Plant Metagenomic Barcoding of pollen loads offers insights on the Foraging Patterns of Queen Bumble Bees in the Southern Rocky Mountains, U.S.A.

Your abstract goes here…

# INTRODUCTION

The inability to reliably identify plants to the level of species often leaves our understanding of ecosystem function and interaction nebulous. Current methods to sidestep this situation include: ignoring these levels of detail, revisiting plots for diagnostic material, assistance from taxonomic specialists, or using barcoding or other molecular techniques. These approaches are unsuitable in light of the benefits offered by species in several complex genera which serve as bioindicators, and preferred partners in ecological interactions. Many genera have species which are well defined based upon ecological rather than morphological properties, the identification of these taxa in degraded areas or without their mutualistic partners is fraught with difficulty, hindering an understanding of the breadth of habitat which some species occupy, and the interactions they have with other species. The identification of many plant species to terminal taxon is an essential component of nearly all land management programs, where many species in the same genus (e.g. Sagebrush - *Artemisia*, Willows - *Salix*, and Sedges - *Carex*) serve as bioindicators, as well as in academic research (Gage & Cooper ([2013](#ref-Gage2013HistoricalRO)), AIM). This endeavour is often mired by lack of diagnostic characters (e.g. flowers, fruits, or roots), and increasingly the description of cryptic species (Janzen *et al.* ([2017](#ref-janzen2017nuclear)), Oliver *et al.* ([2009](#ref-oliver2009cryptic))). Solutions to this problem are wanting, certain programmes have relied increasingly upon revisiting field sites to identify material using morphological or chemical approaches, whereas academic research has often used high copy number plastid genes as barcodes (Rosentreter et al. 2021, ). However, both approaches have significant drawbacks, the former resource intensive at the landscape scale - and often does not work, while the latter seldom works due to a lack of variability in the currently available barcodes (Liu *et al.* ([2014](#ref-liu2014identification))). Recently barcoding, and metabarcoding, have shown much promise in all Kingdoms of life. For example …. . With plants the identification of members of certain clades has been quite successful, whereas with others results have been elusive (Liu *et al.* ([2014](#ref-liu2014identification))), most studies seem to be in between this spectrum (CITE). Herein we have resolved the problem of identifying plant material without morphologically diagnostic character states using the Angiosperms 353 (A353) probes (Johnson *et al.* ([2019](#ref-johnson2019universal))), custom sequence matching databases, and species distribution modelling. A353 probes for hyb seq… Kew Plant and Fungal Tree of Life (PAFTOL) project. BLAST, Kraken and other custom databases can be populated with freely available PAFTOL data, as well as the myriad new phylogenomics projects utilising these probes (SPECIAL ISSUE), and new sequences as required. Species Distribution Modeling (SDM) provides a path forward to develop species lists to create tailored databases for any locale in the world. We tested the ability of A353 probes to identify the plant taxa present in the corbiculae loads of Queen Bumble Bees (*Bombus* sp.) in subalpine settings in the Southern Rocky Mountains. Bumblebees are of concern (Cameron & Sadd ([2020](#ref-cameron2020global)), Goulson *et al.* ([2008](#ref-goulson2008decline))), … Sub-alpine areas of concern ()…

# METHODS

## Study System

Observations and sample collection was conducted at The Rocky Mountain Biological Laboratory (RMBL; 38°57.5” N, 106°59.3” W (WGS 84), 2900 m.a.s.l.), Colorado, USA (see…*APPENDIX 1* for site information). Pollinator observations of *Bombus* spp. (Apidae) were conducted from June - August of 2015. This area is characterised by high-montane/subalpine Parkland vegetation communities.

## Pollen identification

To develop a reference library of pollen grains which may be present in corbiculae loads, an image reference collection of fuchsin-jelly stained (Beattie ([1971](#ref-beattie1971technique))) slides was assembled from slides previously prepared by the authors (n = 2?), and other researchers (n = 38 ) (Brosi & Briggs ([2013](#ref-brosi2013single))). Using 5 years of observational data on Bombus Queen Bee foraging at these studies sites (Ogilvie unpublished), as well as the Vascular Plant Checklist (FRASER BUCK 2007), an additional XXX slides were prepared and imaged at 400x (Leica DMLB, Leica MC170 HD Camera, Leica Application Suite V. 4.13.0) from non accessioned herbarium collections to supplement the number of species and clades covered (Appendix 3).  
In order to determine which plant taxa were distinguishable via light microscopy, and to develop a dichotomous key to pollen morphotypes, Divisive Hierarchical Clustering techniques were used. Ten readily discernible categorical traits were collected from each specimen in the image collection. These traits were transformed using Gower distances, and clustered using Divisive Hierarchical clustering techniques (Maechler *et al.* ([2022](#ref-cluster2022))). Using the cluster dendrogram, elbow plot, and heatmaps (Hennig ([2020](#ref-fpc2022))), of these results morphological groups of pollen which could not be resolved via microscopy were delineated, and a dichotomous key was prepared (APPENDIX NO.). This key was then used to identify the pollen grains sampled from corbiculae loads to morphotypes in a consistent manner. To identify the pollen present in corbiculae loads, light microscopy at 400x (Zeiss Axioscope A1 ) was used. In initial sampling in three transects, each pollen grain was identified to morphotype and counted; an additional two transects were scanned for morphotypes unique to that slide, if either transect contained an unique morphotype than all grains in that transect were also identified and counted. Subsequent to the first round of sampling, non-parametric Species Richness Rarefaction curves (Oksanen *et al.* ([2022](#ref-vegans2022))), and non-parametric Species Diversity rarefaction curves were used to assess the completeness of sampling (Chao *et al.* ([2014](#ref-inextArticle)), Hsieh *et al.* ([2020](#ref-inextPackage))). Slides not approaching the asymptote of the rarefaction curve were then re-sampled, and analysed iteratively for up to a total of seven transects *APPENDIX 2*.

## Pollen Slide Preparation of Corbiculae

To increase sample heterogeneity all corbiculae loads were broken apart and rolled using dissection needlepoints. ~ 0.5mm2 of pollen was placed onto a ~4mm2 fuchsin jelly cube (Beattie ([1971](#ref-beattie1971technique))) atop a graticulated microscope slide, with 20 transects and 20 rows (400 quadrats) (EMS, Hartfield, PA). The jelly was melted until pollen grains were homogeneously spread across the microscope slide. Slides were sealed with Canada Balsam (Rublev Colours, Willits, CA); all samples are noted in *APPENDIX 3*.

## Molecular Lab Work

### Reference Plant Library Generation

Using 5 years of observational data on *Bombus* Queen Bee foraging at these studies sites (Ogilvie unpublished), we identified the plant taxa most frequently visited by Queens across all years. We sequenced the 12 most commonly visited taxa twice using samples from one site within the Gunnison River Drainage and one individual from another population. In addition, for any of these 12 focal species which did not have a congener pair in this filtered sample, we included a congener - or a species from a closely related genus to serve as an outgroup. We also sequenced another 12 abundant taxa commonly visited by *Bombus* workers, based on the aforementioned data set, again using samples from one site within the Gunnison River Drainage and one site from another population (*APPENDIX 4*).

### Plant Genomic DNA Extraction

Plant genomic DNA was isolated from ~ 1 cm2 of leaf tissue from silica-gel dried or herbarium material using a modified cetyltrimethylammonium (CTAB) protocol (Doyle & Doyle ([1987](#ref-doylesCTAB))) that included two chloroform washes. DNA was quantified using a Nanodrop 2000 (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and Qubit fluorometer (Thermo Fisher Scientific).

### Pollen Genomic DNA Extraction

Pollen genomic DNA was extracted from corbiculae using a CTAB based protocol modified from Lahlamgiahi et al. and Guertler et al. (2014, 2014). A SDS extraction buffer (350µL , 100mM Tris-HCl, 50 mM EDTA, 50 mM NaCl, 10% SDS v/v., pH 7.5) was added followed by vortexing to allow dissolution of corbiculae. Pollen grains were then macerated with Kontes Pellet Pestles, and the tip of these washed with 130 µL of the SDS extraction buffer, samples were then incubated for 1 hour at 30°C. This was followed by the addition of 10% CTAB solution (450ul, of 20 mM Tris-Cl pH. 8.0, 1.4 M NaCl, 10 mM EDTA pH 7.5, 10% CTAB, 5% PVP, ~85% Deionized water) and RNAse (10 uL of 10 mg/mL) and samples were incubated for 40 minutes at 37°C, on heat block (Multi-Blok, Thermo Fisher Scientific, Waltham Massachusetts) set to 40°C. After 20 minutes incubation, Proteinase K (15 µL of 20mg/ml) and DTT (12.5 µL of 1M in water) were added, and the samples were further incubated at 60°c for 1 hour. Samples were then incubated overnight at 40°C. 500 µL of Phenol-Chloroform-Isoamyl alcohol (25:24:1) were added, vortexed, and centrifuged at 10,000 rpm for 10 minutes and the aqueous phase was pipetted to a 1.5 ml centrifuge tube.

To precipitate the DNA, chilled Isopropyl alcohol & 3 mM Sodium acetate (5:1) equivalent to 2/3 of the volume of sample were added, with 1 hour of chilling at -20°C, followed by 10 minutes of centrifuging at 13,000 rpm. The supernatant was pipetted to a new 1.5 ml centrifuge tube, and 70% EtOH (400 µL) were added before chilling at -20°C for 20 minutes followed by centrifugation at 13,000 rpm for 10 minutes. Both tubes were then washed with 75% EtOH (400 µL), inverted, centrifuged at 13,000 rpm for 4 minutes, and the solution discarded, then washed with 95% EtOH (400 µL) , inverted, centrifuged at 13,000 rpm for 4 minutes, and the solution discarded. Pellets were dried at room temperature overnight before resuspension in Nuclease free H2O. Extractions were assessed using a Nanodrop 2000 (Thermo Fisher Scientific) and Qubit fluorometer (Thermo Fisher Scientific). DNA extracts were then cleaned using 2:1 v./v. Sera-Mag beads (Cytiva, Little Chalfont, UK) to solute following the manufacturer’s protocol, eluted in 0.5x TE, and the eluent allowed to reduce by half volume in ambient conditions. DNA was quantified using a Qubit fluorometer.

### Fragmentation, Library Preparation & Target Enrichment

Library preparation was performed using the NEBNext Ultra II FS-DNA Library Prep Kit for Illumina (New England BioLabs, Ipswich, Massachusetts, USA) using slightly modified manufacturers recommendation. Fragmentation was performed at ½ volume of reagents and ¼ enzyme mix for 40 minutes at 37\*C, with an input of 500 ng cleaned DNA. Adapter Ligation and PCR enrichment were performed with ½ volumes, while cleanup of products was performed with ½ volume of SPRI beads (Beckman Coulter, Indianapolis, Indiana, USA) and recommended volumes of 80% v./v. ethanol washes. The exception was the herbarium specimens which were not fragmented and only end repaired, with similar library preparation of all samples. Products were analysed on 4% agarose gels, and a Qubit fluorometer.

Libraries were pooled and enriched with the Angiosperms 353 probe kit V.4 (Arbor Biosciences myBaits Target Sequence Capture Kit) by following the manufacturer’s protocol and Brewer et al. 2019. Sequencing was performed using an Illumina mi-Seq with 150-bp end reads, (NUSeq Core, Chicago, Illinois).

## Computational Processes and Analyses.

### Reference Library Data Processing

Sequences were processed using using Trimmomatic, which removed sequence adapters, clipped the first 3 bp, discarding reads less than 36 bp, and removing reads if their average PHRED score dropped beneath 20 over a window of 5 bp (Bolger & Giorgi ([2014](#ref-bolger2014trimmomatic))). Contigs were generated using HybPiper using target files created by M353 (Johnson *et al.* ([2016](#ref-johnson2016hybpiper)), McLay *et al.* ([2021](#ref-mclay2021new))).

### Sequence Identification

A custom Kraken2 database was created by downloading representative species of each genus indicated as being present in the study area by the spatial analyses from the Sequence Read Archive (SRA) NCBI (Wood *et al.* ([2019](#ref-wood2019improved))). These sequences were processed in the same manner as our novel sequences before being placed into the database. The Kraken2 database was built using default parameters. Kraken2 was run on sequences using default parameters (*APPENDIX 5*). Following Kraken2, Bracken was used to classify sequences to terminal taxa (Lu *et al.* ([2017](#ref-lu2017bracken))). Results from both Kraken2 and Bracken, results were reclassified manually to identify terminal taxa. For example, when only a single species of a genus was known in the study area, but our database used a representative of another taxon in the genus, this species was coded as the result. The re-coding of sequences from another representative species for the genus to the sole RMBL representative allowed the identification of XX & % more species.

### Identification of Sequence Matching Loci

A local NCBI database was built using the same processed novel and downloaded sequences (Camacho *et al.* ([2009](#ref-camacho2009blast))).

### Spatial Analyses

In order to develop an ecologically relevant list of vascular plant species which may be present at the study sites all records adjacent to the field site were downloaded from BIEN (Maitner ([2022](#ref-bien2022))), and these taxa had Species Distribution Models generated to infer their suitability. This list of species served as a reference for which species to include in the Kraken2 database.

In order to minimise the number of species for which SDM’s were to be generated, BIEN was queried at a distance of up to 100km from our field site and all plant species records were downloaded. In order to estimate the stochasticity of collections, this dataset was bootstrap re-sampled 250 times, with 90% of samples selected, to create a testing dataset. Using logistic regression, and a Vascular Plant Species list of the species present at RMBL (FRASER BUCK 2007) we determined a distance of ca. 25km was required to meet the centre of the sigmoid line fit (wtf is this called? - where success is 0.5). The species found within the 100 km of RMBL were reduced to those with a record within 25km and these 500 species were subjected to species distribution modelling. LINEAR MODEL FIGURE

Species had all records from BIEN within a 50km border of the Omernik level 4 ecoregion, which the site is located in, downloaded (23k records), (Omernik ([1987](#ref-omernik1987ecoregions))). These records were split into two sets, one for generating machine learning models, and the other for GLM and GAM models. The set for generating GLM and GAM records was thinned to reduce spatial autocorrelation in the dataset. To both datasets an additional 4029 plots collected from a random stratification of 19% of the land cover in the area of analysis were searched to create TRUE ABSENCES (BLM CITATION). To buffer these absences an additional 1000 pseudo-absence records were generated for each taxon, each of which was greater than 10km from an occurrence record. For ML models, these pseudo-absences were reduced so that the number of presence to absence records were 1:1; to achieve this, records inside of 10% of the mean value for any predictor were removed.

Species predictors were a stack of 26 variables at 30m resolution, for linear regression models these predictors underwent both vifstep (theta = 10, max observations = 12,500) and vifcor (theta = 0.7, max observations = 12,500), and collinear features were removed leaving 16 variables (Naimi *et al.* ([2014](#ref-usdm2014))) (*APPENDIX 6*).

Modelling: Random Forest and Boosted Regression Trees, were sub sampled with 30% test and two replicates each before weighted ensemble based on True Skill Statistics (tss) (Naimi & Araujo ([2016](#ref-sdmPackage))). Generalised linear models and Generalised additive models with 30% sub sampling and three replicates each were also ensemble using tss (SDM citation). The results of these models were extracted to the study area, and species from either ensemble with greater than 50% habitat suitability were considered present for further purposes. To evaluate these results GBIF (n = ) results which had not made it to the BIEN database yet were evaluated using logistic regression (GBIF).

# Results

## Microscopic Pollen identification

**Using the fuchsin jelly preparation and light microscopic analyses of grains and scoring of 12 character states resulted in the establishment of XX morphotypes which grains could be reliably classified into. *APPENDIX 7*** . XX Samples were counted and based on rarefaction had over % of expected morphotypes found. The relative abundance of pollen grains in each sample (max % of any species, mean % of all species, min % trace amounts detected).

## Metabarcoding Pollen identification

Kraken2 was able to identify the species richness of pollen samples (mean = , min = , max = ). Bracken was able to estimate the relative abundance of pollen grains in each sample (max % of any species, mean % of all species, min % trace amounts detected).

## Spatial Analyses

Results of Logistic Regression Assessing Accuracy of Species Distribuion Modelling.

Metric

Value

Metric

Value

Accuracy (Training)

83.75

F-Score

0.84

Accuracy (Test)

84.00

AUC

0.92

Recall

81.03

Concordance

0.92

True Neg. Rate

86.97

Discordance

0.08

Precision

88.04

Tied

0.00

The median (24.XX km - or was it 3rd quartile?) of the logistic regression assessing the probability of occurrence of a species record as a function of distance from the study area was used as a threshold distance to include species for distribution modelling. Of the 500 species modelled % were accurately classified as being present at the field station, while % were accurately classified as being absent from the field station *TABLE XX* for more results. Based upon extracting

Species

Future Directions: While at the time of writing this there are limited A353 sequence data, the Plant and Fungal Trees of Life (PAFTOL) project, which is sequencing at least a species of each genera in the plant Kingdom will produce sequence data from over 14,000 species. Given the extant publicly available genomic data, we conservatively estimate that upon completion of PAFTOL there will be no fewer than 15,500 species (4.4% of all ca. 350,000 plant species) for which sequence data of a majority of these loci exist. Accordingly, projects in the near future may increase the number of metagenomics samples while decreasing the need to create their own plant sequence reference libraries. As a result of PAFTOL the first ever comprehensive phylogenetic hypotheses of all plant genera will be presented. In tandem with an increased number of digitised and geo-referenced herbarium specimens, and monitoring programs in natural areas, we believe that geo-informatics, and phylogenetic inference will increase the ability of researchers applying this technique to identifying sequence reads . While our approach emphasises the use of this metagenomic technique for the purpose of identifying pollen, I argue the template and resources we provide here make this approach a suitable candidate for many plant metagenomic tasks. While we did not have the resources to explore the possibility of characterising infraspecific characteristics, preliminary results from others (Wenzell et al. in prep., Loke et al. in prep) indicate a possibility for these probes to also collect data at the level of populations and individuals.

In regards to better understanding the foraging preferences of *Bombus* feeding in subalpine ecosystems. **JANE AND PAUL SET UP FOR NEAR FUTURE RESULTS?**

**Author Contributions:** J.E.O conceived, designed, and conducted all ecological fieldwork, assisted with analyses, and writing. R.C.B conducted botanical collections, conducted all molecular lab work, all analyses, and writing. E.W. prepared, imaged, and collected trait data on pollen reference slides, and assisted with analysis of trait data and writing a dichotomous key. P.J.C & S.T. assisted with analyses and writing. J.B.F. conceived, and designed all lab work, analyses, assisted with writing, and secured funding for molecular work.

**Acknowledgements:** Nyree Zerega for assistance obtaining herbaria loans and accessioning our collections at CHIC. We thank the curators at the following herbaria for supplying tissue: Ben Legler at Stillinger (ID), Charles Williams at Ray J. Davis (IDS), Ernie Nelson at Rocky Mountain (RM). We thank the original collectors of these specimens: D. Knoke, L. Brummer, J. Boyd, C. Davidson, I. Gilman, M. Kirkpatrick, S. McCauley, J. Smith, K. Taylor, & C. Williams. David Giblin for sharing relevant sections of an advanced draft of FNA V. 15. Zoe Diaz-Martinez, Angela McDonnel, & Elena Loke for assistance with genomic library preparation. The Bureau of Land Management is thanked as many plant specimens were collected by R.C.B as a partner or contractor to the agency. Sanda and NEB are gratefully acknowledged for technical support and generously sharing samples.

# References

Beattie, A. (1971). A technique for the study of insect-borne pollen. *The Pan-Pacific Entomologist*, **47**, 82.

Bolger, A. & Giorgi, F. (2014). Trimmomatic: A flexible read trimming tool for illumina NGS data. *Bioinformatics*, **30**, 2114–2120.

Brosi, B.J. & Briggs, H.M. (2013). Single pollinator species losses reduce floral fidelity and plant reproductive function. *Proceedings of the National Academy of Sciences*, **110**, 13044–13048.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. (2009). BLAST+: Architecture and applications. *BMC bioinformatics*, **10**, 1–9.

Cameron, S.A. & Sadd, B.M. (2020). Global trends in bumble bee health. *Annual review of entomology*, **65**, 209–232.

Chao, A., Gotelli, N.J., Hsieh, T.C., Sande, E.L., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014). Rarefaction and extrapolation with hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, **84**, 45–67.

Doyle, J.J. & Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

Gage, E. & Cooper, D.J. (2013). Historical range of variation assessment for wetland and riparian ecosystems, u.s. Forest service rocky mountain region

Goulson, D., Lye, G. & Darvill, B. (2008). The decline and conservation of bumblebees. *Annual review of entomology*, **53**, 191–208.

Hennig, C. (2020). *Fpc: Flexible procedures for clustering*. Retrieved from <https://CRAN.R-project.org/package=fpc>

Hsieh, T.C., Ma, K.H. & Chao, A. (2020). *iNEXT: Interpolation and extrapolation for species diversity*. Retrieved from <http://chao.stat.nthu.edu.tw/wordpress/software_download/>

Janzen, D.H., Burns, J.M., Cong, Q., Hallwachs, W., Dapkey, T., Manjunath, R., Hajibabaei, M., Hebert, P.D. & Grishin, N.V. (2017). Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. *Proceedings of the National Academy of Sciences*, **114**, 8313–8318.

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J. & Wickett, N.J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in plant sciences*, **4**, 1600016.

Johnson, M.G., Pokorny, L., Dodsworth, S., Botigue, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L., Epitawalage, N., Forest, F., Kim, J.T. & others. (2019). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic biology*, **68**, 594–606.

Liu, J., Shi, L., Han, J., Li, G., Lu, H., Hou, J., Zhou, X., Meng, F. & Downie, S.R. (2014). Identification of species in the angiosperm family apiaceae using DNA barcodes. *Molecular ecology resources*, **14**, 1231–1238.

Lu, J., Breitwieser, F.P., Thielen, P. & Salzberg, S.L. (2017). Bracken: Estimating species abundance in metagenomics data. *PeerJ Computer Science*, **3**, e104.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. (2022). *Cluster: Cluster analysis basics and extensions*. Retrieved from <https://CRAN.R-project.org/package=cluster>

Maitner, B. (2022). *BIEN: Tools for accessing the botanical information and ecology network database*. Retrieved from <https://CRAN.R-project.org/package=BIEN>

McLay, T.G., Birch, J.L., Gunn, B.F., Ning, W., Tate, J.A., Nauheimer, L., Joyce, E.M., Simpson, L., Schmidt-Lebuhn, A.N., Baker, W.J. & others. (2021). New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in plant sciences*, **9**.

Naimi, B. & Araujo, M.B. (2016). [Sdm: A reproducible and extensible r platform for species distribution modelling](https://doi.org/10.1111/ecog.01881). *Ecography*, **39**, 368–375.

Naimi, B., Hamm, N. a.s., Groen, T.A., Skidmore, A.K. & Toxopeus, A.G. (2014). [Where is positional uncertainty a problem for species distribution modelling](https://doi.org/10.1111/j.1600-0587.2013.00205.x). *Ecography*, **37**, 191–203.

Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O’Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M.-H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C.J.F. & Weedon, J. (2022). *Vegan: Community ecology package*. Retrieved from <https://CRAN.R-project.org/package=vegan>

Oliver, P.M., Adams, M., Lee, M.S., Hutchinson, M.N. & Doughty, P. (2009). Cryptic diversity in vertebrates: Molecular data double estimates of species diversity in a radiation of australian lizards (diplodactylus, gekkota). *Proceedings of the Royal Society B: Biological Sciences*, **276**, 2001–2007.

Omernik, J.M. (1987). Ecoregions of the conterminous united states. *Annals of the Association of American geographers*, **77**, 118–125.

Wood, D.E., Lu, J. & Langmead, B. (2019). Improved metagenomic analysis with kraken 2. *Genome biology*, **20**, 1–13.