in food web analysis and trophic ecology: diet analysis by DNA and stable isotope analysis. - Molecular Ecology 19: 25–27.
24	Hawes, J.E. and Peres, C.A. 2014. Fruit–frugivore interactions in Amazonian seasonally flooded and unflooded forests. - Journal of Tropical Ecology 30:381–399.
25	Almeida, A. and Mikich, S.B. 2018. Combining plant-frugivore networks for describing the structure of neotropical communities. - Oikos 127: 184–197.
26	Koike, S. and Masaki, T. 2008. Frugivory of Carnivora in central and Southern parts of Japan analyzed by literature search. - Journal of Japanese Forest Science 90: 27–36.
27	Gestich, C.C. et al. 2019. From dropping to dropping: The contribution of a small primate to seed dispersal in Atlantic Forest. - Acta Oecologica, 100: 103464.

F. Supplementary Material References

Beckett, S.J. 2016. Improved community detection in weighted bipartite networks. - Royal Society Open Science 3: 140536.

Bersier, L.F., Banasek-Richter, C. and Cattin, M.F. 2002. Quantitative descriptors of food-web matrices. - Ecology 83: 2394-2407.

Dormann, C. F. et al. 2009. Indices, graphs and null models: Analyzing bipartite ecological networks. - Open Ecology Journal 2: 7–24.

Dormann, C.F. and Strauss, R. 2014. A method for detecting modules in quantitative bipartite networks. - Methods in Ecology and Evolution 5: 90–98.

González-Varo, J. P., Arroyo, J. M. & Jordano, P. 2014. Who dispersed the seeds? The use of DNA barcoding in frugivory and seed dispersal studies. - Methods in Ecology and Evolution 5: 806–814.

Hebert, P.D.N. et al. 2004. Identification of birds through DNA Barcodes. - PLoS Biology 2: 1657–1663.

Jordano, P. 1984. *Relaciones entre plantas y aves frugívoras en el matorral Mediterráneo del área de Doñana*. - Ph. D. Thesis. Universidad de Sevilla. Sevilla, Spain. https://hdl.handle.net/11441/70423

Jordano, P. 1987a. Avian fruit removal: effects of fruit variation, crop size, and insect damage. - Ecology, 68: 1711-1723.

Jordano, P. 1987b. Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. - American Naturalist 129: 657 – 677.

Jordano, P. 1988. Diet, fruit choice and variation in body condition of frugivorous warblers in Mediterranean scrubland. - Ardea, 76: 193-209.

Jordano, P. 1989. Pre-dispersal biology of *Pistacia lentiscus* (Anacardiaceae): cumulative effects on seed removal by birds. - Oikos, 55: 375–386.

Marrero, P. and Nogales, M. 2005. A microhistological survey on the trees of a relict subtropical laurel forest from the Macaronesian Islands as a base for assessing vertebrate plant diet. - Botanical Journal of the Linnean Society 148: 409–426.

Marrero, P. et al. 2009. Extraction of high-quality host DNA from feces and regurgitated seeds: a useful tool for vertebrate ecological studies. - Biological Research 42: 147–151.

Olesen, J.M. et al. 2007. The modularity of pollination networks. - Proceedings of the National Academy of Sciences USA 104: 19891–19896.

Olesen, J.M. et al. 2011. Missing and forbidden links in mutualistic networks. - Proceedings Biological Sciences-The Royal Society 278: 725–732.

R Development Core Team. (2018). R: A Language and Environment for Statistical Computing. R Foundation for Stat. Comput., Vienna, Austria.

Snow, B. K. and Snow, D. W. 1988. Birds and berries. - T. and A.D. Poyser. Calton, UK.

Ulrich, W. et al. 2009. A consumer's guide to nestedness analysis. - Oikos 118: 3–17.

Weiss, N.A. 2015. wPerm: Permutation Tests. R package version 1.0.1. https://CRAN.R-project.org/package=wPerm