DOI: 10.1111/1755-0998.13148

Testing multiple substrates for terrestrial biodiversity monitoring using environmental DNA (eDNA) metabarcoding

van der Heyde, M.[1,2], Bunce, M.[2], Wardell-Johnson, G.[1],Fernandes, K.[2], White, N.E.[2], Nevill, P.[1,2]

[1]ARC Centre for Mine Site Restoration, School of Molecular and Life Sciences, Curtin University, Bentley, GPP Box U1987, Perth, Western Australia, 6845

[2]Trace and Environmental DNA Laboratory, School of Life and Molecular Sciences, Curtin University, GPP Box U1987, Perth, Western Australia, 6845

Corresponding author email:

mieke.vanderheyde@postgrad.curtin.edu.au

keywords: DNA barcoding, metagenomics, biological audit, terrestrial, Biodiversity

Running title: Testing terrestrial metabarcoding substrates

Abstract

Biological surveys based on visual identification of the biota are challenging, expensive, and time consuming, yet crucial for effective biomonitoring. DNA metabarcoding is a rapidly developing technology that can also facilitate biological surveys. This method involves the use of next generation sequencing technology to determine the community composition of a sample. However, it is uncertain as to what biological substrate should be the primary focus of metabarcoding surveys. This study aims to test multiple sample substrates (soil, scat, plant

material and bulk arthropods) to determine what organisms can be detected from each and where they overlap. Samples (n = 200) were collected in the Pilbara (hot desert climate) and Swan Coastal Plain (hot Mediterranean climate) regions of Western Australia. Soil samples yielded little plant or animal DNA, especially in the Pilbara, likely due to conditions not conducive to long-term preservation. In contrast, scat samples contained the highest overall diversity with 131 plant, vertebrate, and invertebrate families detected. Invertebrate and plant sequences were detected in the plant (86 families), pitfall (127 families), and vane trap (126 families) samples. In total 278 families were recovered from the survey, 217 in the Swan Coastal Plain and 156 in the Pilbara. Aside from soil, 22-43% of the families detected were unique to the particular substrate and community composition varied significantly between substrates. These results demonstrate the importance of selecting appropriate metabarcoding substrates when undertaking terrestrial surveys. If the aim is to broadly capture all biota then multiple substrates will be required.

Introduction

There is a growing need for effective biomonitoring with increasing pressure on ecological systems from human population growth, resource use and climate change (Dirzo et al., 2014; Pimm et al., 2014; UNEP, 2011). Biomonitoring is necessary for effective ecosystem management including the early detection of invasive species (Epanchin-Niell, Haight, Berec, Kean, & Liebhold, 2012), measurement of trajectories following ecological restoration (Herrick, Schuman, & Rango, 2006), and the conservation of threatened or endangered species and ecological communities (Campbell et al., 2002). Traditionally, biomonitoring has relied on visual

surveys and traps with species identification based on morphology. However, this presents challenges in some groups due to (i) phenotypic plasticity (Demes, Graham, & Suskiewicz, 2009; Weigand, Jochum, Pfenninger, Steinke, & Klussmann-Kolb, 2011), (ii) juveniles with ambiguous morphology (Ji et al., 2013; Richard et al., 2010), and (iii) taxa having different levels of detectability according to season and time (Fernandes et al., 2018; Thompson & Newmaster, 2014). There has also been a worldwide decline in taxonomic expertise (Pearson, Hamilton, & Erwin, 2011), which further limits traditional approaches. In addition, it is difficult to rely on morphology to monitor across a broad taxonomic range, as expertise and methods tend to be taxon-specific. With the demand for efficient biomonitoring, new technologies are being developed to expand the monitoring "toolkit" to complement traditional methods.

One such method is environmental DNA (eDNA) metabarcoding, a process of sequencing barcode regions from DNA that has been isolated from environmental samples including sediment, water, seawater, bulk arthropods and air (Bohmann et al., 2014). Several reviews have advocated the use of eDNA metabarcoding for biodiversity monitoring (Bohmann et al., 2014; Creer et al., 2016; Fernandes et al., 2018; Taberlet et al., 2012; Thomsen & Willerslev, 2015; Williams et al., 2014) as it has the potential to increase the range of biodiversity detected and to include a broader array of forms (i.e. immature specimens, cryptids, and phoretic individuals). The data are also readily auditable by third parties (Ji et al., 2013) and the cost may be calculated based on number of samples, rather than number of specimens, making it more cost-effective, especially in highly diverse systems (Ji et al., 2013).

The reduction in the cost of high-throughput sequencing has led to a rapid increase in the number of eDNA studies, as well as commercial interest (Supporting Information Koziol et al.

66 2018). Soil microbial researchers have been using eDNA metabarcoding for over two decades
67 (Anderson & Cairney, 2004) and there is now growing evidence that barcoding may be useful to
68 monitor plant communities (de Mattia et al., 2012; Fahner, Shokralla, Baird, & Hajibabaei, 2016;
69 Thompson & Newmaster, 2014), vertebrates (Andersen et al., 2012; Calvignac-spencer, Merkel,
70 & Kutzner, 2013; Fernandes et al., 2019) and invertebrates (Ji et al., 2013; Yang et al., 2014).
71 Researchers have successfully sequenced: top soil (Andersen et al., 2012; Fahner et al., 2016),
72 scat (De Barba et al., 2014), ancient middens (Murray et al., 2012), air (Kraaijeveld et al., 2015),
73 bulk arthropods (Ji et al., 2013; Yu et al., 2012), leaf material (Thompson & Newmaster, 2014),
74 flowers (Thomsen & Sigsgaard, 2019) and more. However, almost all these studies have
75 employed only one eDNA substrate (Koziol et al., 2018). Yang et al. (2014) is one of the few that
76 sampled multiple terrestrial substrates (soil, leaf litter, and insect traps) but they targeted the
77 soil fauna for extraction (separated from the soil) rather than extracting DNA from the soil
78 directly. Yang et al. (2014) also used different PCR assays for their bulk arthropod and soil/leaf
79 litter samples, affecting their comparability. Thus far there has been no study that compares
80 the same barcode across multiple terrestrial substrates. Most samples are able to detect
81 multiple taxonomic groups. However, without a systematic, comparative substrate trial using
82 multiple barcodes it is not possible to determine overlap.
83 Generally, eDNA studies have occurred in temperate regions or colder (Andersen et al.,
84 2012; Fahner et al., 2016), where DNA preservation is more optimal for metabarcoding. Few if
85 any non-microbial barcoding studies have been performed on soils from hotter climates such as
86 those found in parts of Australia. This is true of most biodiversity research; hotter climates are
87 critically understudied (Titley, Snaddon, & Turner, 2017), despite having the highest extinction

rates (Wiens, 2016). As the climate continues to warm and dry (Huang, Yu, Guan, Wang, & Guo, 2016), developing tools to monitor these regions globally has become increasingly important.

In this study, we tested five common terrestrial substrates (soil, bulk scat, bulk plant material and bulk arthropods from pitfall traps and vane traps) with four eDNA barcoding assays to detect a wide range of plants, vertebrates and arthropods. Two well characterized study sites were chosen in Western Australia to examine the feasibility of metabarcoding from substrates collected from hot desert and Mediterranean climates. We aim to improve decision making for terrestrial eDNA surveys by:

1) Examining the diversity within and overlap between commonly sampled substrates. Ideal substrates for monitoring should detect both the greatest richness per sample and greatest overall diversity. The degree of overlap in diversity detected will indicate if multiple substrates are necessary for broad biological surveys.

2) Examining differences between sites as some substrates may perform poorly in certain climates.

3) Comparing DNA sequence and traditional biodiversity survey methods. In particular we compare a DNA sequence based approach with vegetation surveys based on plant morphologyto understand the extent to which they complement one another.

Materials & Methods

Study Sites

Samples (n = 200) were collected from two study sites 1000 km apart in Western Australia; one in the Pilbara, and the other in the south-western Swan Coastal Plain (SCP) (Figure 1). The Pilbara site (-23.19°, 119.24°) is a valley bottom of red clay and strewn pebbles (Burbidge, Johnstone, & Pearson, 2010), has an arid climate with hot summers, mild winters, more than 10 hours sunshine a day, and low but variable rainfall (Sudmeyer, 2016). It is classified as hot desert (Bwh, Beck et al. 2018) and the dominant vegetation consists of *Acacia* shrublands with hummock grasses (*Triodia*) (Burbidge et al., 2010). The Swan Coastal Plain (SCP) site (-31.76°, 115.95°) is in a highly diverse Banksia Woodland on sandy soils. The region has a hot Mediterranean (Csa, Beck et al. 2018) climate with hot, dry summers and cool, wet winters. Study sites with different climates and soil types were chosen because they may affect DNA preservation. Both sites have a broad array of taxa, and the SCP is located in one of Australia's two biodiversity hotspots. The Pilbara also hosts a globally significant resources industry (Argent, 2013) where current and proposed projects require regular biomonitoring or collection of baseline data.

Sample collection

At each site, 5 different substrates were collected; soil, scat, plant material, arthropods from pitfall traps, and arthropods from vane traps. Samples were collected in October and November 2017 (early summer) totally 200 (2 sites x 20 samples per substrate x 5 substrates).

Sample points were 50-60m apart in a grid like pattern across the 15 ha sites. At each sample point, 5 soil subsamples were collected from the surface 5 cm using sterilized equipment and gloves that were re-sterilized with bleach between each sample point. The subsamples were collected randomly within a 10 m x 10 m plot and mixed in the field to form

one sample. Scat was sampled by collecting any visible scat approximately 200 m around each sample point and collecting any scat that was visible. Soil and scat samples were kept cool, and frozen within a few hours. Plant material was collected by a non-specialist but with some training in flora surveys. A leaf was collected from each plant species within a quadrat (50 m x 50 m for the Pilbara, 10 m x 10 m for the SCP) and stored in envelopes in silica gel. Size of quadrats was based on standard monitoring plots used by mining companies in the area. Each sample point also had four pitfall traps (12 cm deep, 4 cm diameter) combined to form one sample, and one yellow vane trap. The traps were left out for 7 days to catch arthropods and contained ethylene glycol in the form of concentrated auto coolant as a capture fluid.

Sample Processing and DNA extraction

All samples were extracted using the Qiacube extraction platform (Qiagen, Germany). Soil samples were manually homogenized and DNA was extracted from 300 mg using the Qiagen DNeasy PowerLyser Powersoil kit (Qiagen, Germany). The Qiagen PowerFecal DNA kit (Qiagen, Germany) was used to extract DNA from 250 mg of each of the scat samples. For the plant samples small sections of each leaf were homogenized dry using Precellys 7ML Hard Tissue Homogenizing Ceramic Beads kit for 8 minutes, then again for 2 minutes with 3 mL of AP1 buffer from the Qiagen DNeasy Plant Mini kit (Qiagen, Germany). They were digested overnight and the DNA extracted using the Plant Mini kit. The arthropod samples were rinsed with de-ionized water using 20 micron sieves that were sterilized in bleach and UV between every sample. They were then homogenized using a hand-held blender (OMNI Tip homogenizer, Kennesaw, GA, USA) and the DNA extracted with a Qiagen QIAmp DNA Mini Kit modified with a

starting volume of 400 µL of digest fluid and a 100 µL elution. DNA extraction controls (blanks) were carried out for every 20 samples using the extraction reagents only.

Assessment of DNA extracts

Quantitative polymerase chain reaction (qPCR) was used to assess the quality and quantity of DNA in the extract, as well as determine the optimal level of DNA input for metabarcoding (Murray, Coghlan, & Bunce, 2015). Four qPCR assays (described below) were run on all samples and all substrates, to determine if there was sufficient amplification to attempt sequencing. Due to the degraded nature of eDNA, all primers used targeted short amplicons (72bp to 157bp) to improve amplification success from samples. ZBJ-ArtF1c/ZBJ-ArtR2c (~157bp, Zeale, Butlin, Barker, Lees, & Jones, 2011) was chosen as a general arthropod primer, with an addition of Ant236/361 (~72bp, Fernandes et al., 2019) to target arthropod orders such as Hymenoptera, which ZBJ-ArtF1c/ZBJ-ArtR2c has shown some bias against (Clarke, Soubrier, Weyrich, & Cooper, 2014; Fernandes et al., 2019). Primer bias may differentially affect sites with different community composition, so the combination of the two invertebrate primers were chosen to control for this bias. Both target sections of cytochrome c oxidase subunit 1 (COI) have extensive reference databases available online to improve taxonomic assignment compared to gene regions with smaller databases. Two plant primer sets were used that target the chloroplast genome: 1) trnL-g/h primers (Taberlet et al., 2007), which produces a PCR amplicon of variable length (10-143 bp) from the P6 loop of the *trn*L (UAA) intron; 2) rbcL-h1aF/h2aR primers (96bp, Poinar et al., 1998) was used only on plant material samples for comparison between the plant sequences and the traditional plant survey.

172 The vertebrate primers targeted the 12S gene 12SV5-F/R (98bp, Riaz et al., 2011). The PCR
173 mix for quantitation contained: 2.5 mM $MgCl_2$ (Applied Biosystems, USA), 1× PCR Gold buffer
174 (Applied Biosystems), 0.25 mM dNTPs (Astral Scientific, Australia), 0.4 mg/ml bovine serum
175 albumin (Fisher Biotec, Australia), 0.4 µmol/L forward and reverse primer, 1 U AmpliTaq Gold
176 DNA polymerase (Applied Biosystems) and 0.6 µl of a 1:10,000 solution of SYBR Green dye (Life
177 Technologies, USA). They were run on a StepOne Plus (Applied BioSystems) real-time qPCR
178 instrument with the following conditions: 5 min at 95°C, 40 cycles of 95°C for 30s, 30s at the
179 annealing temperature (52°C for trnL and rbcL, 53°C for ZBJ-Art, 50°C for Ant236, 60°C for
180 12SV5) and 45s at 72°C, ending with 10 min elongation at 72°C. Contamination was minimised
181 by preparing the PCR mixes in a dedicated clean room and then adding sample in a separate
182 laboratory in specialized UV cabinets.
183 *DNA Amplification and Sequencing*
184 Samples that yielded sufficient amplifiable DNA, as determined by the qPCR screening,
185 were assigned a unique combination of fusion tag primers that contained a unique multiplex
186 identifier (MID) tag between 6-8 bp in length, the gene-specific primer (described above) and
187 Illumina's sequencing adaptors (i.e. P5 and P7). These MID-tag (fusion) primers were then used
188 in qPCR with the same reagents and cycling conditions described above. A single-step fusion
189 protocol was employed with no reuse of index combinations. The MID-tag amplicons were
190 generated in duplicate and then pooled together. Pooled amplicons were cleaned using the
191 QIAquick PCR Purification Kit (Qiagen, Germany) and quantified using the QIAxcel Advanced
192 System (Qiagen, Germany). Pools were combined in approximate equimolar ratios based on
193 this quantitation to create a DNA library for sequencing. Amplicons in this library were size

selected using a Pippin Prep (Sage Science), cleaned using the QIAquick PCR Purification Kit (Qiagen, Germany) and eluted into 50µl. The final DNA library was quantified using Qubit Fluorometric Quantitation (Thermo Fisher Scientific) and sequenced as per Illumina sequencing protocols for single-end sequencing.

Sequence Analysis

Raw sequence reads were demultiplexed (i.e. assigned back to sample using MID-tag primer combos) using 'obitools' (Boyer et al., 2016), then sequences were quality filtered, the errors and chimeras removed using DADA2 (Callahan et al., 2016) on R 3.5.1 (R Core Team, 2018) (script available on https://doi.org/10.5061/dryad.38100f6). DADA2 denoises sequences using error rates estimated from the sequencing run, producing amplicon sequence variants (ASVs) that are not clustered like traditional operational taxonomic units (OTUs), but are analyzed in much the same way. The resulting ASV tables of were then analyzed in R 3.5.1 (R Core Team, 2018). We ignored sequence counts less than 5, and removed samples with less than 200 seq/sample. Any ASVs that were present in the extraction (i.e. laboratory) controls were then removed from the dataset. ASV sequences were matched to a reference database using the Basic Local Alignment Search Tool (BLASTn) on a high performance cluster computer (Pawsey Supercomputing Centre; Perth, WA, Australia) against the online reference database Genbank (https://www.ncbi.nlm.nih.gov/genbank/) for taxonomic assignment. The ZBJ-Art and Ant236 assays were searched against both Genbank and Arthropod COI sequences extracted from the Barcode of Life Database (BOLD: https://www.barcodeoflife.org), because there are arthropod sequences on this database not present in Genbank. BLASTn results returned the top 10 hits with a minimum query coverage of 80% and minimum percent identity or 80%.

Taxonomic identification was assigned to the lowest common ancestor with MEGAN (Huson, Auch, Qi, & Schuster, 2007) with minimum support of 140 (ZBJ-Art), 60 (Ant236), 90 (12SV5), 50, (trnL), and 90 (rbcL).

Vegetation Surveys

Vegetation surveys were conducted by an expert botanist, for each sample point on the SCP in November 2017. Due to time constraints while working remotely, the Pilbara vegetation surveys include 8 sample points conducted in September 2018. The vegetation survey quadrat areas matched those of the sample collection of plant material (10 m x 10 m in SCP and 50 m x 50 m in Pilbara), with the intent of providing a morphological comparison for the plant samples. For the comparison between sequencing data and the morphological surveys, only sample points with both vegetation surveys and sequencing data were included. Identifications of all sampled specimens were confirmed by botanists at the Western Australian Herbarium but no voucher specimens were lodged.

Statistical analysis

Statistical analysis was performed on R 3.5.1 (R Core Team, 2018). For the four assays that were tested on all substrates (ZBJ-Art, Ant236, trnL, 12SV5), we calculated the ASV richness for each substrate (soil, scat, plant material, arthropods from pitfall traps, and arthropods from vane traps) at both sites (SCP and Pilbara) and tested the differences between substrates and site using an two-way analysis of variance (ANOVA) where the data met the assumptions (12SV5). For most assays (trnL, ZBJ-Art, and Ant236), the groups did not have equal variance and so we used permutational analysis of variance (PERMANOVA) from the R package 'vegan' (Oksanen et al., 2019) with 999 permutations. We tested the differences in community

composition between sites and substrates by first combining the four assays into a presence absence matrix that included taxonomic families and the samples in which they were detected. Then the Bray-Curtis similarity was calculated between samples and the differences in community composition were tested using the PERMANOVA with 999 permutations. This was visualized using nonmetric multidimensional scaling (NMDS) from the same package (Oksanen et al., 2019). This matrix was also used to determine the families associated with each substrate by using the R package 'indicspecies' (De Cáceres & Legendre, 2009). Plant families detected from plant material samples were also compared to the families identified in the plant survey.

Results

A total of 26 589 497 metabarcoding sequences were generated from the four PCR assays (See Table 1 for summary). Small quantities of ASVs were present in extraction controls, highest in the trnL assay (5.7%) and lowest in the ZBJ-Art Assay (0.7%). A fish sequence variant was removed as a likely contaminant, it was found in 0.01% of 12SV5 sequences (See Supplementary information for details). In total, there were 278 taxonomic families detected from 87 orders.

Substrate Diversity and Richness

Substrates varied significantly in diversity and ASV richness. Pitfall traps detected the greatest number of ASVs overall (1792 ASVs), while vane traps ranked third overall with 1208 ASVs in total. Approximately the same number of families were detected from both traps

across the two sites (127 in pitfall traps and 126 in vane traps), and they had similar proportions of families unique to that substrate (pitfall traps 35.4%; vane traps 33.3%).

Scat samples were found to have fewer ASVs overall (1333 ASVs) than pitfall traps, but had the highest ASV richness per sample for most assays (Figure 2). There are differences between the per sample richness and the accumulated richness; for example, although scat had higher average ZBJ-Art ASV richness (32.1± 4.4se) than vane samples (18.5 ± 2.0se), there were more ZBJ-Art ASVs in the vane substrate than in scat (542 vs 470). The differences between per sample richness and accumulated richness are related to the overlap between samples. Scat required fewer samples to achieve the same proportion of cumulative diversity than pitfall and vane traps, which have more variation between samples. Overall, scat samples detected the most families overall (131 families), had the most families unique to scat samples (56 families, 42.7%), and was the only substrate that showed successful amplification with the vertebrate 12SV5 assay (Figure 2; Table S1).

Plant samples were found to have similar overall ASV richness (1326 ASVs) to scat (1333 ASVs) and vane traps (1208 ASVs), but fewer families (86 families). The degenerate nature of the Ant236 primers resulted in plant DNA being sequenced, which was confirmed with the trnL assay. Overall, 22.1% of the families detected in the plant samples were unique to this substrate.

Despite numerous optimizations and extraction attempts (modified bead bashing time, pelleting time, alternate extraction method), less than half of the soil samples successfully amplified (Table S1). For trnL and ZBJ-Art, 11/40 soil samples (27.5%) and 18/40 (47%) soil samples amplified respectively. Soil samples were the lowest in the number of total ASVs

detected (224 ASV), and the lowest taxonomic diversity identified (18 families; Figure 3 & 4). Most (63%) of families identified from soil samples were detected in at least two other substrates and only 11.1% of families were unique to soil.

Only 3% of families were detected in all substrates, 10% if soil is excluded from the analysis. All 8 of the families found in all substrates including soil were plant families (Table 2). The results of our indicator analyses showed that 23 out of 157 families in the Pilbara, and 86 out of 217 families in the SCP were significantly associated with one or more substrates ($p < 0.05$, see Supplementary Table S4 for details). Over half (59%) of the families were detected in only one substrate (130 in SCP, 109 Pilbara), and of the families significantly associated with substrates ($p < 0.05$ indicator analysis, Table S4), the majority (63% SCP, 74% Pilbara) were associated with a single substrate. For example, vertebrate families such as Macropodidae (wallabies and kangaroos), and Dromaiidae (emus) were associated with scat samples, while Formicidae (ants) and Termitidae (termites) were associated with pitfall traps. Other families were associated with multiple substrates; i.e. Noctuidae (Owlet moths) in plant and scat, Poaceae (grasses) in plant and scat and soil, and Acrididae (Short-horned grasshopper) in pitfall and vane trap samples. In the SCP, five plant families were significantly associated with all substrates except soil (Myrtaceae, Fabaceae, Asteraceae, Proteaceae, and Dilleniaceae).

Site differences

ASV richness was greater in the SCP than in the Pilbara (Figure 2). There was an interaction between site and substrate for both the trnL and Ant236 assay, and site and substrate were significant terms for all assays ($p < 0.001$, see Supplementary Material for details). The vertebrate assay was the only one where the Pilbara had significantly ($p < 0.001$) greater ASV

richness, and 92% of the ASVs were assigned to the Macropodidae family. The remaining assays showed SCP having higher ASV richness than the Pilbara (Figure 2). Taxonomic diversity was also higher in the SCP (217 families) compared to the Pilbara (156 families)(Figure 4). Additionally, more samples from the SCP were able to be amplified and sequenced; in the trnL assay we were only able to sequence 2 Pilbara soil samples and 9 SCP soil samples. Similarly, in the ZBJ-Art assay, we were able to sequence 16 SCP soil samples and only 2 Pilbara soil samples.

Site influenced the families found in each substrate ($F_{4,168}$ = 10.694, R2 = 0.112, $p < 0.001$, Figure 4A), and site ($F_{1,168}$ = 38.4, R2 = 0.101, $p < 0.001$) and substrate ($F_{4,168}$ = 32.5, R2 = 0.343, $p<0.001$) were also highly significant (See Table S3 for details)

Comparison between plant surveys

The morphology based survey identified 51 families in total (Table S6). Several plant families found in the sequencing (e.g. Mazaceae and Hyacinthaceae) had been reassigned or renamed, and this was accounted for in our comparison with the morphological survey. The Pilbara has the lowest proportion of families found in both the sequencing and morphological survey (48.1%, Figure 5) compared to the SCP (61.7% Fig 5). In total, of the 15 plant families identified in the morphological survey and not found in the sequencing data, 8 were identified in only one sample. Families detected in the SCP sequencing data and not in the morphological survey include Pinaceae, a family of plants not present in the study site but in high abundance in an adjacent pine plantation. Four of the families found in sequencing only were detected at a single sample point. Seven families had less than 700 reads, and the average number of reads per family was 28 580 for the trnL assay and 45 867 for the rbcL assay. In the SCP, almost all

sequencing reads (99%) and morphological plant species (93%) were from families identified in both the sequencing and morphological survey; in the Pilbara, a smaller proportion of reads (81%) and sequences (83%) were from families shared between methods (Figure 5).

Discussion

Use of DNA barcoding and eDNA substrates to monitor biodiversity is on the rise (Koziol et al., 2018). In terrestrial systems researchers are expanding both the range (e.g. Yang et al., 2014) and purpose (Thomsen & Sigsgaard, 2019) of substrates collected. In this study, we tested multiple terrestrial substrates with four barcoding assays and found that terrestrial substrates can detect a broad range of taxonomic groups (Figure 2). Invertebrate and plant DNA was found in all substrates, although scat was the only substrate to consistently yield vertebrate sequences. Some cosmopolitan taxa were shared with multiple substrates. However, many taxa are specific to a particular substrate, and no one substrate was able to detect all taxa.

Substrate Diversity and Richness

Each substrate identified a different biological community in the five substrates and four assays used. Even the most diverse substrate (scat) only comprised half of the total diversity detected (47%). The combination of pitfall trap and scat samples increased detection to 76% of total and with the addition of vane traps, 92% of families detected from all five substrates. For terrestrial biodiversity monitoring, the more substrates chosen, the greater the range of biodiversity that can be detected. This result mirrors a substrate comparison in marine environments and further demonstrates that the manner in which environments are sampled

345 strongly influences both the sensitivity of detection and the assemblages recovered (Koziol et
346 al., 2018).
347 Generally, the literature agrees with what we were able to detect in our substrates. Like Ji
348 et al. (2013) and Yu et al. (2012) we found invertebrate sequences in the pitfall and vane trap
349 samples (Figure 3). Families of flying insects were associated with vane traps (Phoridae,
350 Cicadellidae, and Crambidae) but pitfall traps caught both ground dwellers (Formicidae,
351 Lycosidae) and flying insects (Empididae). While Calvignac-Spencer et al. (2013) were able to
352 retrieve vertebrate DNA from their carrion flies, we were not able to successfully sequence our
353 bulk arthropod samples with the vertebrate assay. Perhaps not enough carnivorous
354 invertebrates were trapped, or the ones trapped had not recently fed. Plant sequences were
355 detected in bulk arthropod samples and may have come from pollinating insects (Pornon et al.,
356 2016), ingested plants (Jurado-Rivera, Vogler, Reid, Petitpierre, & Gómez-Zurita, 2009), or plant
357 material that fell into the traps. The traps differed in the biota that they detected, as we
358 expected based on entomology studies that show the importance of trap type to the
359 biodiversity captured (Prasifka, Lopez, Hellmich, Lewis, & Dively, 2007; Santos, Cabanas, &
360 Pereira, 2007). The traps are complementary, and when used together can detect the majority
361 of families (Figure 4B).
362 In the plant samples, we found both plant and invertebrate sequences. In plant samples,
363 the source of plant sequences is self-evident, while invertebrate DNA may come from larvae, or
364 cells and fecal matter left on plants (Thomsen & Sigsgaard, 2019), spider webs (Blake,
365 McKeown, Bushell, & Shaw, 2016), and empty leaf mines (Derocles, Evans, Nichols, Evans, &
366 Lunt, 2015). As evidence, the arthropod families strongly associated with plant samples are all

367 those that feed and reproduce on living plants (i.e. Gelechiidae, Aphididae, Cecidomyiidae),
368 giving them ample opportunity to leave DNA on the plant. Some plant taxa were commonly
369 found in multiple substrates. These include orders such as Poales, Myrtales, and Proteales,
370 which were found in all five substrates (including soil samples - Figure 3). The cosmopolitan
371 nature of these taxa suggest they may have airborne dispersal, and Kraaijeveld et al. (2015)
372 were able to detect several taxa in these orders from airborne pollen traps.
373 The most taxonomically diverse substrate was scat, which picked up vertebrates,
374 invertebrates and plants (Figure 3). Typically, scat is used to study the diet of a particular
375 organism and researchers choose their barcoding assays accordingly; insectivores' scat are
376 barcoded with invertebrate primers (Zeale et al., 2011), herbivores' scat is barcoded using plant
377 primers (Valentini et al., 2009), and carnivores' scat is barcoded using vertebrate primers
378 (Arteaga Claramunt et al., 2018). Our scat samples were dominated by Macropididae
379 (Kangaroos and Wallabies), likely as a result of sampling bias. Rather than target a specific
380 organism for scat collection, we collected scat along a transect, and Macropodidae scat is both
381 numerous and easy to see, resulting in an overabundance of this family. This sampling bias is
382 exacerbated by site differences, for example, Passeriformes scat is relatively easy to see in the
383 red clay of the Pilbara, but almost impossible to find in the sand of the SCP. In addition,
384 although many samples were from only herbivorous animals (Macropodidae and Dromaiidae)
385 they still picked up invertebrate sequences. This could be from contact with invertebrates on
386 the ground such as beetles (Coleoptera Figure 3). Furthermore, the presence of families of
387 moth within scat samples (Noctuidae, Oecophoridae) can be explained through biological
388 material that could have been eaten (larvae, fecal matter, and other sources) along with plant

material and survived to be sequenced from the scat samples. The richness per sample (Figure 2) and overall diversity suggest that scat samples are appropriate for broad biodiversity surveys. Nevertheless, caution should be applied as there is no guarantee that the diversity detected was not transported from outside the study area. Diet analysis of targeted organisms may be more informative than scats along a transect; for example, analyzing restoration success through diet changes in frugivorous bats (Galimberti et al., 2016).

Site differences

The SCP generated higher biodiversity and greater ASV richness than the Pilbara, as expected based on known biodiversity in these regions (Rix et al., 2015). Patterns of diversity were consistent across site; substrates generated similar levels of unique and overlapping diversity. The greatest difference between sites, which performed poorly overall, but particularly so in the Pilbara where only two samples were successfully amplified in two assays. While other studies using soil samples as a substrate were able to reflect the above ground diversity of plants (Fahner et al., 2016; Yoccoz et al., 2012), vertebrates (Andersen et al., 2012), and other metazoan (Drummond et al., 2015), less than half of our soil samples successfully amplified. This may be a result of our sampling method and the environment (i.e. hot desert, hot mediterranean) from which the samples were sourced. This is the first study investigating non-microbial soil eDNA in hot Mediterranean and desert climates. DNA at our study sites would be relatively more degraded by the heat (Sirois & Buckley, 2019) and high UV radiation (Barnes et al., 2014) than Denmark (Andersen et al., 2012) or New Zealand (Drummond et al., 2015). This might explain why the Pilbara, which is hotter than the SCP and less shaded, had fewer soil samples that amplified successfully. Despite these results, we feel further testing is

411 needed before dismissing soil as a worthwhile substrate for non-microbial eDNA studies in
412 similar environments. We sampled in a relatively hot period and only the surface 0-5cm. It is
413 possible that better results may be achieved by sampling deeper in the soil profile, in cooler
414 weather, and perhaps with more subsamples.

415 *Comparison between plant surveys*

416 The majority (83-93%) of morphological plant species identified were from families also
417 detected in the sequencing data. In total, 7 families in the Pilbara and 8 families in the SCP
418 were identified in the morphological survey and not by metabarcoding (Figure 5). One possible
419 explanation is that assigning taxonomy to a DNA sequence (i.e. barcode) is dependent on the
420 quality of the reference database. Thompson and Newmaster (2014) found that metabarcoding
421 (rbcL and ITS2 gene regions) was more accurate than their morphology based plant survey.
422 However, they had access to a comprehensive and fully referenced barcoding database for the
423 Boreal forest in Canada. The Pilbara and SCP have a much more diverse flora than the Boreal
424 forests of Canada (Rix et al., 2015), and much of the biodiversity is just beginning to enter
425 barcoding libraries (Dormontt et al., 2018; Nevill et al. 2020). Fortunately, reference databases
426 are continuously growing at an impressive rate often through large-scale initiatives (Costa &
427 Carvalho, 2017; Hendrich et al., 2015).

428 Many (7/12) of the families unique to the morphological surveys were found in very low
429 abundance at only one sample point, and may have been missed in the collection of plant
430 material for metabarcoding. However, another consideration is that DNA is more easily
431 extracted from some plant species than others based on the amount of secondary metabolites
432 and variation in leaf structure (Friar, 2005; Khanuja, Shansany, & Kumar, 1999). Some

recalcitrant plant species with low DNA concentrations or poor quality extracts may have been drowned out by the more easily processed plants in the mix, resulting in plants not showing up in sequencing data, despite having been collected.

Limitations

Our results highlight some important limitations in using metabarcoding for terrestrial biodiversity monitoring. We discussed above the importance of reference databases and extraction bias, but our results also bring into focus the difficulty in determining when DNA might have originated from outside the study area (Figure 5). Several plant orders found in the scat and invertebrate samples (Rosales, Vitales, Brassicales) are not present in the plant samples or the plant surveys, indicating they were either missed in surveys or likely originated from nearby suburban gardens. Depending on the study objectives, proponents may want to limit assays and substrates to those that target taxa that are more likely to occur at the sample point (e.g. plant material from plants in a quadrat, invertebrate sequences from bulk invertebrate samples). However, the presence of DNA, even if not proven to be from an organism inhabiting a particular area, does indicate connection between the sample area and those organisms.

Conclusion

We tested five terrestrial eDNA substrates with four metabarcoding assays for a total of 523 sequenced samples, to examine the impact of substrate on eDNA based biodiversity assessment. Our results are consistent with many other eDNA studies by demonstrating the ability of eDNA metabarcoding as a powerful tool for terrestrial biomonitoring, providing a

broad survey of terrestrial environments. While we cannot equivocally state that these results would transfer across all biomes, the consistency in patterns across two sites is highly suggestive of strong spatial fidelity. We showed in our systematic comparison of substrates that the choice of substrate heavily dictates what taxa will be detected and that each additional substrate will increase the number of taxa detected. Therefore, substrates should be selected with care based on the purpose of monitoring and available funding (see Table 3 for recommendations based on target taxa and survey limitations). For example, soil and plant samples identified relatively few unique families. If these substrates were excluded from this study, 92% of the overall diversity would have still been detected. In contrast, scat and pitfall samples had higher per sample richness, making them suitable for surveys of total biodiversity where funding is limited. We cannot make any specific recommendations on which assays to use, as this is beyond the scope of this paper. Instead we can make broad recommendations for appropriate assay targets based on survey target. For example, where there are concerns about the DNA source, proponents may need to limit surveys to substrates and assay combinations more likely to be locally present (invertebrate assays on bulk arthropods, vertebrate assays on scat etc.)

We also show the limitations of metabarcoding where reference databases are depauperate, and that caution should be exercised with regard to the source of DNA in each substrate, which might originate from organisms outside the immediate study area. Our study highlights the utility of eDNA as biomonitoring tool but also cautions that, like other survey

475 methods, its utility, sensitivity and efficacy will be influenced by how studies are designed and
476 executed.

477 Acknowledgements

478 This work was supported by the Australian Research Council Industrial Transformation
479 Training Centre for Mine Site Restoration (ICI150100041) and the Pawsey Supercomputing
480 Centre. We would like to thank Katrina West and Khiraj Bhalsing for help with sample
481 collection. We acknowledge the support of the members of the Trace and Environmental DNA
482 (TrEnD) Laboratory with metabarcoding workflows and bioinformatics.

483 Data Accessibility

484 Sequencing data and DADA2 script is available at the Dryad Digital Repository:
485 https://doi.org/10.5061/dryad.38100f6
486

487 Author Contributions

488 MvH conducted the study and wrote the manuscript. PN, MB, NW, and GW-J were involved
489 in the experimental design; samples were collected by MvH, PN, KF, GW-J, and processed by
490 MvH and KF; molecular and bioinformatics work was performed by MvH; all data was analyzed
491 and processed by MvH; statistical analysis was done by MvH; the manuscript was edited by all
492 authors.

ORCID

Mieke van der Heyde: https://orcid.org/0000-0002-1658-9927

Michael Bunce: https://orcid.org/0000-0002-0302-4206

Grant Wardell-Johnson: https://orcid.org/0000-0002-6751-9224

Kristen Fernandes: https://orcid.org/0000-0003-3388-5354

Nicole White: https://orcid.org/0000-0002-0068-6693

Paul Nevill: https://orcid.org/0000-0001-8238-0534

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

References

Andersen, K., Bird, K. L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjær, K. H., … Willerslev, E. (2012). Meta-barcoding of "dirt" DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, *21*(8), 1966–1979. doi: 10.1111/j.1365-294X.2011.05261.x

Anderson, I. C., & Cairney, J. W. G. (2004). Diversity and ecology of soil fungal communities: Increased understanding through the application of molecular techniques. *Environmental Microbiology*, *6*(8), 769–779. doi: 10.1111/j.1462-2920.2004.00675.x

Argent, N. (2013). Reinterpreting Core and Periphery in Australia's Mineral and Energy Resources Boom: An Innisian perspective on the Pilbara. *Australian Geographer*, *44*(3), 323–340. doi: 10.1080/00049182.2013.817033

Arteaga Claramunt, A. M., White, N. E., Bunce, M., O'Connell, M., Bullen, R. D., & Mawson, P. R. (2018). Determination of the diet of the ghost bat (Macroderma gigas) in the Pilbara region of Western Australia from dried prey remains and DNA metabarcoding. *Australian Journal of Zoology*, *66*(3), 195. doi: 10.1071/ZO18040

Barnes, M. A., Turner, C. R., Jerde, C. L., Renshaw, M. A., Chadderton, W. L., & Lodge, D. M.

(2014). Environmental conditions influence eDNA persistence in aquatic systems. *Environmental Science and Technology*, *48*(3), 1819–1827. doi: 10.1021/es404734p

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*, 1–12. doi: 10.1038/sdata.2018.214

Blake, M., McKeown, N. J., Bushell, M. L. T., & Shaw, P. W. (2016). DNA extraction from spider webs. *Conservation Genetics Resources*, *8*(3), 219–221. doi: 10.1007/s12686-016-0537-8

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., … de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, *29*(6), 358–367. doi: 10.1016/j.tree.2014.04.003

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*(1). doi: 10.1111/1755-0998.12428

Burbidge, A. H., Johnstone, R. E., & Pearson, D. J. (2010). A Biodiversity Survey of the Pilbara Region of Western Australia , 2002 – 2007. *Records of the Western Australiam Museum*, *Suppliment*(78), 247–279.

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

Calvignac-spencer, A., Merkel, K., & Kutzner, N. (2013). *Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity*. 915–924. doi: 10.1111/mec.12183

Campbell, S. P., Clark, J. A., Crampton, L. H., Guerry, A. D., Hatch, L. T., Hosseini, P. R., … O'Connor, R. (2002). An Assessment of Monitoring Efforts in Endangered Species Recovery Plans. *Ecological Applcations*, *12*(3), 674–681.

Clarke, L. J., Soubrier, J., Weyrich, L. S., & Cooper, A. (2014). Environmental metabarcodes for insects: In silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, *14*(6). doi: 10.1111/1755-0998.12265

Costa, F. O., & Carvalho, G. R. (2017). The Barcode of Life Initiative: synopsis and prospective

societal impacts of DNA barcoding of Fish. *Genomics, Society and Policy*, *3*(2), 29–40. doi: 10.1186/1746-5354-3-2-29

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., … Bik, H. M. (2016). The ecologist ' s field guide to sequence-based identification of biodiversity. *Methods in Ecology Adn Evolution*, *7*, 1008–1018. doi: 10.1111/2041-210X.12574

De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Molecular Ecology Resources*, *14*(2). doi: 10.1111/1755-0998.12188

De Cáceres, M., & Legendre, P. (2009). Associations between species and groups of sites and statistical inference. *Ecology*, *90*(12), 3566–3574. doi: 10.1890/08-1823.1

de Mattia, F. D. E., Gentili, R., Bruni, I., Galimberti, A., Sgorbati, S., Casiraghi, M., & Labra, M. (2012). A multi-marker DNA barcoding approach to save time and resources in vegetation surveys. *Botanical Journal of Linnean Society*, *169*, 518–529.

Demes, K. W., Graham, M. H., & Suskiewicz, T. S. (2009). Phenotypic plasticity reconciles incongruous molecular and morphological taxonomies: The giant kelp, macrocystis (laminariales, phaeophyceae), is a monospecific genus. *Journal of Phycology*, *45*(6), 1266–1269. doi: 10.1111/j.1529-8817.2009.00752.x

Derocles, S. A. P., Evans, D. M., Nichols, P. C., Evans, S. A., & Lunt, D. H. (2015). Determining Plant – Leaf Miner – Parasitoid Interactions: A DNA Barcoding Approach. *PLOS ONE*, *10*(2), e0117872. doi: 10.1371/journal.pone.0117872

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J. B., & Collen, B. (2014). Defaunation in the Anthropocene. *Science*, *345*(6195), 401–406. doi: 10.1126/science.1251817

Dormontt, E. E., van Dijk, K., Bell, K. L., Biffin, E., Breed, M. F., Byrne, M., … 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*(September), 1–12. doi: 10.3389/fevo.2018.00134

Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C. M., … Nelson, N. (2015). Evaluating a multigene environmental DNA approach for biodiversity

assessment. *GigaScience*, *4*(46). doi: 10.1186/s13742-015-0086-1

Epanchin-Niell, R. S., Haight, R. G., Berec, L., Kean, J. M., & Liebhold, A. M. (2012). Optimal surveillance and eradication of invasive species in heterogeneous landscapes. *Ecology Letters*, *15*(8), 803–812. doi: 10.1111/j.1461-0248.2012.01800.x

Fahner, N. A., Shokralla, S., Baird, D. J., & 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*(6), 1–16. doi: 10.1371/journal.pone.0157505

Fernandes, K., van der Heyde, M., Bunce, M., Dixon, K., Harris, R. J., Wardell-Johnson, G., & Nevill, P. G. (2018). DNA metabarcoding—a new approach to fauna monitoring in mine site restoration. *Restoration Ecology*, *26*(6), 1098–1107. doi: 10.1111/rec.12868

Fernandes, K., van der Heyde, M., Coghlan, M., Wardell-Johnson, G., Bunce, M., Harris, R., & Nevill, P. (2019). Invertebrate DNA metabarcoding reveals changes in communities across mine site restoration chronosequences. *Restoration Ecology*, 1–10. doi: 10.1111/rec.12976

Friar, E. A. (2005). Isolation of DNA from Plants with Large Amounts of Secondary Metabolites. *Methods in Enzymology*, *395*, 3–14. doi: 10.1016/S0076-6879(05)95001-5

Galimberti, A., Spinelli, S., Bruno, A., Mezzasalma, V., De Mattia, F., Cortis, P., & Labra, M. (2016). Evaluating the efficacy of restoration plantings through DNA barcoding of frugivorous bird diets. *Conservation Biology*. doi: 10.1111/cobi.12687

Hendrich, L., Morinière, J., Haszprunar, G., Hebert, P. D. N., Hausmann, A., Köhler, F., & Balke, M. (2015). A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. *Molecular Ecology Resources*, *15*(4), 795–818. doi: 10.1111/1755-0998.12354

Herrick, J. E., Schuman, G. E., & Rango, A. (2006). Monitoring ecological processes for restoration projects. *Journal for Nature Conservation*, *14*(3–4), 161–171. doi: 10.1016/j.jnc.2006.05.001

Huang, J., Yu, H., Guan, X., Wang, G., & Guo, R. (2016). Accelerated dryland expansion under climate change. *Nature Climate Change*, *6*(2), 166–171. doi: 10.1038/nclimate2837

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data.

606 *Genome Research*, *17*(3), 377–386. doi: 10.1101/gr.5969107
607 Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., … Yu, D. W. (2013).
608 Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology*
609 *Letters*, *16*(10), 1245–1257. doi: 10.1111/ele.12162
610 Jurado-Rivera, J. A., Vogler, A. P., Reid, C. A. M., Petitpierre, E., & Gómez-Zurita, J. (2009). DNA
611 barcoding insect-host plant associations. *Proceedings of the Royal Society B: Biological*
612 *Sciences*, *276*(1657), 639–648. doi: 10.1098/rspb.2008.1264
613 Khanuja, S. P. S., Shansany, A. K., & Kumar, S. (1999). Rapid isolation of DNA from and fresh
614 samples of plants producing large amounts of secondary metabolites and essential Oils.
615 *Plant Molecular Biology Reporter*, *17*, 1–7. doi: 10.1023/A:1007528101452
616 Koziol, A., Stat, M., Simpson, T., Jarman, S., Harvey, E. S., McDonald, J., … Bunce, M. (2018).
617 Environmental DNA metabarcoding studies are critically affected by substrate selection.
618 *Molecular Ecology Resources*, (September), 1–11. doi: 10.1111/1755-0998.12971
619 Kraaijeveld, K., de Weger, L. A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P. S., &
620 den Dunnen, J. T. (2015). Efficient and sensitive identification and quantification of
621 airborne pollen using next-generation DNA sequencing. *Molecular Ecology Resources*,
622 *15*(1), 8–16. doi: 10.1111/1755-0998.12288
623 Murray, D. C., Coghlan, M. L., & Bunce, M. (2015). From Benchtop to Desktop: Important
624 Considerations when Designing Amplicon Sequencing Workflows. *PLOS ONE*, *10*(4),
625 e0124671. doi: 10.1371/journal.pone.0124671
626 Murray, D. C., Pearson, S. G., Fullagar, R., Chase, B. M., Houston, J., Atchison, J., … Bunce, M.
627 (2012). High-throughput sequencing of ancient plant and mammal DNA preserved in
628 herbivore middens. *Quaternary Science Reviews*, *58*, 135–145. doi:
629 10.1016/J.QUASCIREV.2012.10.021
630 Nevill, P. G., Zhong, X., Tonti-Filippini, J., Byrne, M., Hislop, M., Thiele, K., van Leeuwen, S.,
631 Boykin, L. M., Small, I. (2020). Large scale genome skimming from herbarium material for
632 accurate plant identification and phylogenomics. *Plant Methods, 16*(1), 1-8.
633 doi: 10.1186/s13007-019-0534-5
634 Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., … Maintainer, H.

W. (2019). vegan: Community Ecology Package. R package version 2.5-5. https://CRAN.R-project.org/package=vegan. *Community Ecology Package*, *2*(9), 1–297. Retrieved from https://cran.r-project.org/web/packages/vegan/vegan.pdf

Pearson, D. L., Hamilton, A. L., & Erwin, T. L. (2011). Recovery Plan for the Endangered Taxonomy Profession. *BioScience*, *61*(1), 58–63. doi: 10.1525/bio.2011.61.1.11

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., … Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*(6187). doi: 10.1126/science.1246752

Poinar, H. N., Hofreiter, M., Spaulding, W. G., Martin, P. S., Stankiewicz, B. A., Bland, H., … Pääbo, S. (1998). Molecutar Coproscopy : Dung and Diet of the Extinct Ground Stoth. *Science*, *281*(5375), 402–406.

Pornon, A., Escaravage, N., Burrus, M., Holota, H., Khimoun, A., Mariette, J., … Andalo, C. (2016). Using metabarcoding to reveal and quantify plant-pollinator interactions. *Scientific Reports*, *6*(February), 27282. doi: 10.1038/srep27282

Prasifka, J. R., Lopez, M. D., Hellmich, R. L., Lewis, L. C., & Dively, G. P. (2007). Comparison of pitfall traps and litter bags for sampling ground-dwelling arthropods. *Journal of Applied Entomology*, *131*(2), 115–120. doi: 10.1111/j.1439-0418.2006.01141.x

R Core Team. (2018). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org*.

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). EcoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, *39*(21), 1–11. doi: 10.1093/nar/gkr732

Richard, B., Decaëns, T., Rougerie, R., James, S. W., Porco, D., & Hebert, P. D. N. (2010). Re-integrating earthworm juveniles into soil biodiversity studies: Species identification through DNA barcoding. *Molecular Ecology Resources*, *10*(4), 606–614. doi: 10.1111/j.1755-0998.2009.02822.x

Rix, M. G., Edwards, D. L., Byrne, M., Harvey, M. S., Joseph, L., & Roberts, J. D. (2015). Biogeography and speciation of terrestrial fauna in the south-western Australian biodiversity hotspot. *Biological Reviews*, *90*(3), 762–793. doi: 10.1111/brv.12132

Santos, S. A. P., Cabanas, J. E., & Pereira, J. A. (2007). Abundance and diversity of soil arthropods in olive grove ecosystem (Portugal): Effect of pitfall trap type. *European Journal of Soil Biology*, *43*(2), 77–83. doi: 10.1016/j.ejsobi.2006.10.001

Sirois, S. H., & Buckley, D. H. (2019). Factors governing extracellular DNA degradation dynamics in soil. *Environmental Microbiology Reports*, *11*(2), 173–184. doi: 10.1111/1758-2229.12725

Sudmeyer, R. (2016). Climate in the Pilbara. *Bulletin 4873*, (Department of Agriculture and Food, Western Australia, Perth).

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., … Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, *35*(3), e14–e14. doi: 10.1093/nar/gkl938

Taberlet, P., Coissac, E., Pompanon, F. F., Brochmann, C., Willerslev, E., PIERRE TABERLET, ERIC COISSACFRANC, FRANCOIS POMPANON, C. B. and E. W., … Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. doi: 10.1111/j.1365-294X.2012.05470.x

Thompson, K. A., & Newmaster, S. G. (2014). Molecular taxonomic tools provide more accurate estimates of species richness at less cost than traditional morphology-based taxonomic practices in a vegetation survey. *Biodiversity and Conservation*, *23*(6), 1411–1424. doi: 10.1007/s10531-014-0672-z

Thomsen, P. F., & Sigsgaard, E. E. (2019). Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. *Ecology and Evolution*, *9*(4), 1665–1679. doi: 10.1002/ece3.4809

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18. doi: 10.1016/j.biocon.2014.11.019

Titley, M. A., Snaddon, J. L., & Turner, E. C. (2017). Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. *PLoS ONE*, *12*(12), 1–14. doi: 10.1371/journal.pone.0189577

UNEP. (2011). Report of the Tenth Meeting of the Conference of the Parties to the Convention

on Biological Diversity. *UNEP/CBD/COP/10/27.*, 18–29 October 2010, Nagoya, Japan.

Valentini, A., Miquel, C., Nawaz, M. A., Bellemain, E., Coissac, E., Pompanon, F., … Taberlet, P. (2009). New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: The trnL approach. *Molecular Ecology Resources*, *9*(1), 51–60. doi: 10.1111/j.1755-0998.2008.02352.x

Weigand, A. M., Jochum, A., Pfenninger, M., Steinke, D., & Klussmann-Kolb, A. (2011). A new approach to an old conundrum-DNA barcoding sheds new light on phenotypic plasticity and morphological stasis in microsnails (Gastropoda, Pulmonata, Carychiidae). *Molecular Ecology Resources*, *11*(2), 255–265. doi: 10.1111/j.1755-0998.2010.02937.x

Wiens, J. J. (2016). Climate-Related Local Extinctions Are Already Widespread among Plant and Animal Species. *PLoS Biology*, *14*(12), 1–18. doi: 10.1371/journal.pbio.2001104

Williams, A. V., Nevill, P. G., & Krauss, S. L. (2014). Next generation restoration genetics: applications and opportunities. *Trends in Plant Science*, *19*(8), 529-537. doi.org/10.1016/j.tplants.2014.03.011

Yang, C., Wang, X., Miller, J. A., Blécourt, M. De, Ji, Y., Yang, C., … Yu, D. W. (2014). Using metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity indicator. *Ecological Indicators*, *46*, 379–389. doi: 10.1016/j.ecolind.2014.06.028

Yoccoz, N. G., Brathen, K. A., Gielly, L., Haile, J., Edwards, M. E., Goslar, T., … Taberlet, P. (2012). DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, *21*, 3647–3655. doi: 10.1111/j.1365-294X.2012.05545.x

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, *3*(4), 613–623. doi: 10.1111/j.2041-210X.2012.00198.x

Zeale, M. R. K., Butlin, R. K., Barker, G. L. A., Lees, D. C., & Jones, G. (2011). Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*, *11*(2), 236–244. doi: 10.1111/j.1755-0998.2010.02920.x

722 **Table 1: Summary of sequencing results for each assay**

Assay	*Sequences*	*Samples (>200reads)*	*ASVs*	*Substrates*
12SV5	268,038	40	221	Scat
ZBJ-Art	5,676,094	160	2,253	Scat, Plant, Pitfall, Vane, Soil
Ant236	4,491,408	119	2,145	Scat, Plant, Pitfall, Vane
trnL	11,660,421	164	546	Scat, Plant, Pitfall, Vane, Soil
rbcL	7,080,330	40	1,064	Plant*
Total	29,176,291	523	6,229	

723 *rbcL was only applied to plant samples and was not tested for the other substrates

724 **Table 2: Taxa of interest**

Taxa	**Common names**	**Substrate**	**Interest**
Dilleniaceae, Fabaceae, Loranthaceae, Myrtaceae, Pinaceae, Poaceae, Proteaceae, Lauraceae	Plant families	All	These families were detected in all substrates
Noctuidae, Geometridae, and Pyralidae	Owlet moths, Geometer moths, Grass moths	Scat	Invertebrate families associated with scat samples ($p<0.05$) that reproduce on plant tissue
Aphididae, Cecidomyiidae, Noctuidae, Clubionidae, Curculionidae, Galumnidae, and Zyganidae.	Aphids, Gallmidges, Owlet moths, Club spiders, Weevils, Mites/ticks, Burnet moths	Plant	Invertebrate families associated with plant samples ($p<0.05$) that all live and reproduce on plant tissue
Macathuria		Plant	Cryptid, often difficult to find because of growth habit. This taxa was found in sequencing data from plant material samples, but not the plant survey results
Vulpes vulpes	Red Fox	Scat	Invasive species that has led to the decline of native fauna

Taxa	Common names	Substrate	Interest
Poaceae, Fabaceae, Asteraceae	Grasses, Legumes, Daisies	Scat	Found in every scat sample in both SCP and the Pilbara, likely common food sources for Macropodidae (Kangaroos and Wallabies)
Pinaceae	Pines	Plant	Detected in plant material sequences but not morphological survey, likely sourced from pine plantations near the SCP site

725

726 **Table 3**: Substrates recommended for eDNA surveys based on target taxa and survey
727 limitations.

	Target taxa			
Limitations	*Invertebrates*	*Plants*	*Vertebrates*	*Total Biodiversity*
None	Pitfall+Vane	Scat+Plant	Scat	Pitfall+Vane+Scat+Plant
DNA Source	Pitfall+Vane	Plant	Scat	Pitfall+Vane+Scat+Plant
Funding	Pitfall/Vane	Scat	Scat	Scat+Pitfall

728 * Colour of substrate recommended indicates appropriate assay: Invertebrate assay, Plant assay,
729 Vertebrate Assay, Multiple assays
730

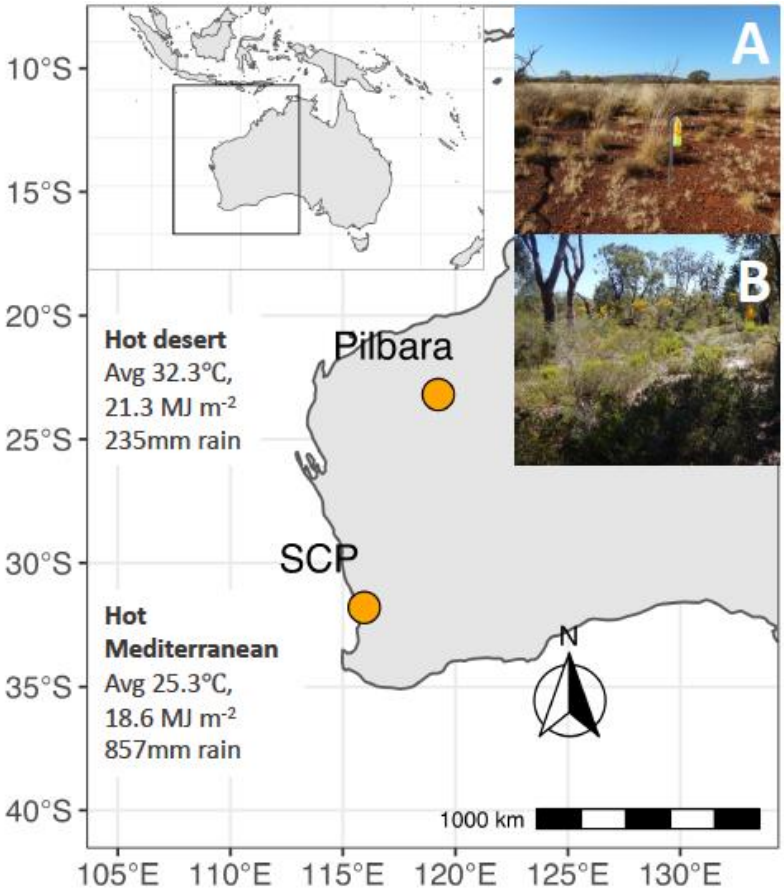


Figure 1: Map of the two study sites in Western Australia. The Pilbara (A) and the Swan Coastal Plain (B) are shown in the photos. Information on the left contains the climate type (Koppen Classification), the average daily maximum temperature, average daily solar exposure, and the total rainfall for the sample year.

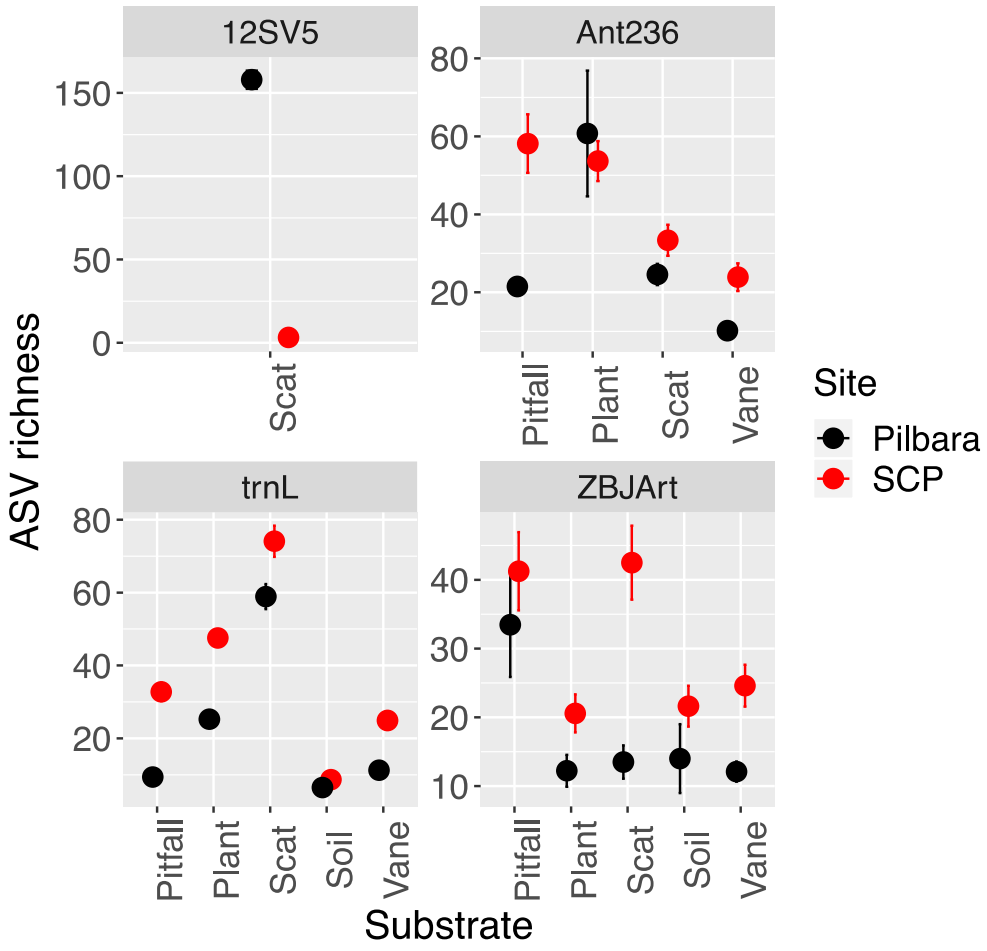
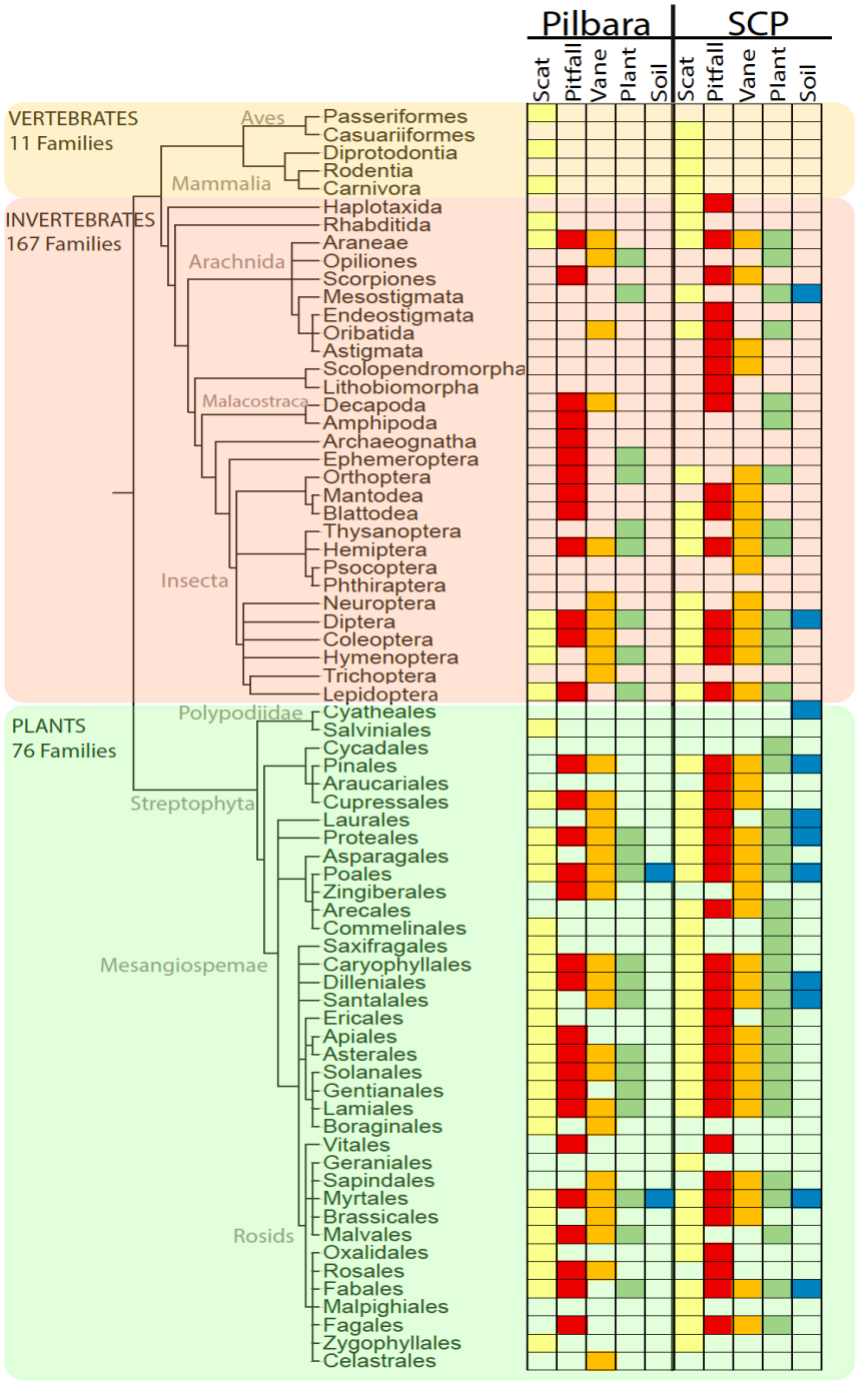


Figure 2: ASV richness of all four barcoding assays in all the substrates. Error bars indicate standard error, red points are SCP samples and black points are from the Pilbara.



740

741 **Figure 3:** Taxonomic orders detected in each substrate. Fungal and algal orders were
742 removed as the assays are not equipped to properly detect fungal diversity. Orders were
743 chosen for this figure because there were too many families to fit in one figure.
744

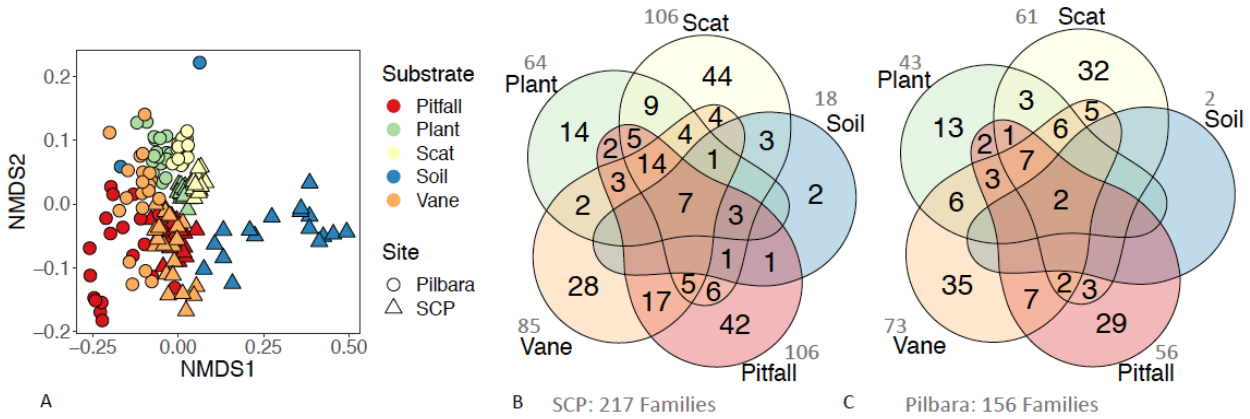


Figure 4: Ordination and venn diagrams of the families detected in the various substrates. A) Nonmetric multidimensional scaling (NMDS) ordination of a presence-absence matrix of families detected in each sample (stress = 0.166, similarity = Bray-Curtis). Venn diagram of families shared between substrates at B) SCP and C) Pilbara site. Gray numbers indicate the total number of families in that substrate.

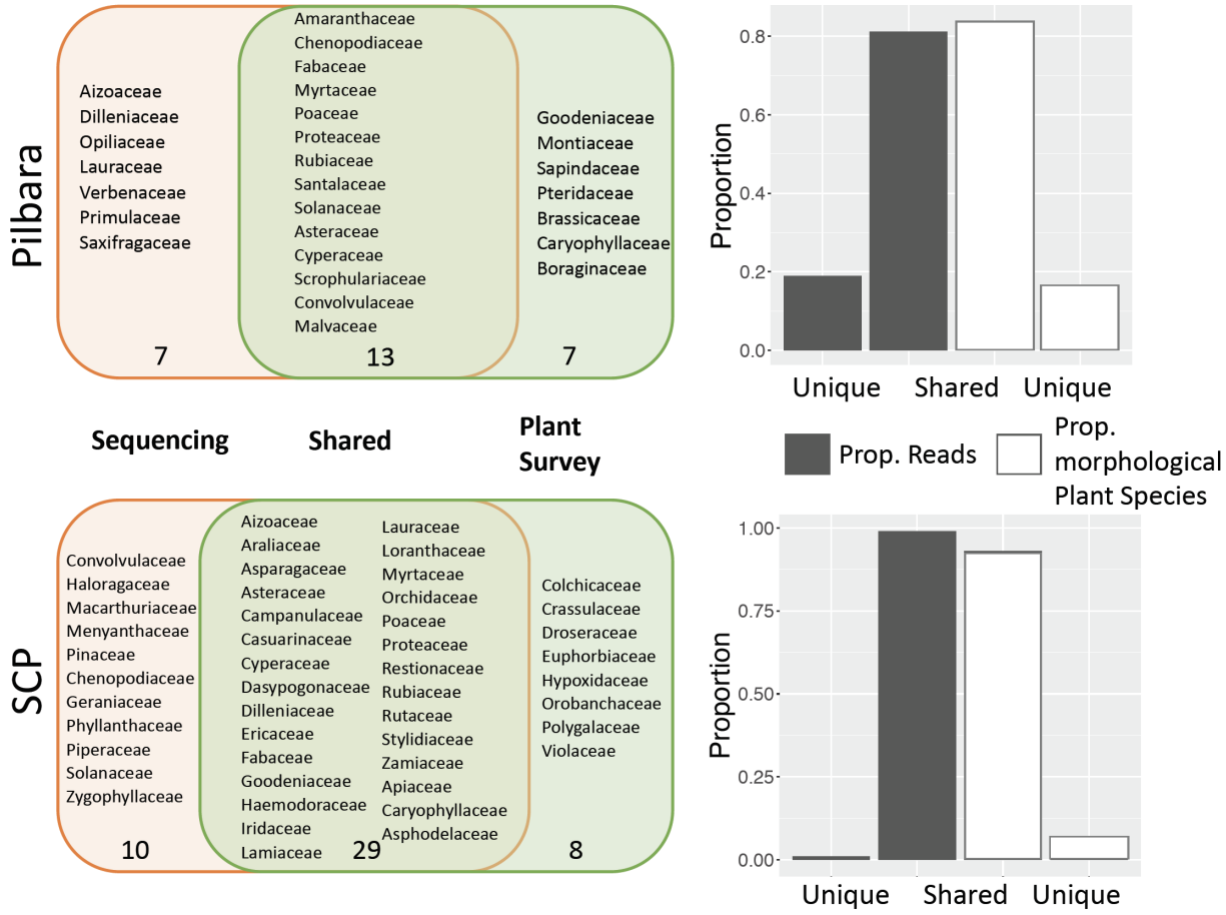


Figure 5: Comparison between plant families found in the morphological plant survey and using sequencing. Numbers indicate the number of families in each category. Bar plots show the proportion (reads for sequencing data, or plant species for morphological survey) from families that were common between sequencing or the morphological plant survey.