

Optimisation of a pollen DNA metabarcoding method for diet analysis of flying-foxes (*Pteropus* spp.)

Karen L. Bell ^{A,B,E}, Kathryn L. Batchelor^A, Matt Bradford^C, Adam McKeown^D,
Stewart L. Macdonald ^C and David Westcott^C

^ACSIRO Health and Biosecurity, Floreat, WA 6014, Australia.

^BUniversity of Western Australia, School of Biological Sciences, Crawley, WA 6008, Australia.

^CCSIRO Land and Water, Atherton, Qld 4883, Australia.

^DCSIRO Land and Water, Waite Campus, Adelaide, SA 5064, Australia.

^ECorresponding author. Email: karen.bell@uwa.edu.au

Abstract. Determining the diet of flying-foxes can increase understanding of how they function as pollinators and seed dispersers, as well as managing any negative impacts of large roosts. Traditional methods for diet analysis are time consuming, and not feasible to conduct for hundreds of animals. In this study, we optimised a method for diet analysis, based on DNA metabarcoding of environmental DNA (eDNA) from pollen and other plant parts in the faeces. We found that existing eDNA metabarcoding protocols are suitable, with the most useful results being obtained using a commercial food DNA extraction kit, and sequencing 350–450 base pairs of a DNA barcode from the internally transcribed spacer region (ITS2), with ~550 base pairs of the chloroplast rubisco large subunit (*rbcL*) as a secondary DNA barcode. A list of forage plants was generated for the little red flying-fox (*Pteropus scapulatus*), the black flying-fox (*Pteropus alecto*) and the spectacled flying-fox (*Pteropus conspicillatus*) from our collection sites across Queensland. The diets were determined to comprise predominantly Myrtaceae species, particularly those in the genera *Eucalyptus*, *Melaleuca* and *Corymbia*. With more plant genomes becoming publicly available in the future, there are likely to be further applications of eDNA methods in understanding the role of flying-foxes as pollinators and seed dispersers.

Keywords: pollen, DNA metabarcoding, DNA barcoding, environmental DNA, *Pteropus*, nectar, Myrtaceae, plant-animal interactions, diet, foraging.

Received 19 October 2020, accepted 4 October 2021, published online 28 October 2021

Introduction

Flying-foxes (*Pteropus* spp.) provide important ecosystem services as pollinators and seed dispersers (Marshall 1983, 1985; Fleming *et al.* 2009). However, as fruit feeders, flying-foxes can also be an agricultural pest (Aziz *et al.* 2021). They are increasingly roosting in urban areas, where they impact on amenity through noise, odour, and faeces (Tait *et al.* 2014), and act as reservoirs for pathogens of veterinary and medical importance (Halpin *et al.* 1999). To manage the conflicting negative and positive sociocultural values surrounding flying-foxes requires an understanding of what drives their movement and feeding behaviour.

Determining the diet of flying-foxes can inform on what plant species are attracting them to particular locations. Traditionally, this is done by radio-tracking (Palmer *et al.* 2000; Markus and Hall 2004), satellite tracking (Tidemann and Nelson 2004), direct observation of feeding behaviour in the field (Markus and Hall 2004), microscopic examination of

plant parts in excrement (Parry-Jones and Augee 2001; Parsons *et al.* 2006), or a combination of all these methods (Palmer *et al.* 2000). All of these techniques are time consuming and costly if large numbers of data points need to be obtained. Microscopic examination may be unable to provide good taxonomic resolution if the required diagnostic characters are not present, and may completely miss species that are present in liquid form only (Aziz *et al.* 2017).

New environmental DNA (eDNA) techniques have shown promise for the identification of plant species in the faeces of nectivorous bats (Aziz *et al.* 2017; Lim *et al.* 2018a, 2018b). These techniques are based on DNA barcoding, a method that uses the DNA sequence of a short, standardised gene region, compared with a database of the same gene region from known species (Hebert *et al.* 2003). Taxonomic resolution is usually high, generating species-level matches in 70–90% of taxa depending on the barcode(s) being used (CBOL Plant Working Group 2009; Chen *et al.* 2010). DNA metabarcoding uses high

throughput DNA sequencing to identify all species in a mixture using DNA barcoding (Cristescu 2014). Combined with DNA extraction methods that include a step that lyses pollen grains, these methods can taxonomically identify mixtures of various plant parts in faeces. While DNA metabarcoding is somewhat expensive on a per sample basis, this approach overall is likely less expensive and time consuming and permits processing of a higher number of samples, relative to microscopic identification (Bell *et al.* 2019). Moreover, these methods are well suited to flying-foxes that roost in very large numbers, making sample collection efficient. Limitations to these methods include the need to develop a reference database for the study system, the need to confirm that the selected barcode(s) are useful for the taxonomic groups present in the study system, and the inability to determine which plant part was consumed without information from other sources.

High diversity ecosystems, such as tropical rainforests, present particular challenges to DNA barcoding, due to the large number of species that need to be included in reference databases (Parmentier *et al.* 2013; Costion *et al.* 2016; Lima *et al.* 2018). In Australia, ~53% of plant species have been included in reference databases, and only 14% of species have data for the three core DNA barcodes, *matK*, *rbcL*, and ITS2 (Dormontt *et al.* 2018). These limitations could be overcome by generating reference sequences for documented food plants for the target species in the region of interest, as recommended by Chan *et al.* (2021). This process could be aided by publicly available databases of bat–plant interactions (Aziz *et al.* 2021). A further challenge to plant DNA barcoding in Australia comes from the large number of species in the family Myrtaceae, particularly in *Eucalyptus* and other genera in the tribe Eucalypteae. The genus and species delimitations within the Eucalypteae are the subject of active research (Brooker and Kleinig 2004; Centre for Australian National Biodiversity Research 2020), and introgression has led to a mismatch between phylogenies inferred from chloroplast DNA (the source of most barcode markers), nuclear rDNA (also frequently used as DNA barcodes), and taxonomic classification (Schuster *et al.* 2018). A particular issue with chloroplast DNA is the tendency for geographic sharing of chloroplast clades across taxonomic groups (Schuster *et al.* 2018). A particular issue with the ITS2 barcode is the presence of multiple paralogues, a problem in several taxonomic groups, and well documented in the Eucalypteae (Bayly and Ladiges 2007), although this may be less of an issue with high throughput sequences as all paralogues can be sequenced simultaneously. It is unknown to what extent these processes affect the accuracy of different DNA barcodes in the identification of species from Australian tropical rainforests.

Here we have developed and described a protocol for DNA metabarcoding of faecal samples, generated reference sequences for likely diet plant species of flying-foxes in Queensland, Australia, to add to reference sequence databases, and used this method and databases to document the diet of flying-foxes. Specifically, we determined whether existing DNA metabarcoding protocols are suitable; assessed the relative performance of different DNA extraction methods and DNA barcode primers for amplification and sequencing of

faecal eDNA; and determined the accuracy of different DNA barcodes for identification to species level, through the generation of new DNA barcode reference sequences. Finally, we show that a list of food plant species can be generated for multiple flying-fox species across a broad range of locations.

Methods

Samples

We collected 276 faecal samples from 11 roosts of the little red flying-fox (*Pteropus scapulatus*), three roosts of the black flying-fox (*Pteropus alecto*), two of which overlapped with *P. scapulatus*, and one roost of the spectacled flying-fox (*Pteropus conspicillatus*) across Queensland between July 2017 and November 2019 (Fig. 1; Supplementary Table S1). Faecal samples were taken directly from captured bats using a sterile swab, and new plastic sheets were placed on the ground below roosts to collect additional faecal samples without handling animals. Faecal samples were generally collected before 12 noon to ensure adequate quantity of fresh sample material. However, this was not always possible in remote roosts. Faecal pellets and swabs were placed in sterile 1.5 mL microcentrifuge tubes with 95% ethanol to preserve plant DNA and to neutralise any pathogens. Negative controls were also taken in the field, and comprised a sterile swab placed into a sterile 1.5 mL tube of ethanol at the sample collection site.

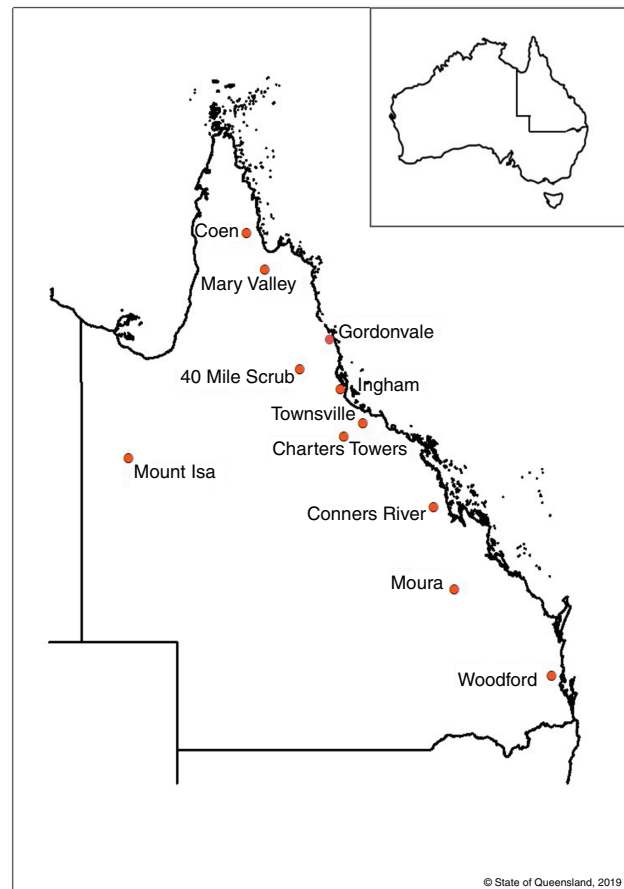


Fig. 1. Faecal sample collection locations.

Full details of samples can be found in Supplementary Table S1.

Leaf samples were collected from any tree species in flower or fruit near sampling sites, to obtain reference sequences of likely food species. Leaves were dried and preserved in sealed bags of silica gel. Species were identified by CSIRO botanists. Full details of leaf samples can be found in Supplementary Table S2.

DNA extraction

Samples were preprocessed by vortexing to dislodge as much of the sample from the swab as possible and to homogenise the sample, then centrifuged at 15000g for 2 min, after which the ethanol was set aside in a separate tube. Up to 200 mg of the sample was then transferred to a new tube containing 200 µL of ultrapure water, scraping the swab against the side of the tube to transfer as much of the sample as possible. The remaining sample was reunited with the original ethanol and stored at -80°C, to retain any residual DNA for future work.

Following a preliminary method comparison between two different kits (see Supplementary Methods and Results, Table S3), DNA was extracted using the Nucleospin DNA Food Kit (Macherey-Nagel, Düren, Germany), following the manufacturer's protocol with the following modifications: samples were homogenised using new, sterile 3 mm tungsten carbide beads for 1 min in a TissueLyser II (Qiagen, Hilden, Germany). Following homogenisation, we then followed the manufacturer's protocol for extraction of DNA using the NucleoSpin DNA Food Kit except the volume of Elution Buffer CE was decreased to 50 µL to increase the concentration of the final extraction. A negative control (no faecal sample added) was included in each batch of 12–24 DNA extractions. Leaf DNA extraction was conducted using the DNeasy Plant Kit (Qiagen). Dry leaf was homogenised in a TissueLyser II for 1 min with two 3 mm tungsten carbide beads, cleaned with 0.4M HCl in between samples. This step was repeated 2–3 times until the leaf was powdered. If the leaf was still not powdered after three repetitions, it was ground using a sterilised mortar and pestle. We extracted DNA from the dry, powdered leaf following the manufacturer's protocol. We used only the DNA from the first 100 µL elution to maximise the concentration of DNA going into PCR. A negative control was included in each batch of DNA extractions.

To ensure that our methods were able to accurately detect and identify plant species in mixtures of similar species-richness to faecal samples, we created mock communities. Each mock community contained five plant species, from leaf DNA extractions, in approximately equal concentrations. Mock Community 1 comprised *Ficus racemosa* (Moraceae), *Bauhinia hookeri* (Fabaceae), *Eucalyptus platyphylla* (Myrtaceae), *Corymbia clarksoniana* (Myrtaceae), and *Melaleuca leucadendra* (Myrtaceae). Mock Community 2 comprised *Syncarpia glomulifera* subsp. *glomulifera*, *Eucalyptus camaldulensis*, *Melaleuca bracteata*, *Melaleuca viridiflora*, *Melaleuca nervosa* (all Myrtaceae). Mock Community 3 comprised *Eucalyptus platyphylla* (Myrtaceae), *Corymbia torelliana* (Myrtaceae), *Brachychiton bidwillii*

(Malvaceae), *Terminalia playtphylla* (Combretaceae), and *Melaleuca viridiflora* (Myrtaceae). Mock Community 4 comprised *Eucalyptus camaldulensis* (Myrtaceae), *Eucalyptus tereticornis* (Myrtaceae), *Alyxia ruscifolia* (Apocynaceae), *Syzygium suborbiculare* (Myrtaceae), and *Eucalyptus megasepala* (Myrtaceae). All of these species were present in the reference databases for both *rbcL* and ITS2.

DNA sequencing

The concentration of each faecal or leaf DNA extraction was determined using the Qubit 1X dsDNA HS Assay Kit on a Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Samples with a concentration greater than 20 ng/µL were diluted to 10–20 ng/µL before amplification and sequencing.

Two DNA barcode markers were used, ITS2 and *rbcL*. Following a preliminary trial of two different primer pairs for each marker (Supplementary Methods and Results, Table S4), we conducted PCR amplification using the primers in Table 1. The two DNA barcode markers were amplified in separate reactions for each sample. We used an indexing strategy that gave each sample a unique combination of two indices, that were used for both DNA barcodes. The PCR reactions contained primers at a final concentration of 200 nM, 12.5 µL of KAPA HiFi ReadyMix (KAPA Biosystems, Boston, MA, USA), and 8.5 µL of the template DNA (up to 20 ng/µL) in a 25 µL reaction. To increase the chance of detecting all species in the mixture, each DNA extraction was included in three PCR reactions that were amplified separately, i.e. the reaction contents described above were divided between three PCR tubes. The PCR conditions included an initial period of heat activation for 3 min at 95°C; followed by 35 cycles of 30 s at 95°C, 30 s at 55°C, and 1 min at 72°C; followed by a final extension of 10 min at 72°C and then held at 10°C. For the short fragments, following a gradient PCR to determine optimal annealing temperature, PCR was conducted with an initial denaturing step of 95°C for 3 min; followed by 35 cycles of: 95°C for 30 s, 55°C for 30 s, 72°C for 1 min; and a final extension at 72°C for 10 min. A negative control (with ultrapure water as DNA template) was included in each PCR plate. We used our constructed mock communities as positive controls, and these were amplified in duplicate with different indices.

After amplification, PCR products were purified with Agencourt AMPure XP magnetic beads (Beckman Coulter, Danvers, MA, USA) to remove any remaining primers and PCR artefacts. The presence of a PCR product of the correct size was assessed by running 2 µL on a QIAxcel (Qiagen), using the 15bp–3kb alignment marker and 50bp–800bp size marker. The DNA concentration was quantified using the Qubit 1X dsDNA HS Assay Kit on a Qubit 4 Fluorometer. Within each sequencing run, all PCRs were pooled at equal concentrations and run on a single flow cell of the Illumina MiSeq instrument by AGRF (Melbourne), using a 2 × 250-cycle paired-end run for the first sequencing run, and a 2 × 300-cycle paired-end run for subsequent batches. Each sample had an index combination that was unique across the entire project to avoid cross-contamination between sequencing runs. A subset of eight

Table 1. PCR and sequencing primers used in this study

PCR product lengths include the fusion primers. Illumina tag is shown in bold; index, pad and linker sequences in plain text; and marker-specific primer in *italics*

Region	Primer name	Primer type	Primer sequences (index sequence shown as XXXXXXXXX)	References
ITS2 long (450–550 bp PCR product)	ITS-S2F-fusion	PCR	AATGATACGGCGACCACCGAGATCTACAC XXXXXXXXXXCCTGGTGTGCTGGT <i>ATGCGATACTTGGTGTGAAT</i>	Chen <i>et al.</i> (2010); fusion primer as in Sickel <i>et al.</i> (2015)
	ITS4R-fusion	PCR	CAAGCAGAAGACGGCATACGAGAT XXXXXXXXXXAGTCAGTCAGCC <i>TCCTCCGCTTATTGATATGC</i>	White <i>et al.</i> (1990); fusion primer as in Sickel <i>et al.</i> (2015)
	Read 1 ITS	Sequencing	CCTGGTGTGCTGGTATGCGATACTTGGTGTGAAT	Sickel <i>et al.</i> (2015)
	Read 2 ITS	Sequencing	AGTCAGTCAGCCTCCTCCGCTTATTGATATGC	Sickel <i>et al.</i> (2015)
	Index ITS	Sequencing	GCATATCAATAAGCGGAGGAGGCTGACTGACT	Sickel <i>et al.</i> (2015)
<i>rbcL</i> long (~600 bp PCR product)	rbcL2-fusion	PCR	AATGATACGGCGACCACCGAGATCTACA XXXXXXXXXXCCTGGTGTGCTGGT <i>TGGCAGCATTTCGAGTAACTC</i>	Palmieri <i>et al.</i> (2009); fusion primer as in Bell <i>et al.</i> (2017b)
	rbcLaR-fusion	PCR	CAAGCAGAAGACGGCATACGAGAT XXXXXXXXXXAGTCAGTCAGCC <i>GTAAATCAAGTCCACCRGC</i>	Kress <i>et al.</i> (2009); fusion primer as in Bell <i>et al.</i> (2017b)
	Read 1 <i>rbcL</i>	Sequencing	CCTGGTGTGCTGGTTGGCAGCATTTCGAGTAACTC	Bell <i>et al.</i> (2017b)
	Read 2 <i>rbcL</i>	Sequencing	AGTCAGTCAGCCGTAAATCAAGTCCACCRGC	Bell <i>et al.</i> (2017b)
	Index <i>rbcL</i>	Sequencing	CGYGGTGGACTTGATTTTACGGCTGACTGACT	Bell <i>et al.</i> (2017b)

index combinations (two for each batch) was not used for a sample, and was demultiplexed by AGRF as a sequencing negative control.

Taxonomic identification

For each sequenced sample, the ITS2 and *rbcL* sequencing reads were sorted into separate fastq files using scripts provided by Anya Cutler (Emory University, USA) for analysis in R (R Core Team 2016). This script uses the first four base-pairs of the expected sequence to identify reads as either ITS2 or *rbcL* and sort them into separate files. The dada2 package in R (Callahan *et al.* 2016) was used to remove likely sequencing errors and generate a set of unique DNA sequences. For the subset of samples in the preliminary run, analysed from long amplicons with 250 bp paired-end sequences, the *rbcL* amplicon was too long for forward and reverse sequences to merge. Therefore, these samples were analysed based on the single reads without merging, and in subsequent runs we used 300 bp paired-end sequences. Sequences that were likely to be the result of contamination were flagged using the decontam package (Davis *et al.* 2018) in R, using the prevalence method which compares presence/absence of sequences in samples and negative controls. These were removed from further analyses. To determine if adequate sequencing depth had been used, species accumulation curves were generated for a subset of 10 samples from one sampling event (*P. scapulatus*, Charters Towers, September 2017) using the vegan package in R (Oksanen *et al.* 2019), with Amplicon Sequence Variants (ASVs; unique DNA sequences) in place of species and number of sequence reads as units.

Sequences from leaf samples were separated from the dada2 output. Unique ITS2 sequences that were represented by at least 10% of the total sequence reads in the sample were

used as reference sequences for taxonomic identification, as these may represent multiple paralogues, or alternate diploid alleles. Only one *rbcL* sequence from each leaf sample was retained as a reference sequence, as this locus is haploid and not expected to have paralogous copies. Remaining sequences were compared with the NCBI database (www.ncbi.nlm.gov) using a BLAST search to confirm that they were the designated species and not contaminants. For *rbcL*, we appended our new sequences to the database of Bell *et al.* (2017b), and formatted the database for the dada2 functions assignSpecies and assignTaxonomy. The revised *rbcL* database is available from <https://doi.org/10.6084/m9.figshare.c.5504193.v1>. A new reference database was generated containing all publicly available ITS2 sequences. Flowering plant ITS2 sequences were obtained from the NCBI database using the following search terms: 'Magnoliopsida'[Organism] AND ITS2[All Fields]. The program ITSx was used to extract 5.8S+ITS2 sequences from longer sequences (Bengtsson-Palme *et al.* 2013). Publicly available reference databases, such as the NCBI database, rely on researchers to perform their own quality control before submitting sequences, and contaminating sequences are known to occur in the database (Breitwieser *et al.* 2019). To identify any erroneous sequences that could cause problems for our analysis, we conducted BLAST searches of the 50 most abundant ITS2 ASVs and 20 most abundant *rbcL* ASVs from our eDNA sequencing. Sequences were removed from our reference database if they were in the top 200 matches for one of these ASVs, but with taxonomy fields on NCBI that were outliers relative to all other sequences in the top 200 matches (ITS2: FM887018 *Terminalia bellerica*; *rbcL*: JX856776 *Strychnos potatorium*, KU564752 *Actephilia sessifolia*, KU564758 *Antirhea putaminosa*, MH598849 *Datura stramonium*, JX856673 *Callistemon polandii*, KU564753 *Alectryon*

pubescens, KU564797 *Eupotamia bennettii*, KU564800 *Freycinetia excelsa*, KU564801 *Freycinetia scandens*, and KU564874 *Sophora fraseri*). There are likely other erroneous sequences remaining in our database, but these are not a close match with our high-frequency ASVs, and so are unlikely to influence our results. The database was then formatted for the dada2 functions `assignSpecies` and `assignTaxonomy`, and our reference leaf sequences were appended to this database. The ITS2 database is available from <https://doi.org/10.6084/m9.figshare.c.5203703.v1>.

Matrices of DNA sequence by sample generated by dada2 analysis of each sequencing run were merged into a single data matrix for each marker (with the unmerged *rbcL* long reads from the preliminary sequencing run treated as a different marker, due to the inability to merge reads) following the dada2 instruction manual. Samples containing fewer than 1000 total sequencing reads were removed from further analysis. Identification to higher taxonomic level (up to genus) was conducted using the `assignTaxonomy` function in dada2. Species-level identification was determined using the `assignSpecies` function, which requires exact matches and can output multiple exact matches if present. Because ITS2 sequences from the same species typically vary by up to 1% (Chen *et al.* 2010), we also considered any ITS2 sequence on the NCBI database or among our reference sequences with at least 99% identity to be a possible species-level match. Only exact matches were considered for the less variable *rbcL* barcode, and where multiple exact matches occurred, all were considered plausible. Finally, for sequences with multiple

plausible species-level identifications (multiple matches with at least 99% identity with ITS2 or multiple exact matches with *rbcL*), based on sequence similarity, species distributions and, where appropriate, flowering and fruiting records and observations were used to shorten the list to the most likely species. Specific observation of flowering and fruiting allowed us to confidently identify the plant part in most collections, particularly in woodland settings where species dominance is high. In rainforest settings with low species dominance, this is more difficult.

Results

DNA sequencing results

The preliminary run of Illumina MiSeq generated 2 304 061 paired-end 250 bp sequencing reads. Of these, 1 854 641 were from the 25 samples sequenced with long PCR products. After separating sequences into ITS2 and *rbcL* reads, and analysis via the dada2 bioinformatics pipeline, there remained 456 195 long ITS2 reads, 139 295 long *rbcL* forward reads, and 139 152 *rbcL* reverse reads. Two subsequent Illumina MiSeq runs of long ITS2 and long *rbcL* PCR products generated 11 608 382 and 11 490 757 paired-end 300 bp sequencing reads. After separation of ITS2 and *rbcL* sequences, filtering and cleaning with dada2, and contaminant removal with decontam, 3 129 419 and 3 261 146 ITS2 sequences and 1 973 615 and 3 579 131 *rbcL* sequences remained, for each sequencing run respectively (Supplementary Table S5). Of these, 99.3% of ITS2 sequences were able to be identified to family and 98.9%

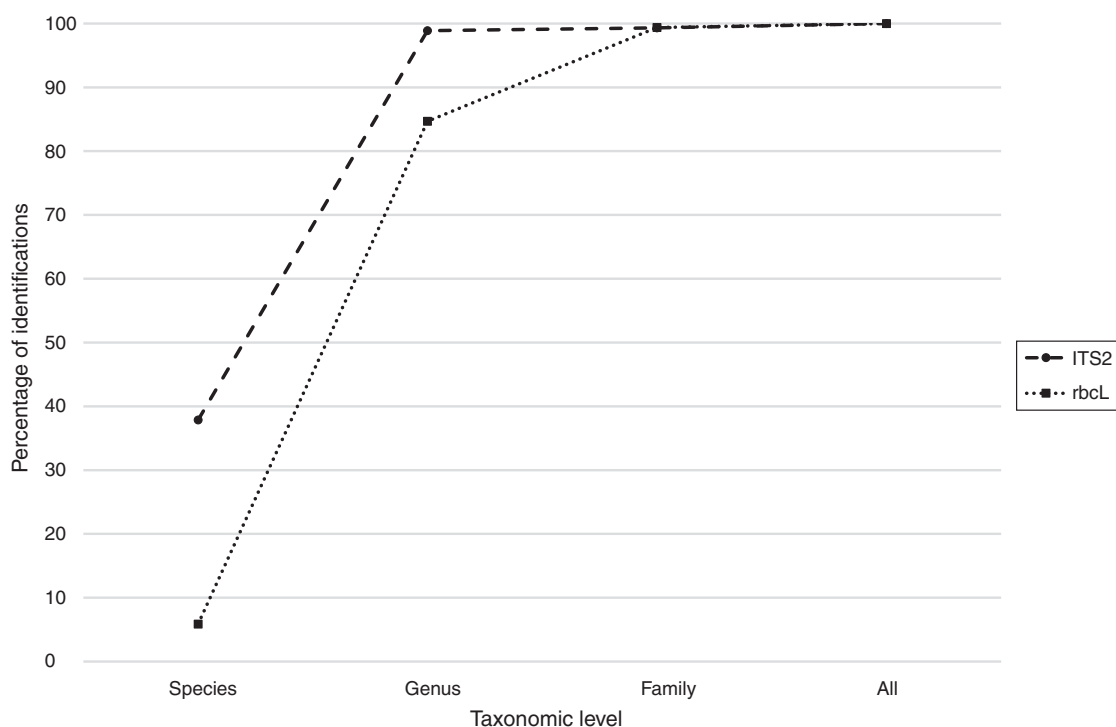


Fig. 2. Taxonomic level of identification with ITS2 and *rbcL*. Percentage of Illumina MiSeq reads across all samples identified to each taxonomic level using long DNA sequences (350–450 bp of ITS2; ~550 bp of *rbcL*), with taxonomic matching conducted using the `assignTaxonomy` and `assignSpecies` functions in dada2.

to genus with the `assignTaxonomy` function, while 37.9% were identified to exactly one species with the `assignSpecies` function. The *rbcL* sequences tended to have lower taxonomic resolution, with 99.4% able to be identified to family, 84.7% to genus, and 5.9% to exactly one species (Fig. 2).

The decontam function identified 18 and 13 ITS2 sequences as contaminants in sequencing runs 2 and 3, respectively, representing 64 728 and 78 371 sequencing reads (2.1 and 2.4% of input reads). For *rbcL*, 4 and 6 sequences were identified as contaminants, representing 60 802 and 50 793 sequencing reads (3.1 and 1.4% of input reads), respectively. These sequences were removed from subsequent analyses, but, due to their low proportions in mixtures, they would have been unlikely to affect our results if they were retained.

Species accumulation curves of a subset of 10 samples, considering ASVs as species and sequencing reads as sampling units, showed most ASVs within a sample were detected with fewer than 5000 sequencing reads (Fig. 3). For many samples, this could be achieved with 1000 sequencing reads or fewer.

Taxonomic identification

Species known to be present in mock communities were almost always able to be detected with both ITS2 and *rbcL*, but were not always identified to the species level, despite all being present in reference databases (Fig. 4). Using ITS2, all species were detected except for *Ficus racemosa* (Moraceae) in Mock Community 1 and *Melaleuca nervosa* in Mock Community 2. We failed to amplify *F. racemosa* ITS2 on its own during the reference database supplementation step, which suggests there could be a primer mismatch. Sequence of ITS2 is available for this species on NCBI, so we were still able to include it in the reference database. *Melaleuca nervosa* was likely indistinguishable from *M. viridiflora*, which was also present, rather than being undetectable. *Eucalyptus camaldulensis* and *E. tereticornis* were indistinguishable but detected in all samples where present. Using *rbcL*, all species were detected except for *Corymbia clarksoniana* in Mock Community 1. *Brachychiton bidwillii* (Malvaceae) in Mock

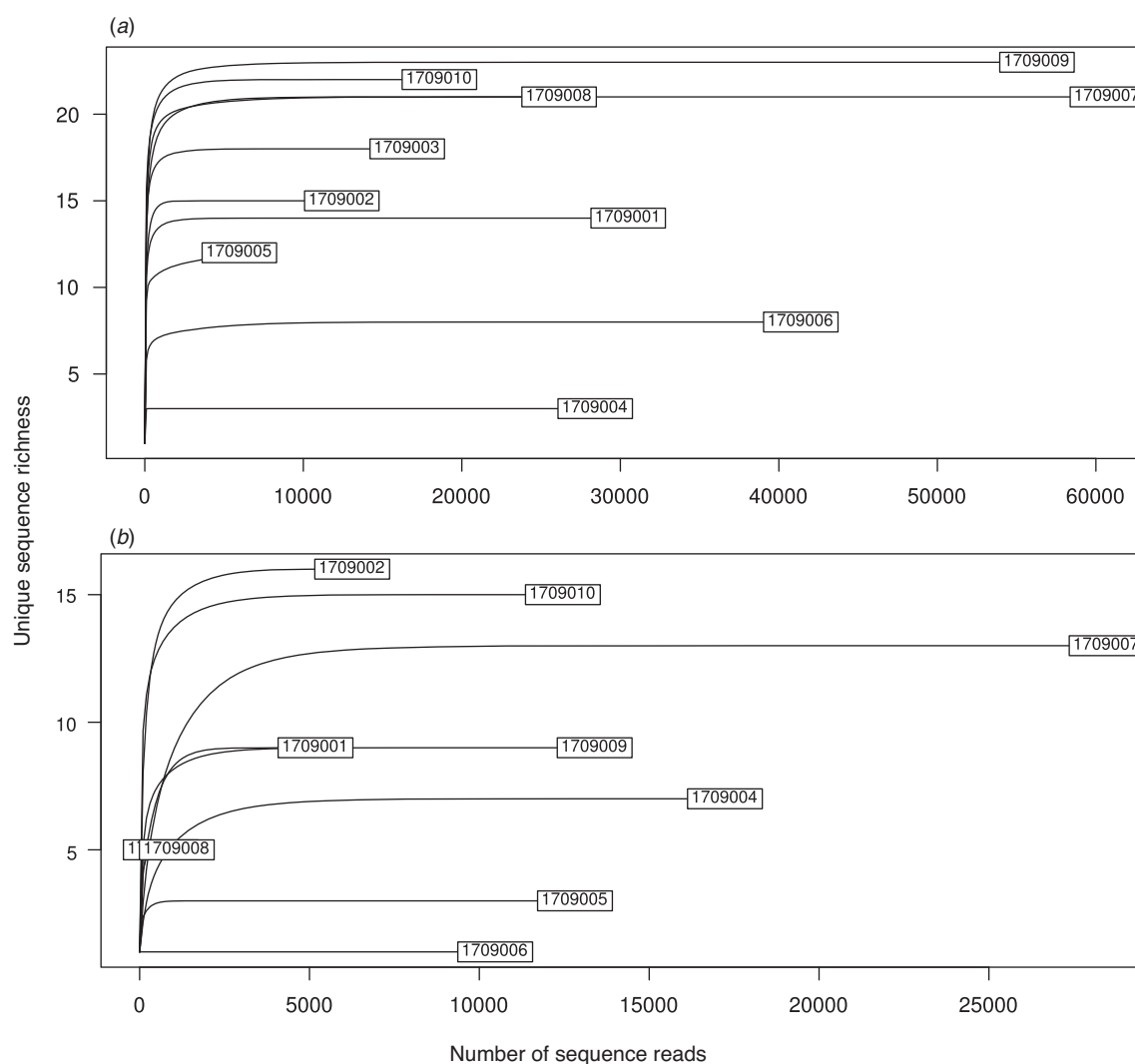


Fig. 3. Species accumulation curves of amplicon sequence variants (ASVs) from each sample, after quality control and filtering with `dada2`, as a function of the number of sequencing reads. (a) ITS2; (b) *rbcL*.

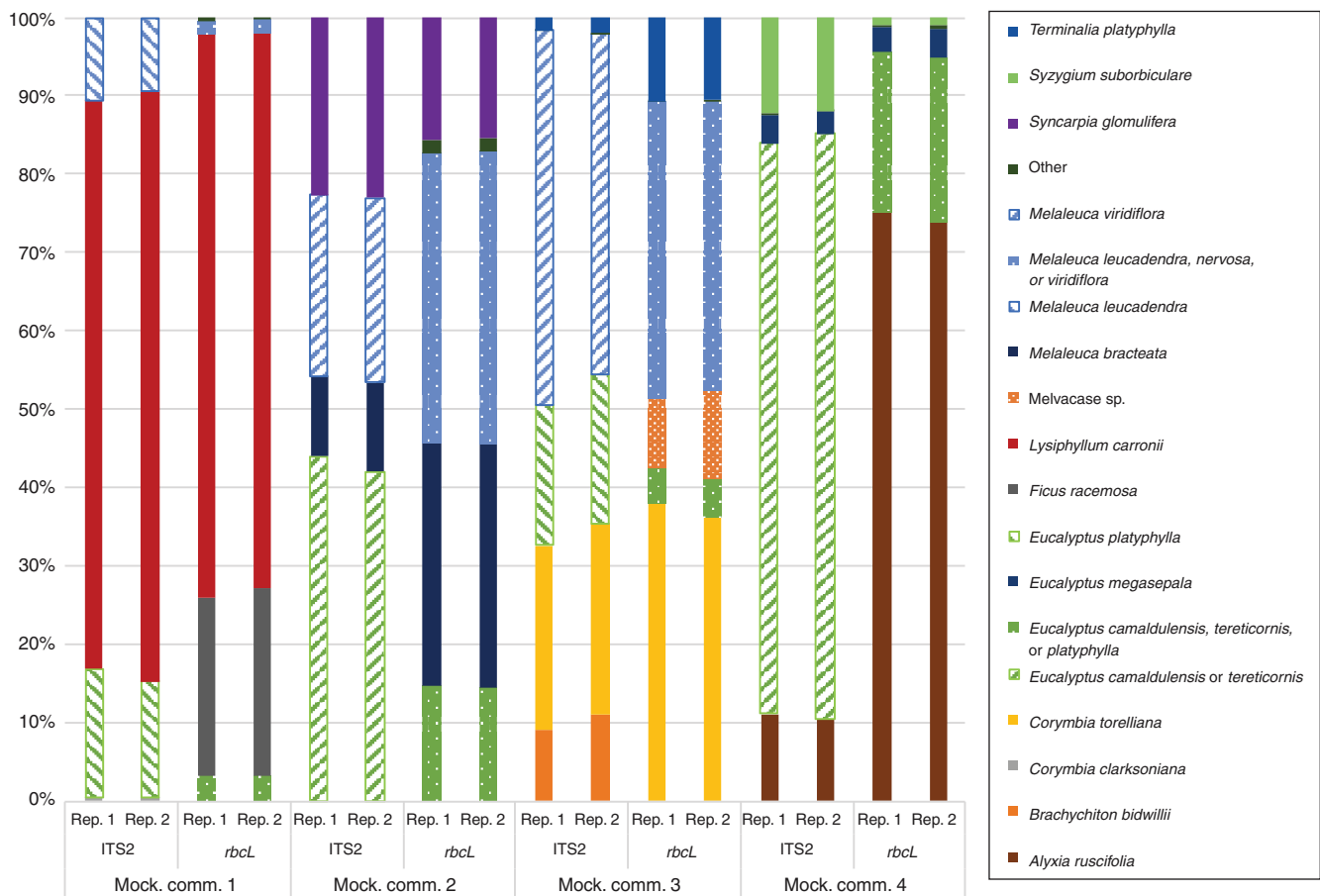


Fig. 4. Taxonomic identifications of mock communities. Four mock communities of known species composition, as described in the text, were sequenced using long amplicons of ITS2 and *rbcL* and the sequences were identified using the assignTaxonomy and assignSpecies functions in dada2.

Community 3 was detected but misidentified as *Ochroma* sp. (Malvaceae). *Eucalyptus camaldulensis*, *E. tereticornis* and *E. platyphylla* were indistinguishable but detected where present. Likewise, *M. leucadendra*, *M. nervosa* and *M. viridiflora* were indistinguishable but detected where present.

Species and genera from the family Myrtaceae were the predominant components identified in faecal samples. The most abundant ITS2 ASV in faecal samples was identified as *Eucalyptus camaldulensis* (Myrtaceae). Other abundant ASVs were identified as *Eucalyptus tereticornis*, *Melaleuca viridiflora* (Myrtaceae), *Eucalyptus* sp., and *Melaleuca leucadendra* (Table 2). The most abundant *rbcL* ASV was identified as *Melaleuca* sp. Other abundant ASVs were identified as *Eucalyptus* sp., Myrtaceae sp., *Bauhinia* sp. (Fabaceae), and *Corymbia* sp. (Myrtaceae). Some ASVs matched plant species that were wind-pollinated, mostly in the Poaceae and Casuarinaceae. These were likely contaminants from the field or laboratory and were usually present only in a single sample. The most frequently detected wind-pollinated species was *Axonopus* (Poaceae), which was identified from ITS2 in 11 samples and from *rbcL* in 14 samples. A full list of ITS2 and *rbcL* taxonomic identifications from faecal samples,

can be found in the Supplementary Material (Tables S5 and S6, respectively). These data, along with data from additional faecal samples, is being used to understand flying-fox foraging behaviour and the results of this work will be published elsewhere.

The taxonomic composition of a subset of samples is shown in Fig. 5. Within the same samples, a similar taxonomic composition was determined with both DNA barcodes. In most cases, the same genera were detected with both ITS2 and *rbcL*, but with ITS2 taxonomic identification was more likely to be to species-level, or a small subset of the species within a genus. Occasionally a taxon was detected with only one or the other marker. For example, *Musa* sp. (Musaceae) in 1709005 and *Vitex* sp. (Lamiaceae) in 1709010 were identified only from *rbcL* sequences, and *Shotia brachypetala* (Fabaceae) in 1709010 was identified only with ITS2 sequences.

Discussion

We were successful in identifying food plants of flying-foxes in Queensland, Australia, using faecal eDNA metabarcoding. We found that standard pollen DNA metabarcoding methods were effective, with no need to use a shorter DNA barcode to

Table 2. Taxonomic identification of commonly occurring amplicon sequence variants (ASVs; unique DNA sequence variants) in flying-fox faecal samples

Taxonomic identifications were based on a combination of the following method(s): (1) the taxonomic classifications generated by dada2 using the assignTaxonomy and assignSpecies functions, (2) 99% sequence identity at the species level to sequences on NCBI and our reference sequences, and/or (3) knowledge of geographic range and flowering phenology of the species				
Amplicon sequence variant (ASV)	Dada2 identification	(a) Identifications using ITS2 Interpreted identification	Taxonomic identification method(s)	No. of sequencing reads
ITS2seq0002	<i>Eucalyptus camaldulensis</i>	<i>Eucalyptus camaldulensis</i>	assignSpecies	359 198
ITS2seq0001	<i>Eucalyptus camaldulensis</i> / <i>tereticornis</i>	<i>Eucalyptus camaldulensis</i> / <i>tereticornis</i>	assignSpecies	348 699
ITS2seq0004	<i>Melaleuca viridiflora</i>	<i>Melaleuca viridiflora</i>	assignSpecies	286 995
ITS2seq0005	<i>Eucalyptus</i> sp.	<i>Eucalyptus</i> sp.	assignTaxonomy	252 880
ITS2seq0007	<i>Melaleuca leucadendra</i>	<i>Melaleuca leucadendra</i>	assignSpecies	214 010
ITS2seq0003	<i>Eucalyptus camaldulensis</i> / <i>tereticornis</i>	<i>Eucalyptus camaldulensis</i> / <i>tereticornis</i>	assignSpecies	181 620
ITS2seq0008	<i>Eucalyptus platyphylla</i>	<i>Eucalyptus platyphylla</i>	99% sequence identity	143 180
ITS2seq0014	<i>Bauhinia</i> sp.	<i>Bauhinia</i> sp.	assignTaxonomy	136 970
ITS2seq0012	<i>Corymbia dallachiana</i>	<i>Corymbia dallachiana</i>	assignSpecies	136 796
ITS2seq0016	<i>Corymbia aparerrinja</i> / <i>dallachiana</i>	<i>Corymbia aparerrinja</i> / <i>dallachiana</i>	assignSpecies	110 566
ITS2seq0013	<i>Corymbia brachycarpa</i> / <i>clarksoniana</i> / <i>erythrophloia</i> / <i>nesophila</i> / <i>xanthope</i>	<i>Corymbia erythrophloia</i> / <i>clarksoniana</i> / <i>nesophila</i>	assignSpecies, geographic range and flowering	109 661
ITS2seq0006	<i>Eucalyptus platyphylla</i>	<i>Eucalyptus platyphylla</i>	assignSpecies	104 625
ITS2seq0009	<i>Melaleuca</i> sp.	<i>Melaleuca leucadendra</i>	99% sequence identity	88 475
ITS2seq0026	<i>Eucalyptus</i> sp.	<i>Eucalyptus platyphylla</i>	99% sequence identity	79 178
ITS2seq0033	<i>Erythrina</i> sp.	<i>Erythrina vespertilio</i>	99% sequence identity	74 257
ITS2seq0018	<i>Melaleuca</i> sp.	<i>Melaleuca leucadendra</i>	99% sequence identity	68 118
ITS2seq0025	<i>Melaleuca</i> sp.	<i>Melaleuca leucadendra</i>	99% sequence identity	64 699
ITS2seq0024	<i>Eucalyptus</i> sp.	<i>Eucalyptus platyphylla</i>	99% sequence identity	62 646
ITS2seq0021	<i>Corymbia dallachiana</i>	<i>Corymbia dallachiana</i>	assignSpecies	62 607
ITS2seq0022	<i>Eucalyptus coolabah</i> / <i>leucophylla</i> / <i>microtheca</i> / <i>ravaretiana</i>	<i>Corymbia microtheca</i> / <i>leptophleba</i>	assignSpecies, geographic range and flowering	59 161
(b) Identifications using <i>rbcl</i>				
Amplicon sequence variant (ASV)	Dada2 identification	Interpreted identification	Source of evidence	No. of sequencing reads
rbclSeq0002	<i>Melaleuca cajuputi</i> / <i>leucadendra</i> / <i>nervosa</i> / <i>quinquenervia</i> / <i>stenostachya</i> / <i>viridiflora</i>	<i>Melaleuca</i> sp.	assignSpecies	678 007
rbclSeq0001	<i>Eucalyptus camaldulensis</i> / <i>degupta</i> / <i>diversicolor</i> / <i>globulus</i> / <i>gomphocephala</i> / <i>grandis</i> / <i>major</i> / <i>pellita</i> / <i>platyphylla</i> / <i>tereticornis</i>	<i>Eucalyptus</i> sp.	assignSpecies	663 334
rbclSeq0004	Myrtaceae sp.	Myrtaceae sp.	assignTaxonomy	473 740
rbclSeq0003	<i>Eucalyptus cladocalyx</i> / <i>coolabah</i> / <i>crebra</i> / <i>globulus</i> / <i>lehmannii</i> / <i>leucophylla</i> / <i>melanophloia</i> / <i>melliodora</i> / <i>microtheca</i> / <i>polybractea</i> / <i>populnea</i> / <i>pruinosa</i> / <i>ravaretiana</i>	<i>Eucalyptus</i> sp.	assignSpecies	426 444
rbclSeq0007	<i>Bauhinia carronii</i> / <i>gilvum</i> / <i>hookeri</i>	<i>Bauhinia</i> sp.	assignSpecies	261 327
rbclSeq0006	<i>Corymbia citriodora</i> / <i>dallachiana</i> / <i>eximia</i> / <i>henryi</i> / <i>tessellaris</i> / <i>torelliana</i>	<i>Corymbia</i> sp.	assignSpecies	203 667
rbclSeq0008	Myrtaceae sp.	Myrtaceae sp.	assignTaxonomy	135 488
rbclSeq0010	<i>Vitex melicopea</i>	<i>Vitex melicopea</i>	assignSpecies	116 223
rbclSeq0015	<i>Pouteria chartacea</i> / <i>cotinifolia</i> / <i>gongripii</i> / <i>grandifolia</i> / <i>myrsinifolia</i> / <i>pohlmanniana</i> / <i>venulosa</i>	<i>Pouteria</i> sp.	assignSpecies	69 883

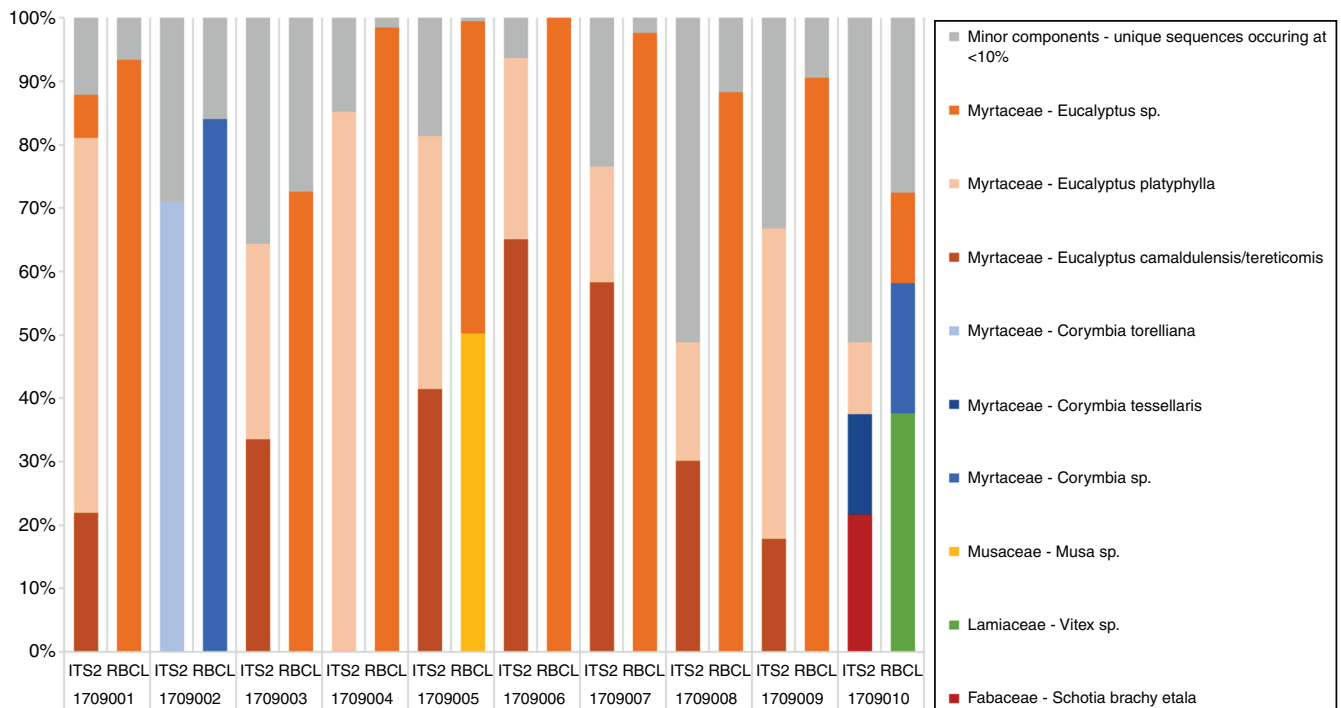


Fig. 5. Taxonomic composition of a subset of flying-fox faecal samples. Faecal samples were sequenced using long amplicons of ITS2 and *rbcL* and the sequences were identified using the `assignTaxonomy` and `assignSpecies` functions in `dada2`.

allow for DNA degradation. Using ITS2 and *rbcL* in combination enabled the detection of all species in mock communities, while each marker alone was able to detect most species. We found the taxonomic resolution to be finer for identifications based on ITS2, relative to *rbcL*. Most taxa were identified to species level with ITS2, whereas most taxa could not be discriminated below genus level with *rbcL*. Finally, we used this newly developed method to identify the forage plants for flying-foxes in Queensland, noting a predominance of Myrtaceae species.

Performance of standard pollen DNA metabarcoding methods

We adapted our method from a protocol previously developed for the identification of plant species in bee-collected pollen (Bell *et al.* 2017a). Faecal samples can contain enzymes involved in digestion, which may inhibit PCR. For this reason, we trialled an alternative DNA extraction kit, specifically designed for faecal samples. We extracted higher quantities of DNA with the method previously optimised for extraction of DNA from pollen, with no PCR failure. Faecal samples may also be expected to contain degraded DNA due to the digestion process. At least one previous study using DNA metabarcoding for diet analysis of a flying-fox used primers specifically designed to sequence a shorter DNA fragment than typical DNA barcodes (Aziz *et al.* 2017). However, we found that full-length DNA barcodes could be successfully amplified and sequenced. This is consistent with the results of Lim *et al.* (Lim *et al.* 2018a, 2018b) and Chan *et al.* (2021), who were able to successfully amplify and sequence faecal samples from

nectivorous and frugivorous bats with full-length DNA barcodes. The gut passage time of flying-foxes is usually fast, although some seeds can be retained for up to 24 h (Westcott *et al.* 2001; Aziz *et al.* 2021). Fast gut passage times may mean that the DNA is less degraded than faecal DNA of other species. Lack of degradation could also be because the majority of DNA in the faecal samples was present within pollen grains.

We found that, in combination, ITS and *rbcL* have been able to detect all species in a mixture. This is consistent with the results of Bell *et al.* (2019), where mixtures of known pollen species composition were all able to be identified using these two DNA barcodes. This combination of barcodes has not been previously tested in north-eastern Australian study systems. The standard chloroplast markers *rbcL* and *matK*, in addition to *trnH-psbA* as a supplementary marker (also from the chloroplast genome) have been successfully used for DNA barcoding in south-east Queensland (Shapcott *et al.* 2015). As a study system dominated by Myrtaceae species, particularly *Eucalyptus* and related genera, we may have expected some difficulties with DNA barcoding. Previous studies have shown that distantly related *Eucalyptus* species can share chloroplast genomes due to hybridisation and introgression (Schuster *et al.* 2018). For this reason, we supplemented the more traditional chloroplast barcode *rbcL*, with the nuclear ribosomal barcode ITS2. However, DNA barcoding via nuclear rDNA could also be complicated, due to paralogous gene copies. Combining the DNA barcoding data with field observations of flowering, and fruiting where appropriate, and a knowledge of plant species distributions, allowed us to overcome these issues. For example, *Eucalyptus tereticornis* and *E. camaldulensis*

partially overlap in range and flowering season and share ITS2 sequences. For most samples, assignment to species was possible based on the distribution of each species, then our flowering records.

Relative performance of ITS2 and rbcL

The taxonomic resolution was found to be higher for ITS2 than *rbcL*, with many ITS2 sequences being identified to species level (37.9%), and almost all (98.9%) identified to genus. On the other hand, only 5.9% of *rbcL* sequences were identified to species, and only 84.7% identified to genus. A higher rate of species-level identification with ITS2 compared with *rbcL* has been observed elsewhere in DNA metabarcoding of environmental samples, such as bee pollen (Bell *et al.* 2017a), and soil (Fahner *et al.* 2016). For single-species DNA barcoding, ITS2 has also been shown to have higher taxonomic resolution than *rbcL* (Chen *et al.* 2010; Li *et al.* 2011), especially if multiple sequences per species are included (Kolter *et al.* 2021), and even with shorter amplicons species-level resolution has been shown to be 86.1% (Moorhouse-Gann *et al.* 2018). For this reason, we recommend that both markers be used, or if only one marker is used that ITS2 be preferred over *rbcL* if high-level taxonomic resolution is required.

Food plants of Queensland flying-foxes

Species and genera from the family Myrtaceae were the predominant components identified in faecal samples. Frequently occurring species included *E. camaldulensis*, *E. tereticornis*, *M. viridiflora*, *M. leucadendra*, other *Eucalyptus* and *Melaleuca* spp., and *Corymbia* spp. Other studies have also shown flying-fox species to frequently include flowers of Myrtaceae species in their diets. For example, Parry-Jones and Augee (2001) found Myrtaceae pollen in the faeces of *Pteropus poliocephalus* from Sydney in all months of the year. This is likely due to the prevalence of Myrtaceae species in Australian vegetation, as well as the nectar and pollen resources provided by these species. The food plants identified in this study are examined in more detail in Bradford *et al.* (2020).

Shortcomings and priorities for future research

We validated our method using mock communities of known species composition, made from DNA extracted from leaf material, of species that were present in reference databases. Using both DNA barcodes in combination, all species could be detected. However, occasionally species were missed with one barcode or the other. If analysing only ITS2 data, it is possible that some species such as *Ficus* sp. may be missed, but this can be resolved by combining the ITS2 with *rbcL* sequences, or redesigning primers for ITS2 to improve amplification. Our mock community analysis demonstrated that if DNA is present in a faecal sample, then it can be detected with either or both of ITS2 and *rbcL*. However, we have not confirmed that there will be adequate DNA in faeces for all species eaten to be detected. When gut passage time is short, this could mean that only species consumed within the last few hours are detected. Finally, eDNA analysis alone cannot determine which plant part is being consumed. Combining eDNA analysis with microscopic analysis of faecal samples could assist with determining which plant part is

consumed, although if plant parts are eaten in liquid form, such as fruit juice or nectar, these may not be detectable in the faeces (Aziz *et al.* 2017). Combining eDNA analysis with knowledge of plant phenology should provide insight on whether flowers or fruits are being eaten, and behavioural observation would determine whether feeding could result in pollination or seed dispersal.

Both ITS2 and *rbcL* are unable to detect and identify all plant taxa to species level. However, they are currently our best options for taxonomic identification in species-rich ecosystems, based on their reasonable taxonomic resolution and reasonably complete reference databases. Several prospective methods for eDNA-based identifications of species mixtures, including faecal and pollen samples, may become more feasible in the future. Methods that sequence the whole genome or a reduced-representation of the whole genome, or that skim organellar (chloroplast and mitochondrion) sequences from whole genome data have demonstrated improvement in taxonomic resolution over single DNA barcodes as well as quantification of the proportions of species in mixtures (Gómez-Rodríguez *et al.* 2015; Bista *et al.* 2018; Lang *et al.* 2019; Ji *et al.* 2020). This improved resolution could even allow identification at intraspecific levels, allowing analyses of pollen dispersal between populations. However, for diet analysis in species-rich ecosystems, particularly where the target taxa are highly mobile, the amount of DNA sequence required for reference databases could make such studies prohibitively expensive. There are currently several programs underway aimed at sequencing near-complete genomes of a large representation of the planet's biodiversity (Lewin *et al.* 2018). As more genomic data become publicly available, and DNA sequencing technologies continue to improve, the use of whole genome methods for diet analysis will become more feasible, providing a powerful method for fine-scale taxonomic identification and quantification.

Conclusions

We have optimised an eDNA sequencing protocol for the DNA metabarcoding of flying-fox diet, as well as constructing a reference database to enable sequence matching to identified plant species. Methods based on eDNA have practical advantages over existing methods, as they can be scaled up to high-throughput analysis, with minimal increase in labour. These methods, especially when combined with observational methods, will be useful for research into the understanding of the role of flying-foxes in ecosystems, as well as understanding behaviours that can lead to human–flying-fox conflict. The opportunities for applying these methods will increase as they are further developed to allow finer scale taxonomic resolution, quantification, and population-level identification.

Data availability statement

Reference databases are available at: <https://doi.org/10.6084/m9.figshare.c.5504193.v1> and <https://doi.org/10.6084/m9.figshare.c.5203703.v1>. New reference sequences have been submitted to NCBI under accession numbers MW178132–MW178183 and MW139070–MW139223.

Conflicts of interest

The authors declare no conflicts of interest.

Declaration of funding

This research was funded by the Queensland Department of Environment and Science.

Acknowledgements

This research was permitted under Section 173P of the Queensland *Nature Conservation Act 1992* and Section r72 of the Western Australian *Biosecurity and Agricultural Management Act 2007*, using CSIRO animal ethics approvals AEC2016-17 and AEC2019-14. Botanical sampling in National Parks was conducted under permit no. WITK18679017. We thank Eric Vanderduys (CSIRO) for faecal and botanical field collections and Philipp Bayer (University of Western Australia) for assistance with development of the ITS2 reference database. The manuscript was improved with comments from Suzanne Metcalfe and three anonymous reviewers.

References

- Aziz, S. A., Clements, G. R., Peng, L. Y., Campos-Arceiz, A., McConkey, K. R., Forget, P.-M., and Gan, H. M. (2017). Elucidating the diet of the island flying fox (*Pteropus hypomelanus*) in Peninsular Malaysia through Illumina Next-Generation Sequencing. *PeerJ* **5**, e3176. doi:10.7717/peerj.3176
- Aziz, S. A., McConkey, K. R., Tanalgo, K., Sritongchuay, T., Low, M.-R., Yong, J. Y., Mildenstein, T. L., Nuevo-Diego, C. E., Lim, V.-C., and Racey, P. A. (2021). The critical importance of Old World fruit bats for healthy ecosystems and economies. *Frontiers in Ecology and Evolution* **9**, 641411.
- Bayly, M. J., and Ladiges, P. Y. (2007). Divergent paralogues of ribosomal DNA in eucalypts (Myrtaceae). *Molecular Phylogenetics and Evolution* **44**, 346–356. doi:10.1016/j.ympev.2006.10.027
- Bell, K. L., Fowler, J., Burgess, K. S., Dobbs, E. K., Gruenewald, D., Lawley, B., Morozumi, C., and Brosi, B. J. (2017a). Applying pollen DNA metabarcoding to the study of plant–pollinator interactions. *Applications in Plant Sciences* **5**, 1600124. doi:10.3732/apps.1600124
- Bell, K. L., Loeffler, V. M., and Brosi, B. J. (2017b). An rbcL reference library to aid in the identification of plant species mixtures by DNA metabarcoding. *Applications in Plant Sciences* **5**, 1600110. doi:10.3732/apps.1600110
- Bell, K. L., Burgess, K. S., Botsch, J. C., Dobbs, E. K., Read, T. D., and Brosi, B. J. (2019). Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. *Molecular Ecology* **28**, 431–455. doi:10.1111/mec.14840
- Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., Amend, A. S., Jumpponen, A., Unterseher, M., Kristiansson, E., Abarenkov, K., Bertrand, Y. J. K., Sanli, K., Eriksson, K. M., Vik, U., Veldre, V., and Nilsson, R. H. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution* **4**, 914–919. doi:10.1111/2041-210X.12073
- Bista, I., Carvalho, G. R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., Shokralla, S., Seymour, M., Bradley, D., Liu, S., Christmas, M., and Creer, S. (2018). Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Molecular Ecology Resources* **18**, 1020–1034. doi:10.1111/1755-0998.12888
- Bradford, M., Venz, M., Hogan, L., Smith, G., Bell, K. L., McKeown, A., Vanderduys, E., Macdonald, S. L., Ford, A., Eyre, T., and Westcott, D. (2020). The diet of the little red flying-fox (*Pteropus scapulatus*) in Queensland. In 'The little red flying-fox: ecology and management of Australia's most enigmatic flying-fox'. (Eds D. Westcott, et al.) pp. 245–277. A report to the Queensland Department of Environment and Science, CSIRO.
- Breitwieser, F. P., Perteu, M., Zimin, A. V., and Salzberg, S. L. (2019). Human contamination in bacterial genomes has created thousands of spurious proteins. *Genome Research* **29**, 954–960. doi:10.1101/gr.245373.118
- Brooker, M. I. H., and Kleinig, D. A. (2004). 'Field Guide to Eucalypts. Volume 3. Northern Australia.' (Bloomings Books: Melbourne, Australia.)
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**, 581–583. doi:10.1038/nmeth.3869
- CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12794–12797. doi:10.1073/pnas.0905845106
- Centre for Australian National Biodiversity Research (2020). 'EUCLID Eucalyptus of Australia.' 4th edn. (Centre for Australian Biodiversity Research (CANBR): Canberra.)
- Chan, A. A. Q., Aziz, S. A., Clare, E. L., and Coleman, J. L. (2021). Diet, ecological role and potential ecosystem services of the fruit bat, *Cynopterus brachyotis*, in a tropical city. *Urban Ecosystems* **24**, 251–263. doi:10.1007/s11252-020-01034-x
- Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y., and Leon, C. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* **5**, e8613. doi:10.1371/journal.pone.0008613
- Costion, C. M., Lowe, A. J., Rossetto, M., Kooyman, R. M., Breed, M. F., Ford, A., and Crayn, D. M. (2016). Building a plant DNA barcode reference library for a diverse tropical flora: an example from Queensland, Australia. *Diversity* **8**, 5. doi:10.3390/d8010005
- Cristescu, M. E. (2014). From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution* **29**, 566–571. doi:10.1016/j.tree.2014.08.001
- Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., and Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226. doi:10.1186/s40168-018-0605-2
- Dormontt, E. E., van Dijk, K.-j., Bell, K. L., Biffin, E., Breed, M. F., Byrne, M., Caddy-Retallac, S., Encinas-Viso, F., Nevill, P. G., Shapcott, A., Young, J. M., Waycott, M., and Lowe, A. J. (2018). Advancing DNA barcoding and metabarcoding applications for plants requires systematic analysis of herbarium collections – an Australian perspective. *Frontiers in Ecology and Evolution* **6**, 134. doi:10.3389/fevo.2018.00134
- Fahner, N. A., Shokralla, S., Baird, D. J., and Hajibabaei, M. (2016). Large-scale monitoring of plants through environmental DNA metabarcoding of soil: recovery, resolution, and annotation of four DNA markers. *PLoS ONE* **11**, e0157505. doi:10.1371/journal.pone.0157505
- Fleming, T. H., Geiselman, C., and Kress, W. J. (2009). The evolution of bat pollination: a phylogenetic perspective. *Annals of Botany* **104**, 1017–1043. doi:10.1093/aob/mcp197
- Gómez-Rodríguez, C., Crampton-Platt, A., Timmermans, M. J. T. N., Baselga, A., and Vogler, A. P. (2015). Validating the power of mitochondrial metagenomics for community ecology and phylogenetics of complex assemblages. *Methods in Ecology and Evolution* **6**, 883–894. doi:10.1111/2041-210X.12376

- Halpin, K., Young, P. L., Field, H., and Mackenzie, J. S. (1999). Newly discovered viruses of flying foxes. *Veterinary Microbiology* **68**, 83–87. doi:10.1016/S0378-1135(99)00063-2
- Hebert, P. D. N., Cywinska, A., Ball, S. L., and DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings. Biological Sciences* **270**, 313–321. doi:10.1098/rspb.2002.2218
- Ji, Y., Huotari, T., Roslin, T., Schmidt, N. M., Wang, J., Yu, D. W., and Ovaskainen, O. (2020). SPIKEPIPE: a metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes. *Molecular Ecology Resources* **20**, 256–267. doi:10.1111/1755-0998.13057
- Kolter, A., Gemeinholzer, B., and Boatwright, J. S. (2021). Plant DNA barcoding necessitates marker-specific efforts to establish more comprehensive reference databases. *Genome* **64**, 265–298. doi:10.1139/gen-2019-0198
- Kress, W. J., Erickson, D. L., Jones, F. A., Swenson, N. G., Perez, R., Sanjur, O., and Bermingham, E. (2009). Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 18621–18626. doi:10.1073/pnas.0909820106
- Lang, D., Tang, M., Hu, J., and Zhou, X. (2019). Genome-skimming provides accurate quantification for pollen mixtures. *Molecular Ecology Resources* **19**, 1433–1446. doi:10.1111/1755-0998.13061
- Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V., Forest, F., Gilbert, M. T. P., Goldstein, M. M., Grigoriev, I. V., Hackett, K. J., Haussler, D., Jarvis, E. D., Johnson, W. E., Patrinos, A., Richards, S., Castilla-Rubio, J. C., van Sluys, M.-A., Soltis, P. S., Xu, X., Yang, H., and Zhang, G. (2018). Earth BioGenome Project: sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 4325–4333. doi:10.1073/pnas.1720115115
- Li, D.-Z., Gao, L.-M., Li, H.-T., Wang, H., Ge, X.-J., Liu, J.-Q., Chen, Z.-D., Zhou, S.-L., Chen, S.-L., Yang, J.-B., Fu, C.-X., Zeng, C.-X., Yan, H.-F., Zhu, Y.-J., Sun, Y.-S., Chen, S.-Y., Zhao, L., Wang, K., Yang, T., and Duan, G.-W. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 19641–19646. doi:10.1073/pnas.1104551108
- Lim, V.-C., Clare, E. L., Littlefair, J. E., Ramli, R., Bhassu, S., and Wilson, J.-J. (2018a). Impact of urbanisation and agriculture on the diet of fruit bats. *Urban Ecosystems* **21**, 61–70. doi:10.1007/s11252-017-0700-3
- Lim, V.-C., Ramli, R., Bhassu, S., and Wilson, J.-J. (2018b). Pollination implications of the diverse diet of tropical nectar-feeding bats roosting in an urban cave. *PeerJ* **6**, e4572. doi:10.7717/peerj.4572
- Lima, R. A. F. d., Oliveira, A. A. d., Colletta, G. D., Flores, T. B., Coelho, R. L. G., Dias, P., Frey, G. P., Iribar, A., Rodrigues, R. R., Souza, V. C., and Chave, J. (2018). Can plant DNA barcoding be implemented in species-rich tropical regions? A perspective from São Paulo State, Brazil. *Genetics and Molecular Biology* **41**, 661–670. doi:10.1590/1678-4685-gmb-2017-0282
- Markus, N., and Hall, L. (2004). Foraging behaviour of the black flying-fox (*Pteropus alecto*) in the urban landscape of Brisbane, Queensland. *Wildlife Research* **31**, 345–355. doi:10.1071/WR01117
- Marshall, A. G. (1983). Bats, flowers and fruit: evolutionary relationships in the Old World. *Biological Journal of the Linnean Society* **20**, 115–135. doi:10.1111/j.1095-8312.1983.tb01593.x
- Marshall, A. G. (1985). Old World phytophagous bats (Megachiroptera) and their food plants: a survey. *Zoological Journal of the Linnean Society* **83**, 351–369. doi:10.1111/j.1096-3642.1985.tb01181.x
- Moorhouse-Gann, R. J., Dunn, J. C., de Vere, N., Goder, M., Cole, N., Hipperson, H., and Symondson, W. O. C. (2018). New universal ITS2 primers for high-resolution herbivory analyses using DNA metabarcoding in both tropical and temperate zones. *Scientific Reports* **8**, 8542. doi:10.1038/s41598-018-26648-2
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H. (2019). Community Ecology Package Version 2.6. Available at: <https://cran.r-project.org/web/packages/vegan/vegan.pdf>
- Palmer, C., Price, O., and Bach, C. (2000). Foraging ecology of the black flying fox (*Pteropus alecto*) in the seasonal tropics of the Northern Territory, Australia. *Wildlife Research* **27**, 169. doi:10.1071/WR97126
- Palmieri, L., Bozza, E., and Giongo, L. (2009). Soft fruit traceability in food matrices using real-time PCR. *Nutrients* **1**, 316–328. doi:10.3390/n1020316
- Parmentier, I., Dumini, J., Kuzmina, M., Philippe, M., Thomas, D. W., Kenfack, D., Chuyong, G. B., Cruaud, C., and Hardy, O. J. (2013). How effective are DNA barcodes in the identification of African rainforest trees? *PLoS One* **8**, e54921. doi:10.1371/journal.pone.0054921
- Parry-Jones, K. A., and Augee, M. L. (2001). Factors affecting the occupation of a colony site in Sydney, New South Wales by the grey-headed flying-fox *Pteropus poliocephalus* (Pteropodidae). *Austral Ecology* **26**, 47–55.
- Parsons, J. G., Cairns, A., Johnson, C. N., Robson, S. K. A., Shilton, L. A., and Westcott, D. A. (2006). Dietary variation in spectacled flying foxes (*Pteropus conspicillatus*) of the Australian Wet Tropics. *Australian Journal of Zoology* **54**, 417–428. doi:10.1071/ZO06092
- R Core Team (2016). 'R: A Language and Environment for Statistical Computing.' (R Foundation for Statistical Computing: Vienna, Austria.)
- Schuster, T. M., Setaro, S. D., Tibbits, J. F. G., Batty, E. L., Fowler, R. M., McLay, T. G. B., Wilcox, S., Ades, P. K., and Bayly, M. J. (2018). Chloroplast variation is incongruent with classification of the Australian bloodwood eucalypts (genus *Corymbia*, family Myrtaceae). *PLoS One* **13**, e0195034. doi:10.1371/journal.pone.0195034
- Shapcott, A., Forster, P. I., Guymer, G. P., McDonald, W. J. F., Faith, D. P., Erickson, D., and Kress, W. J. (2015). Mapping biodiversity and setting conservation priorities for SE Queensland's rainforests using DNA barcoding. *PLoS One* **10**, e0122164. doi:10.1371/journal.pone.0122164
- Sickel, W., Ankenbrand, M. J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., Steffan-Dewenter, I., and Keller, A. (2015). Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology* **15**, 20. doi:10.1186/s12898-015-0051-y
- Tait, J., Perotto-Baldivieso, H. L., McKeown, A., and Westcott, D. A. (2014). Are flying-foxes coming to town? Urbanisation of the spectacled flying-fox (*Pteropus conspicillatus*) in Australia. *PLoS One* **9**, e109810. doi:10.1371/journal.pone.0109810
- Tidemann, C. R., and Nelson, J. E. (2004). Long-distance movements of the grey-headed flying fox (*Pteropus poliocephalus*). *Journal of Zoology* **263**, 141–146. doi:10.1017/S0952836904004960
- Westcott, D. A., Dennis, A. J., McKeown, A., Bradford, M., and Margules, C. R. (2001). 'The Spectacled Flying Fox, *Pteropus conspicillatus*, in the Context of the World Heritage Values of the Wet Tropics World Heritage Area.' (CSIRO Sustainable Ecosystems and the Rainforest Cooperative Research Centre: Atherton, Australia.)
- White, T. J., Bruns, T., Lee, S., and Taylor, J. W. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In 'PCR Protocols: a Guide to Methods and Applications'. (Eds M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White.) pp. 315–322. (Academic Press: New York.)

Handling Editor: Paul Cooper