

RESEARCH ARTICLE

Investigating the use of pollen DNA metabarcoding to quantify bee foraging and effects of threshold selection

Katherine A. Arstingstall^{1*}, Sandra J. DeBano^{1,2}, Xiaoping Li^{2,3}, David E. Wooster^{1,2}, Mary M. Rowland⁴, Skyler Burrows⁵, Kenneth Frost^{2,3}

1 Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, Oregon, United States of America, **2** Hermiston Agricultural Research and Extension Center, Oregon State University, Hermiston, Oregon, United States of America, **3** Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon, United States of America, **4** United States Forest Service, Pacific Northwest Research Station, La Grande, Oregon, United States of America, **5** Bee Biology and Systematics Lab, Utah State University, Logan, Utah, United States of America

* kaarst01@gmail.com



OPEN ACCESS

Citation: Arstingstall KA, DeBano SJ, Li X, Wooster DE, Rowland MM, Burrows S, et al. (2023) Investigating the use of pollen DNA metabarcoding to quantify bee foraging and effects of threshold selection. PLoS ONE 18(4): e0282715. <https://doi.org/10.1371/journal.pone.0282715>

Editor: Wolfgang Blenau, University of Leipzig Faculty of Life Sciences: Universitat Leipzig Fakultät für Lebenswissenschaften, GERMANY

Received: June 10, 2022

Accepted: February 22, 2023

Published: April 18, 2023

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: Sequencing results have been deposited in the NCBI Sequence Read Archive under BioProject PRJNA852949. The trained databases, command line arguments and Python code have been deposited in the Figshare repository: <https://doi.org/10.6084/m9.figshare.20280051.v2>.

Funding: Funding for this project was provided by the USDA Forest Service, Pacific Northwest Research Station (SJD): <https://www.fs.usda.gov/>

Abstract

DNA metabarcoding of pollen is a useful tool for studying bee foraging ecology. However, several questions about this method remain unresolved, including the extent to which sequence read data is quantitative, which type of sequence count removal threshold to use and how that choice affects our ability to detect rare flower visits, and how sequence artefacts may confound conclusions about bee foraging behavior. To address these questions, we isolated pollen from five plant species and created treatments comprised of pollen from each species alone and combinations of pollen from multiple species that varied in richness and evenness. We used ITS2 and *rbcL* metabarcoding to identify plant species in the samples, compared the proportion of pollen by mass to the proportion of sequencing reads for each plant species in each treatment, and analyzed the sequencing data using both liberal and conservative thresholds. We collected pollen from foraging bees, analyzed metabarcoding data from those samples using each threshold, and compared the differences in the pollinator networks constructed from the data. Regardless of the threshold used, the relationship between the proportion of pollen by mass and sequencing reads was inconsistent, suggesting that the number of sequence reads is a poor proxy for pollen abundance in mixed-species samples. Using a liberal threshold resulted in greater detection of original plant species in mixtures but also detected additional species in mixtures and single-species samples. The conservative threshold reduced the number of additional plant species detected, but several species in mixtures were not detected above the threshold, resulting in false negatives. Pollinator networks produced using the two thresholds differed and illustrated tradeoffs between detection of rare species and estimation of network complexity. Threshold selection can have a major effect on conclusions drawn from studies using metabarcoding of bee pollen to study plant-pollinator interactions.

pnw/, the Northwest Potato Research Consortium (KF): <https://www.nwpotatoresearch.com/>, the Oregon Department of Forestry (SJD): <https://www.oregon.gov/odf/pages/index.aspx>, Oregon State University's Agricultural Research Fund (SJD): <https://agresearchfoundation.oregonstate.edu/>, a USDA Western Sustainable Agriculture Research and Education Graduate Student Grant (#GW19-188) (KAA): <https://western.sare.org/>, a Mazamas Research Grant (KAA): <https://mazamas.org/>, the Oregon Zoo Foundation's Future for Wildlife Pacific Northwest Fund (SJD): <https://www.oregonzoo.org/>, and the Oren Pollak Memorial Student Research Grant for Grassland Science, The Nature Conservancy (KAA): <https://www.nature.org/en-us/>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Bee populations worldwide are currently experiencing significant declines [1, 2], and one way to reverse these declines is to increase the quantity and quality of their habitat by planting flowering species that are important food sources for bees [3]. However, determining which plant species are significant food sources for bees can be challenging. Traditional methods for describing bee-flower interactions (e.g., visual observation, microscopy) are time-consuming, can require specialized expertise, and often lack taxonomic resolution [4–6]. However, the recent use of DNA metabarcoding on bee pollen is potentially a more effective method for identifying food sources for bees than traditional methods, resulting in higher taxonomic resolution and revealing a more detailed record of bee foraging behavior [7–13].

Pollen metabarcoding has become a widely used method that is applicable to many different organisms and areas of research including, but not limited to, allergen monitoring [14], plant biodiversity assessments [15], and biomonitoring of plant pathogens [16]. A number of questions remain about pollen metabarcoding analyses in general and potential limitations of metabarcoding to establish floral resource preferences of bees, specifically. First, the issue of whether sequence read data can be used to quantify the amount of each plant species in a mixed pollen load is unresolved. Over the past decade, researchers have investigated the quantitative abilities of pollen DNA metabarcoding using six different gene regions, three nuclear ribosomal and three plastid, and have yet to come to a consensus (Table 1). Furthermore, studies that use the same barcode markers continue to find contradictory results. For example, Keller et al. [7] found similar relative abundances when comparing ITS2 sequencing reads to microscopic pollen counts, while Richardson et al. [8] found no association.

Sequence read data may not accurately quantify the amount of pollen from each plant species in a mixed pollen load when certain plant species are over- or underrepresented [8, 11, 18]. Misrepresentation might occur for several reasons including variable gene copy number [22, 23], differences in DNA extraction efficiency [24], and primer amplification bias [25, 26]. Brooks et al. [24] found that different combinations of DNA extraction kits and number of

Table 1. Results of studies examining quantitative relationship between attributes of pollen in mixed samples and reads of genetic barcode markers.

Study	Barcode Marker	Support of Quantitative Relationship
Keller et al. [7]	ITS2	Yes—similar relative abundance of reads compared to microscopic counts
Kraaijeveld et al. [17]	<i>trnL</i>	Yes—found that correcting the number of sequencing reads by the total number of microscopically counted pollen gave an accurate estimate of the absolute number of each pollen type in a sample
Richardson et al. [8]	ITS2	No—no relationship in relative abundance of reads compared to microscopic counts
Richardson et al. [9]	ITS2, <i>matK</i> , and <i>rbcL</i>	No—certain species were consistently over- or underrepresented in the molecular results relative to the microscopic results
Pornon et al. [10]	ITS1 and <i>trnL</i>	Yes—positive relationship between amount of DNA isolated from each plant species and reads of both markers
Bell et al. [18]	ITS2 and <i>rbcL</i>	No—even though significant correlations between sequence reads and proportion of pollen grains for each species, variance explained was low
Richardson et al. [19]	ITS2, <i>rbcL</i> , <i>trnH</i> , and <i>trnL</i>	Yes—statistically significant relationship between microscopy pollen counts and median number of sequence reads for family level
Baksay et al. [20]	ITS1 and <i>trnL</i>	Yes—found a significant positive relationship between the number of DNA sequences and the number of pollen grains in mock solutions
Polling et al. [21]	nrITS2 and <i>trnL</i>	Yes—found highly significant positive relationships between the relative abundance of sequencing reads and the relative abundance of microscopically obtained pollen concentrations

<https://doi.org/10.1371/journal.pone.0282715.t001>

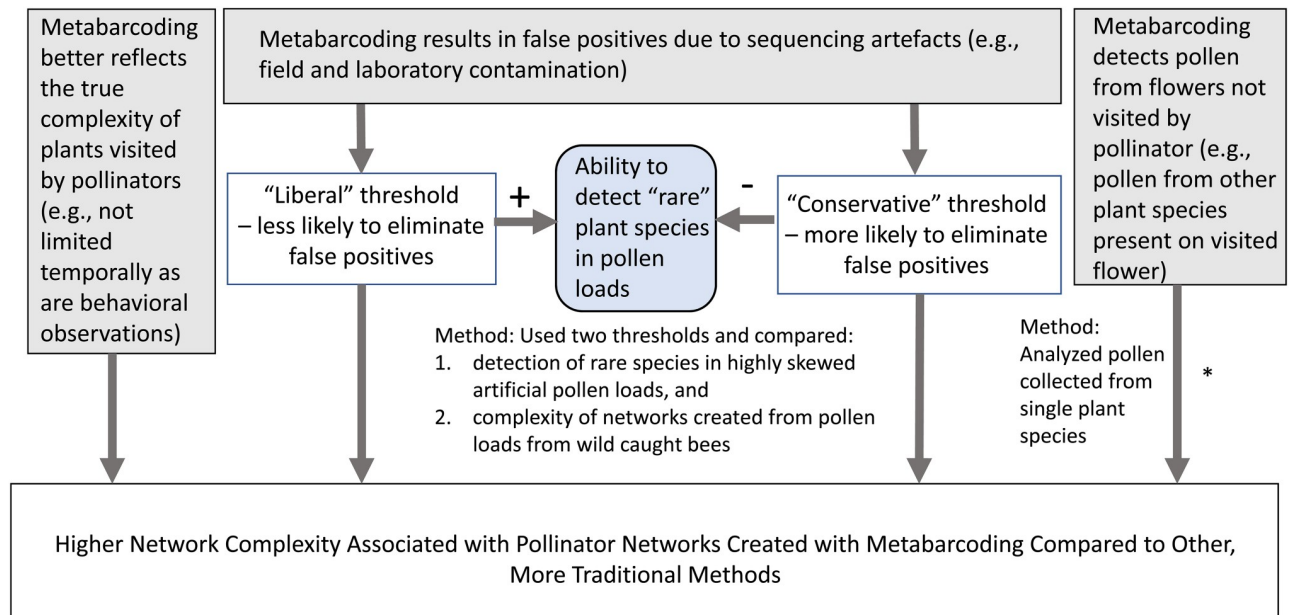


Fig 1. Three general explanations for greater complexity associated with pollinator networks derived from metabarcoding. Three general explanations for the pattern of higher complexity often associated with pollinator networks based on metabarcoding data compared to more traditional methods. Possible explanations are shown in gray boxes. Explanations are not mutually exclusive. The potential impacts of “liberal” and “conservative” thresholds (employed to decrease false positives) on the ability to detect rare species (shown in the blue rounded box) are depicted with signed arrows. “Rare” plant species in the network are those that may be visited by only a few individuals of a particular pollinator species, or may be visited by many species, but at relatively lower rates compared to other plant species. *This process is also influenced by threshold selection, but is not shown to simplify the figure.

<https://doi.org/10.1371/journal.pone.0282715.g001>

polymerase chain reaction (PCR) cycles resulted in dramatically different proportions of sequencing reads per taxa. Additionally, sequence variation at the barcode priming site can affect the amplification efficiency of a plant species, resulting in false negatives [26, 27]. Nevertheless, several studies have found definitive quantitative relationships and proposed that number of sequence reads are a good proxy for the relative abundance of a species in a pollen load [7, 10, 17]. Resolving the issue of whether DNA metabarcoding of bee pollen produces accurate quantitative results will allow us to determine the types of questions that can be answered using this approach and rule out those that cannot.

Another category of issues that must be addressed relates to how well plant-pollinator networks produced from plant species assignments obtained using metabarcoding data represent pollinator foraging behavior (Fig 1). Several studies have found that plant-pollinator networks based on metabarcoding data are more complex, containing more plant taxa and interactions than networks created using foraging observations [12, 13, 28]. Although this pattern may occur simply because metabarcoding studies provide a more accurate depiction of bee foraging behavior than traditional methods, the tendency of metabarcoding networks to be more complex than those based on foraging observations alone could be influenced by other factors (Fig 1). One potential factor is sequence artefacts associated with field and laboratory processes (e.g., contamination) (Fig 1). Contamination that occurs in the field and laboratory is often addressed by using negative controls at each stage of the metabarcoding process (DNA isolation, PCR, plating), which can be used in various ways to quantify the number of sequencing reads in a sample that can be attributed to contamination alone. The number of sequencing reads found in negative controls can be used to create a sequence count removal threshold,

and any taxonomic assignment whose read count falls below this threshold can be considered contamination or other “background noise” and removed from further analysis [18]. However, there is currently no standardized protocol for creating these thresholds in pollen metabarcoding studies, despite the fact that different threshold protocols can potentially yield different interpretations of the same results.

One general approach is to remove entire plant taxa from the dataset if the total number of reads for the taxa is less than a threshold. For example, Richardson et al. [19] discarded “taxonomic groups represented by <0.01% of the data” after first removing any taxonomic groups that were assigned using only one of the four barcode markers in the study. Similarly, Pornon et al. [10] eliminated plant species with less than 1,000 sequencing reads. In contrast, another general approach is to compare the number of reads found in each individual taxonomic assignment to a threshold. Bell et al. [18] took this approach by using the maximum number of sequencing reads in their negative controls as their threshold, and setting any taxonomic assignment in a sample with fewer reads than this threshold to zero. Macgregor et al. [29] also used the maximum number of reads in their controls to create their sequence count removal threshold and set any individual assignment falling below it to zero. Because one approach removes entire taxonomic groups while the other removes specific plant-pollinator interactions, different results and interpretations of pollinator behavior may result. Although multiple studies have used similar versions of the same techniques, threshold choice appears to be somewhat arbitrary. Identifying a standardized threshold protocol would help eliminate these problems and allow for comparison of pollinator foraging behavior (e.g., plant-pollinator networks) across multiple pollen metabarcoding studies.

An additional consideration in selecting a contamination threshold approach is that its choice also influences the likelihood of eliminating rare interactions occurring at low frequencies [22] (Fig 1). One way to examine the potential trade-off associated with threshold selection and rare species detection is to sequence laboratory mixtures with plant species present in amounts small enough to represent a rare plant-pollinator interaction and compare each threshold’s ability to detect these species. This approach could help determine whether rare interactions can be detected using DNA metabarcoding of pollen or if the number of sequencing reads produced by the pollen collected during these rare interactions might fall below the sequence count removal threshold.

Another way that metabarcoding could potentially overestimate pollinator network complexity is by detecting pollen dispersed by wind and other insects [30–32]. A flower could have pollen present from several other plant species before a pollinator visits it (Fig 1). Thus, a pollinator could inadvertently pick up pollen from multiple plant species in one flower visit. Despite this possibility, no studies have isolated pollen collected from flowering plants in natural settings and sequenced their pollen to determine if additional plant species were detected and how that detection is affected by the type of contamination threshold used.

Here, we collected pollen by hand from five plant species to create single-species pollen samples and five mixtures of known concentrations, varying in species richness (3–5 species) and evenness (uniform to highly skewed). We used DNA metabarcoding of the ITS2 gene and *rbcL* region, separately, to identify the plant species in the samples, and analyzed the sequencing data using both a liberal and a conservative sequence count removal threshold. We also collected pollen from foraging bees and used metabarcoding data from those samples to construct plant-pollinator networks to compare consequences of using the two thresholds. The specific objectives of this study were to determine whether 1) the proportions of plant species in a pollen mixture based on mass correspond with proportions of plant species in a pollen mixture based on number of sequencing reads, 2) the two types of thresholds yield different results with regard to detecting rare plant-bee interactions and portraying pollinator network

complexity, and 3) metabarcoding detects additional plant species in single-species pollen samples, and if so, how that is influenced by threshold selection.

Materials and methods

Plant and invertebrate materials were collected on public lands managed jointly by the U.S. Forest Service Wallowa-Whitman National Forest and Pacific Northwest Research Station, with their permission. Such collections do not require a field permit or special use authorization.

Flower collection

Flowers were collected in July 2017 from the United States Forest Service (USFS) Starkey Experimental Forest and Range (Starkey) and from Hermiston, both located in eastern Oregon. Slender cinquefoil (*Potentilla gracilis*), Oregon checkerbloom (*Sidalcea oregana*), and mountain goldenbanner (*Thermopsis montana*) flowers were collected at Starkey (45.2332°N, 118.5511°W). Starkey is a long-term research site in Union County (elevation 1,130–1,500 m) that was established in 1940 [33]. Sampling took place at sites along Meadow Creek, a major tributary of the Upper Grande Ronde River that flows through Starkey. Hairy vetch (*Vicia villosa*) and Scotch thistle (*Onopordum acanthium*) flowers were collected in Hermiston (45.8169°N, 119.2846°W). Flowers from each plant species were collected by hand and stored in separate plastic bags. Each bag contained flowers sampled from multiple plants of the same species. Flowers were stored at -20°C until further processing.

Isolating pollen from flowers

All tools used during pollen isolation were placed in 10% bleach for at least one min prior to use and in between each bout of pollen isolation to reduce contamination. Stamens were removed from each flower using forceps and placed in a glass vial filled with sterilized water. All stamens from each plant species were placed in the same vial. Vials were shaken vigorously for one min to detach pollen from anthers. The solution was poured through a fine mesh sieve into a vacuum filtration system. The pollen was collected on a 5-μm mixed cellulose ester filter. The filters were then placed in a drying oven at 65°C for approximately two hours. Once dry, filters were placed in individual 50 mL centrifuge tubes, and 10 mL of acetone was added to each tube. The centrifuge tubes were vortexed until the filters were completely dissolved, and the solution was mixed thoroughly. The tubes were then centrifuged for 3 min at 2,000 rpm (448 x g), and the supernatant was discarded. Then, 1 mL of acetic acid was added to each tube, and the pellet was re-suspended. The solution was transferred to a 1.5 mL microcentrifuge tube. The solution was centrifuged for 30 s at 12,500 rpm (14,403 x g), and the supernatant was discarded. Two washes were performed with 1 mL of sterilized water for 30 s at 12,500 rpm (14,403 x g). This was followed by two washes with 1 mL of ethanol for 30 s at 12,500 rpm (14,403 x g) and 3 min at 13,500 rpm (15,555 x g). The supernatant was discarded, and the pollen pellet was dried for 30 min. The mass of the dried pollen pellet was calculated as ((mass of pollen pellet + microcentrifuge tube)—mass of microcentrifuge tube) and recorded.

Preparation of single-species samples and mixtures

Single-species samples were created for each of the five plant species (three replicates each) by transferring ~1.5 mg of the stock pollen to a 1.5 mL screw cap microcentrifuge tube using a scoopula and a balance scale. All tools were placed in 10% bleach for at least one min prior to use and between different plant species. Next, a ~10 mg stock was created for each of the five

pollen mixtures representing a range of species richness (3–5 species) and evenness (uniform to highly skewed). Mixture 1 was a slightly skewed mixture with pollen from three different plant species, mixture 2 was a slightly skewed mixture with pollen from four different plant species, mixture 3 was a uniform mixture with pollen from all five plant species, and mixtures 4 and 5 were highly skewed mixtures with pollen from all five plant species. See table in [S1 Table](#) for total mass and proportions of each mixture stock. Mixture stocks were vortexed for 30 s to ensure even concentration of each species. We transferred ~2.5 mg of pollen from each mixture stock to a 1.5 mL screw cap microcentrifuge tube as described above (three replicates of each mixture). Mass of samples and rank abundance of plant species within the samples were determined based on the amount of pollen isolated from each plant species.

Bee sampling

Bees of varying taxa were sampled during peak foraging hours (0900–1800) from Starkey on June 25, 2017. Each bee was either collected directly into an individual glass vial or caught with an insect net and then transferred into a vial. Each bee was given a unique identification number so that it could be associated with the metabarcoding data obtained from its pollen load. Bees were pinned, sexed and identified to species using the methods described in Kuhlman and Burrows [34].

Isolating pollen from bees

Pollen grains were isolated from each bee's body by either shaking vigorously in a vial filled with sterilized water or scraping pollen loads from the corbicula directly into a vial for small/medium sized bees and large bees, respectively. Small/medium bees were removed from the vials, and the pollen solution was transferred to a 50 mL centrifuge tube. Each 50 mL tube was centrifuged for 4 min at 2,000 rpm (448 x g), and the supernatant was discarded. We added 1 mL of water to each 50 mL centrifuge tube to resuspend the pollen pellet. The solution was transferred to a 1.5 mL microcentrifuge tube that was weighed beforehand. The 1.5 mL centrifuge tubes were centrifuged for 3 min at 13,500 rpm (15,555 x g), and the supernatant was discarded. The pollen pellets were dried for 1 hour in an Eppendorf vacufuge. The mass of the dried pollen pellet was calculated as described above and recorded. Bee specimens and pollen pellets were stored at -20°C until further processing.

DNA extraction and PCR

DNA were extracted from pollen samples using the Macherey-Nagel Nucleospin Food kit (Macherey-Nagel, Bethlehem, Pennsylvania, USA), following the “isolation of genomic DNA from honey or pollen” supplementary protocol. The samples were homogenized in a Mini-BeadBeater-24 (Biospec Products, Bartlesville, Oklahoma, USA). Negative controls (i.e., sterilized water in place of pollen) were included with each round of DNA extraction. For library preparation, we used internal transcribed spacer (ITS2) primers ITS S2F and ITS4R [35] and universal ribulose-1,5-biphosphate carboxylase/oxygenase large subunit (*rbcL*) primers, *rbcL2* [36] and *rbcLaR* [37]. The ITS2 and *rbcL* barcodes were amplified via a two-step PCR protocol. In the first step, Illumina overhang adapter sequences were added to each primer [38]. Each PCR reaction contained 10 µL of 5X Green GoTaq[®] Reaction Buffer, 0.3 µL of 10 mM dNTPs, 1 µL of each primer, 0.2 µL of GoTaq[®] DNA Polymerase, 35.5 µL of water, and 2 µL of the template DNA with a total volume of 50 µL per reaction. The PCR began with an initial heat activation period of 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. A final extension of 10 min at 72°C was included after the last cycle, and the samples were held at 10°C until further processing. Negative controls consisting of water instead

of template DNA were included in each round of PCR. A 5 μ L sample of each PCR product was electrophoresed in a 2% agarose gel, stained with GelRed™ (Biotium Inc, Fremont, CA), and visualized under UV light to confirm the presence of the appropriately sized amplicons. PCR products were purified using the Promega Wizard SV Gel and PCR Clean-Up System and quantified using a Nanodrop 2000 Spectrophotometer (Thermo Scientific, Waltham, MA). Samples were diluted with sterilized water in 96-well plates then shipped to the Center for Genome Research and Biocomputing at Oregon State University for index PCR and sequencing (standard Illumina MiSeq paired-end 300 bp run).

Bioinformatics

Read quality filtering and denoising. A total of 4,530,819 and 3,579,511 raw, paired-end reads were retrieved across 112 samples for ITS2 and *rbcL* respectively. The open-source Quantitative Insights Into Microbial Ecology (QIIME2 version 2020.11 and 2021.2, [39]) was used as our pipeline environment for sequence processing and analysis. The QIIME2 DADA2 plugin [40] was used to filter read quality and denoise reads (mixture ITS: --p-trunc-len-f 254, --p-trunc-len-r 199, mixture *rbcL*: --p-trunc-len-f 300, --p-trunc-len-r 269, 13-pollen samples ITS: --p-trunc-len-f 254, --p-trunc-len-r 202, mixture *rbcL*: --p-trunc-len-f 300, --p-trunc-len-r 269). Noisy reads and reads with a median sequencing quality score below 30 were omitted. Paired-end reads were joined to create contigs, and the duplicated sequences and sequences with chimera were removed. This resulted in a feature table containing optimized sequences and amplicon sequence variants (ASV), which was used for downstream analyses (e.g., taxonomic classification). After filtering and denoising, 2,637,043 (58%) and 2,639,554 (74%) paired-end reads were retained for ITS2 and *rbcL*, respectively.

Reference database construction. We used a list of plant species known to occur at Starkey, developed over multiple years by botanists, and generated a list of vascular plant species, including cultivated specimens, known to occur in Umatilla County via the Consortium of Pacific Northwest Herbaria (<https://www.pnwherbaria.org>). After removing sedges, rushes, and grasses, the lists were comprised of 492 and 1,435 plant species known to be present at Starkey and in Umatilla County respectively. A master plant list was developed that included the 1,607 unique species occurring at both sampling locations. After removal of subspecies and cultivars, the master list was comprised of 1,277 unique species. Reference sequences of the ITS2 and *rbcL* region for the plants included in this list were then retrieved from the National Center for Biotechnology Information (NCBI) database using a python package: NCBI-Companion (available at https://github.com/lixiaopi1985/NCBI_Companion). The downloaded reference sequences were further cleaned to remove duplicates. The ITS2 database contained 13,306 sequences with a median length of 607 bp and the *rbcL* database contained 10,766 sequences with a median length of 607 bp. Respectively, the ITS2 and *rbcL* databases covered 80.4% and 77.1% of the total species on the master list.

Training classifier and taxonomy classification. The Naïve Bayes algorithm provided by the QIIME2 feature-classifier plugin [41] was used to train a classifier for taxonomic assignment based on the ITS2 and *rbcL* databases. The default settings were applied (i.e., kmer length = 7, confidence threshold = 0.7) with high accuracy and low recall. The classify-sklearn tool of the QIIME2 feature-classifier plugin [41] was used to import the reference sequences and then assign taxonomy to each ASV using the default settings.

Sequence count removal thresholds. We used two types of sequence count removal thresholds to filter the sequencing data: a liberal threshold and a conservative threshold. For the liberal threshold, we removed any taxa whose total number of DNA sequences was less than 0.01% of the total number of DNA sequences for all taxa combined, similar to the

methods used by Pornon et al. [10] and Richardson et al. [19]. We categorize this approach as “liberal” because only plant species that are globally rare in the dataset are removed (not interactions that may be rare in a particular pollen load).

For the conservative threshold, we found the average and the standard deviation of the number of sequencing reads in the negative control samples. We added 1.645 standard deviations to the average and used this number as our sequence count removal threshold, accounting for 95% of possible “background noise” detected in the pollen samples [42]. We set the sequence count to zero for any taxonomic assignment whose read count fell below the threshold, similar to the methods used by Bell et al. [18] and Macgregor et al. [29]. We categorize this approach as “conservative” because taxonomic assignments that are rare in each pollen load are removed.

Comparing proportion of pollen mass and sequencing reads

For each mixture sample, we filtered the sequence reads using both the liberal and conservative threshold. Then, we created stacked bar plots from each data set comparing the proportion of pollen by mass to the proportion of ITS2 and *rbcL* sequencing reads for each plant species within the mixture. Stacked bar plots were created using the ggplot2 package in R [43, 44].

For each of the five plant species included in the mixture samples, we created scatter plots with proportion of pollen by mass on the x-axis and proportion of sequence reads on the y-axis. These scatter plots allowed us to visually compare the relationship between mass of pollen and number of sequence reads to a one-to-one relationship for each plant species, which is what we would expect to see if number of sequence reads were a good proxy for amount of pollen in a sample. Scatter plots were created for both the liberal and conservative threshold data sets in Excel (Version 2211).

Plant-pollinator network comparison

We filtered the ITS2 and *rbcL* metabarcoding data from the bee pollen loads using both liberal and conservative thresholds and created plant-pollinator networks from both data sets using the bipartite package in R [43, 45]. The total number of plant species and average number of plant species per pollen load were compared among networks.

Results

ITS2 metabarcoding

Using plant species assignments obtained from ITS2 metabarcoding, three of the five plant species included in the laboratory-prepared mixtures (*O. acanthium*, *S. oregana*, and *V. villosa*) were correctly identified to the species level. *P. gracilis* was assigned only at the genus level, and *T. montana* was assigned as *Thermopsis rhombifolia*. These results were consistent for all replicates in each of the five mixtures.

Using the liberal threshold, nine additional plant taxa that were not used to create the laboratory-prepared mixtures were detected above the sequence count removal threshold in the single-species samples (S2 Table). Trace amounts of *O. acanthium* and *T. montana* (misidentified as *T. rhombifolia*) were detected in all single-species samples (S2 Table). All “rare” plant species in mixture 5 were detected above the liberal threshold (S3 Table).

Using the conservative threshold, some plant species in the mixtures were not consistently detected above the sequence count removal threshold, resulting in false negatives (S3 Table). *O. acanthium* and *T. montana* (misidentified as *T. rhombifolia*) were always detected above the sequence count removal threshold (S3 Table). *S. oregana*, however, was not detected above the

sequence count removal threshold in mixtures 2–5 (S3 Table). *P. gracilis* and *V. villosa* were not detected above the sequence count removal threshold in mixture 5, and the number of sequencing reads for *P. gracilis* fell below the threshold for one replicate of mixture 3 (S3 Table). Only one additional plant species that was not used to create the laboratory-prepared mixture was detected above the sequence count removal threshold in the *P. gracilis* single-species samples (S2 Table).

***rbcL* metabarcoding**

Only one of the five plant species included in the laboratory-prepared mixtures (*V. villosa*) was correctly identified at the species level using plant species assignments obtained from *rbcL* metabarcoding. *O. acanthium* was assigned as *Hieracium* sp., *S. oregana* was assigned as *Malva neglecta*, and *P. gracilis* and *T. montana* were assigned only at the genus level. These results were consistent across all replicates in each mixture.

Using the liberal threshold, 19 additional plant taxa that were not used to create the laboratory-prepared mixtures were detected above the sequence count removal threshold in the single-species samples (S2 Table). Trace amounts of *Thermopsis* sp. were detected in all single-species samples (S2 Table). All “rare” plant species in mixture 5 were detected above the threshold (S3 Table).

Using the conservative threshold, *S. oregana* (misidentified as *M. neglecta*) was not detected above the sequence count removal threshold in mixtures 2–5, and the number of sequence reads for *P. gracilis* fell below the threshold for replicate 3 of mixture 5 (S3 Table). *O. acanthium* (misidentified as *Hieracium* sp.), *Thermopsis* sp., and *V. villosa* were always detected above the sequence count removal threshold (S3 Table). Four additional plant species that were not used to create the laboratory-prepared mixtures were detected above the sequence count removal threshold in the single-species samples (S2 Table), and trace amounts of *P. gracilis* were detected above the sequence count removal threshold in the *S. oregana* (misidentified as *M. neglecta*) single-species samples (S2 Table).

Comparing proportion of pollen mass and sequencing reads

Although there was a positive relationship between the proportion of pollen by mass and the proportion of sequencing reads for each species across all mixtures, regardless of the type of threshold used or the region examined, the proportion of pollen of some species was consistently over- or underrepresented by sequence reads (Fig 2). *T. montana* was consistently overrepresented by the number of sequencing reads for both regions and *S. oregana* was consistently underrepresented by the number of sequencing reads in mixtures for both regions (Fig 2). The consequences of certain species being over- and underrepresented resulted in a poor correspondence between the proportion of pollen by weight and sequence reads for many of the mixtures (Fig 3), with even the ranked abundance of read proportions for species differing from the rank abundance of weight proportions.

Plant-pollinator network comparison

We used pollen collected from 13 individual bees representing six species to construct plant-pollinator networks (Fig 4 and S5 Table). When using the liberal threshold to filter ITS2 and *rbcL* metabarcoding results, 44 plant taxa were detected and individual pollen loads contained an average of 16.5 taxa (Fig 4a and S6 Table). When using the conservative threshold, 15 plant taxa were detected and individual pollen loads contained an average of 2.7 taxa (Fig 4b and S7 Table).

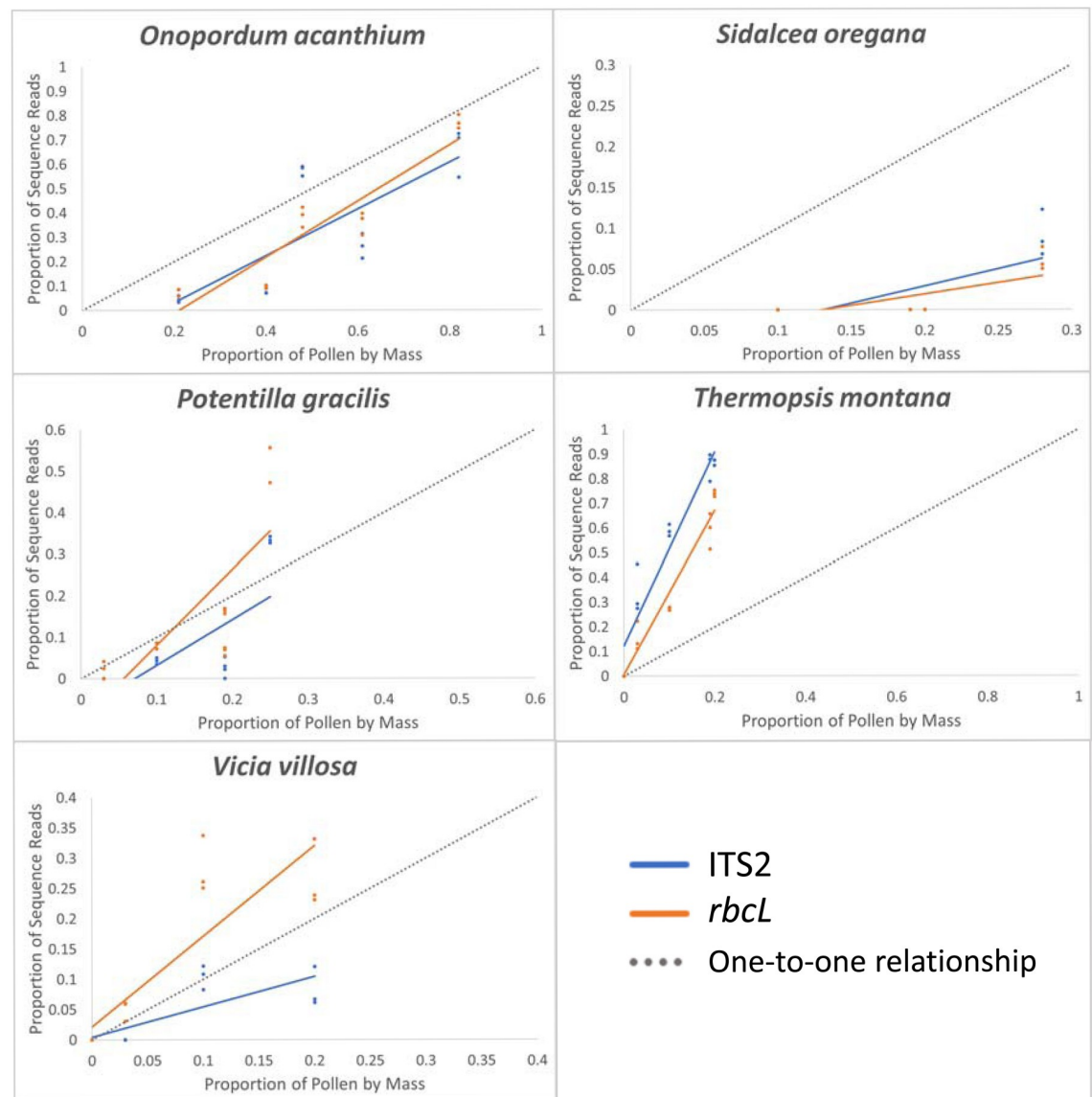


Fig 2. Proportion of pollen by mass in mixtures vs. proportion of sequence reads using the conservative threshold. Gray dotted line represents expected one-to-one relationship. For ITS2, *Potentilla gracilis* = *Potentilla* sp.; *Thermopsis montana* = *Thermopsis rhombifolia*. For *rbcL*, *Onopordum acanthium* = *Hieracium* sp.; *Sidalcea oregana* = *Malva neglecta*; *Potentilla gracilis* = *Potentilla* sp.; *Thermopsis montana* = *Thermopsis* sp. See scatterplots in S1 Fig for liberal threshold results.

<https://doi.org/10.1371/journal.pone.0282715.g002>

Discussion

The results of our study further document that the number of sequence reads cannot be reliably used to estimate the proportion of pollen from different plant species in a pollen mixture. Some species were consistently overrepresented while others were consistently underrepresented (Figs 2 and 3). One overrepresented species was *T. montana*, which had proportions of reads that were on average 4–6 times greater than the proportion of pollen by mass for ITS2 and *rbcL* results regardless of the threshold type. In contrast to *T. montana*, the proportion of sequencing reads produced for *S. oregana* was on average 33–39 times lower than the proportion of pollen by mass when using the liberal threshold for ITS2 and *rbcL*. Moreover, this

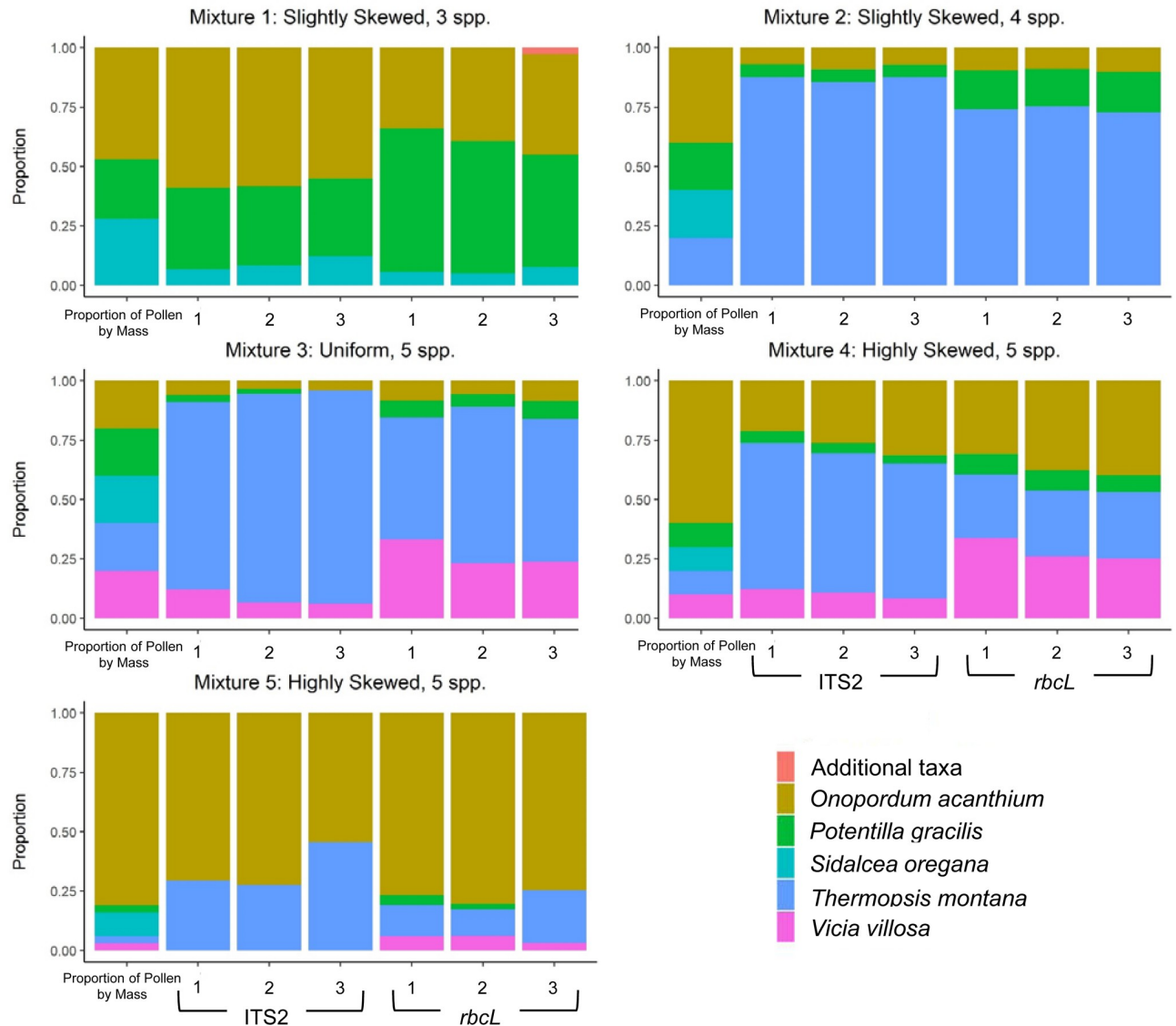


Fig 3. Actual proportion of pollen by mass vs. proportion of sequencing reads using the conservative threshold. The first column of each graph shows the actual proportion of pollen by mass for each plant species in the mixture and the following six columns show the proportion of sequencing reads produced for each species for each of three replicates per mixture obtained with DNA metabarcoding for ITS2 and *rbcL* regions, respectively. For ITS2, *Potentilla gracilis* = *Potentilla* sp.; *Thermopsis montana* = *Thermopsis rhombifolia*. For *rbcL*, *Onopordum acanthium* = *Hieracium* sp.; *Sidalcea oregana* = *Malva neglecta*; *Potentilla gracilis* = *Potentilla* sp.; *Thermopsis montana* = *Thermopsis* sp. See graphs in S2 Fig for liberal threshold results.

<https://doi.org/10.1371/journal.pone.0282715.g003>

species was not even detected above the sequence count removal threshold in 80% of the mixture samples when using the conservative threshold. These deviations from a one-to-one relationship occurred for both barcode regions we examined, even though several studies have suggested that chloroplast barcode markers may be more useful than nuclear ribosomal markers for interpreting metabarcoding results quantitatively [8, 18].

There are several reasons why the number of sequencing reads produced for a species may not be proportional to the mass of pollen including differences in DNA extraction efficiency, amplification bias [26, 27], and differences in marker copy number [23, 46]. Although we used the same DNA extraction kit for all samples, extraction bias can relate to the samples

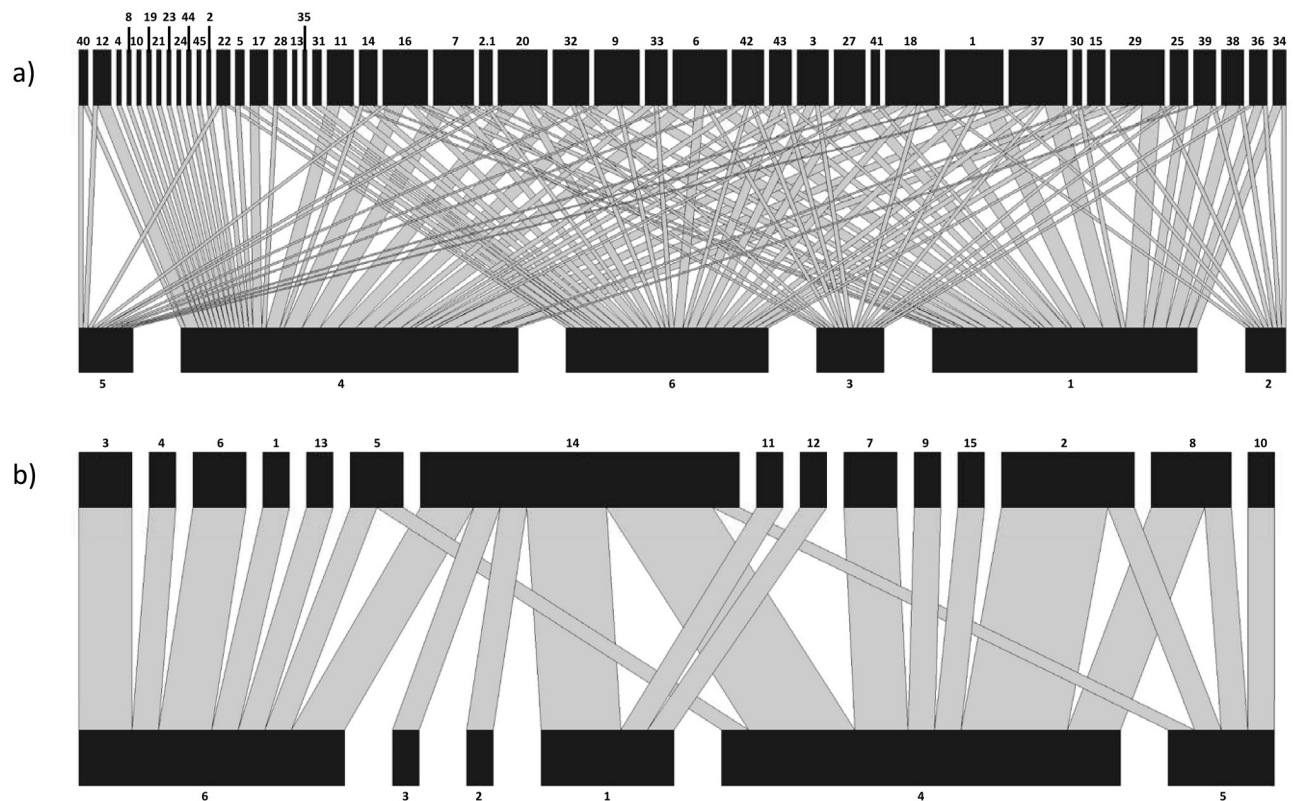


Fig 4. Plant-pollinator networks based on ITS2 and *rbcL* metabarcoding data using liberal and conservative thresholds. (a) Metabarcoding data filtered using the liberal threshold. (b) Metabarcoding data filtered using the conservative threshold. In each network, the top row represents plant taxa and the bottom row represents bee species. The thickness of the lines represents the frequency of the interactions. A complete list of bee species can be found in the table in [S5 Table](#). A complete list of plant species in Fig 4a and 4b can be found in the tables in [S6](#) and [S7](#) Tables, respectively.

<https://doi.org/10.1371/journal.pone.0282715.g004>

themselves. We were able to extract about eight times the amount of DNA, on average, from *T. montana* samples compared to *S. oregana* samples of approximately the same size ([S8 Table](#)), which corresponds to the over-representation of *T. montana* and under-representation of *S. oregana*. Bias can also be related to the success of amplification resulting from the barcode's universal primers. Depending on the combination of barcode marker and species, sequence variation at the priming site can prevent efficient annealing of the primer for that species, resulting in false negatives. Furthermore, amplification efficiency can differ from species to species depending on the composition of their variant of the barcode sequence, resulting in proportions of sequence reads that differ from the actual proportion of the species in the sample [12]. Finally, marker copy number could have contributed to over- and under-representation of *T. montana* and *S. oregana*, respectively. ITS2 is a multi-copy gene region, with high variation in copy number among different plant species [47], and marker copy number bias for chloroplasts is likely to be high when working with pollen because of the various modes of chloroplast inheritance for different plant lineages [48].

Regardless of the mechanism responsible for the over- or under-representation of sequence reads for particular plant species relative to the actual proportion of pollen in a mixed load, our findings show that the issue presents a serious obstacle to using sequence read data quantitatively to assess mixed species samples. When researchers use DNA metabarcoding to determine which plant species are in a pollen sample, the reference library to which they compare the DNA sequences in their samples may contain hundreds or even thousands of plant species,

depending on the study. Attempts to derive quantitative information from pollen samples would require predetermination of which plant species may be over- or underrepresented. Since quantitative biases of certain species are dependent on the other species present in the sample [10, 18], every combination of species would need to be tested prior to analysis, a task that would be practically impossible. Several researchers have proposed the use of correction factors to derive quantitative information from DNA metabarcoding sequence reads [8, 17, 49]. However, the proportions of plant species in a mixed pollen sample are affected by negative correlation bias and thus cannot be examined independently [50]. Although sequencing reads produced by pollen metabarcoding cannot be used to quantify the amount of a plant species in a mixed pollen sample using current methods, sequencing technologies are rapidly advancing and this type of quantification may be possible in the near future using other methods like genome-skimming and whole-genome shotgun sequencing [51, 52].

The results of this study also illustrate the importance of developing reference libraries that correspond as closely as possible to the sampling areas of interest to increase the likelihood of correctly identifying pollen to the species level. Although the five plant species used to create the laboratory-prepared mixtures were consistently detected above the liberal threshold using both ITS2 and *rbcL* barcode markers, several species were misidentified (S9 Table). These misidentifications were likely due to the lack of strong interspecific genetic divergence at the ITS2 and *rbcL* barcoding sites. This explanation is more likely than misidentified reference sequences in the database because genus and family level identifications were consistently correct for ITS2 and *rbcL* DNA sequences, respectively. In our study, plant species used to create the pollen mixtures were collected from two locations in eastern Oregon. One location (Starkey) has a well-documented plant list specific to the area. In contrast, the plant list generated for the other site (Hermiston) was based on a much larger geographic area (i.e., the 1,435 plant species known to occur in Umatilla County). In our study, it is likely that a large number of plants included on the county list do not occur in the particular area that we sampled. Using regional lists instead of site-specific lists to create reference databases increases the likelihood of erroneous identifications when using metabarcoding [13]. Using a narrower list of plant species known to occur in our sampling location rather than a county-wide list of plant species when creating our reference database would have likely led to more accurate plant species assignments.

Another major issue that our study addressed was the degree to which methods used to decrease erroneous conclusions about bee foraging and pollinator network complexity influence rare species detection (Fig 1). When using DNA metabarcoding to identify plant species in a pollen mixture, a certain amount of contamination is inevitable. Contamination can occur at any stage of the DNA metabarcoding process, so precautions are taken in the field and laboratory (e.g., using clean nets to catch each bee, bleaching workspaces and tools). Additional steps are then taken to remove potential sequencing artefacts (e.g., negative controls, sequence count removal thresholds). However, methods that are designed to eliminate sequencing artefacts are also likely to decrease the detection of rare plant species that occur in lower quantities in a pollen load (e.g., from rare flower visits by bees), as these species may fall below the sequence count removal threshold and subsequently be removed from analysis.

Our results suggest that the detection of rare bee-flower interactions depends on the type of threshold used. Our most highly skewed mixture contained 0.075 mg (3%) of pollen from each of *P. gracilis*, *T. montana*, and *V. villosa*, which represent rare bee-flower interactions in a pollen load. When using our liberal threshold, which removes any plant taxa whose total number of sequencing reads falls below 0.01% of total reads, all “rare” plant species were detected above the threshold. However, when using the conservative threshold, *S. oregana* was not detected above the sequence count removal threshold in any of the highly skewed mixture

replicates (mixtures 4 and 5). Furthermore, *P. gracilis* and *V. villosa* were not detected above the sequence count removal threshold in any of the mixture 5 replicates when using the ITS2 barcode marker and the conservative threshold. Therefore, it appears that we can consistently detect “rare” bee-flower visits when using the liberal threshold, but not when using the conservative threshold. However, using a conservative threshold may not always present a serious limitation when studying plant-pollinator interactions because rare bee-flower visits may not be ecologically significant. Certain steps can be taken to reduce stochastic PCR effects such as performing library-prep PCRs in replicate and pooling them [7, 15, 25]. This may have helped us detect more rare species when using the conservative threshold, and we recommend that future studies perform this step, if feasible.

While many precautions can be taken in the laboratory to avoid and control for contamination issues, contamination is more difficult to control in the field. Wind and other insects can move pollen around the environment before sampling occurs. Pollen from other plant species could be present on a flower before a bee visits, resulting in the bee picking up pollen from multiple plant species in one visit. If this were the case, we would expect to see multiple plant species that were not part of our artificial mixtures in our single-species samples. When using the liberal threshold, nine and 19 additional plant species were detected in the single-species samples when using ITS2 and *rbcL* barcode markers, respectively. When using the conservative threshold, only one and four additional plant species were detected in the single-species samples based on ITS2 and *rbcL* sequencing results, respectively. It is important to note that we pooled pollen from flowers from several different individual plants when creating our single-species pollen samples, which likely contributed to the number of additional plant species detected in these samples. However, regardless of the number of additional species detected, the dataset filtered with the liberal threshold consistently included a greater number of additional species than the dataset filtered with the conservative threshold, suggesting that the use of a liberal threshold may overestimate resources use by bees.

We see the consequences of this in the plant-pollinator networks created using each threshold type (Fig 4). The network created using data filtered by the liberal threshold had three times as many plant species than the network created using data filtered by the conservative threshold. In addition, bees, on average, were found to be carrying six times the number of plant species in their pollen loads when using the liberal threshold compared to the conservative one. The ecological interpretations of these results are very different: in Fig 4a, the bee species appear to be exhibiting generalist foraging behavior, whereas in Fig 4b, they appear to be much more selective in their choice of foraging resources. Furthermore, these interpretations lead to different management actions regarding the creation and maintenance of pollinator habitat. Overestimation of network complexity may be a common problem in studies that use DNA metabarcoding to examine bee foraging behavior, and it is important to be aware of the trade-off between using a liberal or conservative sequence count removal threshold with regard to the detection of “rare” plant species and false positives.

Conclusions

In this study, certain plant species were consistently over- and underrepresented by the number of ITS2 and *rbcL* sequence reads produced for the laboratory-created pollen mixtures and rank abundance of sequencing reads did not correspond to rank abundance of pollen mass. This suggests that DNA metabarcoding of the ITS2 and *rbcL* gene region cannot be used to estimate the abundance of a plant species in a mixed pollen sample and rank abundance of sequencing reads should not be used as a proxy for rank abundance of pollen. Furthermore, although one could argue that the plastid barcode marker used in this study appeared to

perform slightly better than the nuclear ribosomal marker, its performance was not improved enough to justify using plastid barcode markers to interpret pollen metabarcoding data quantitatively. Although some misidentification occurred, we were able to consistently detect plant species present in amounts representing rare bee-flower visits when using a liberal threshold, but several additional plant taxa that were not used to create the laboratory-prepared mixtures were also detected in the single-species samples. When using a conservative threshold, some false negatives occurred, but we detected significantly fewer additional plant taxa in the single-species samples. Based on the information presented here, a conservative threshold seems to be most appropriate for the study of most plant-pollinator interactions. When describing these interactions, we suggest that it would generally be better to not capture a few rare flower visits that may not be ecologically relevant than to include potential field contamination and other background noise that could significantly skew results. However, we realize that certain research questions may require the use of a liberal sequence count removal threshold (e.g., detection of a specific compound collected in small amounts). We encourage future researchers using pollen metabarcoding to study plant-pollinator interactions to critically examine their threshold choice to determine whether it is appropriate for the questions being addressed. Finally, we recommend that all pollen metabarcoding studies clearly report the methods used to create and use sequence count removal thresholds so that results can be compared across multiple studies and used to provide accurate and consistent data to inform management decisions regarding pollinator habitat.

Supporting information

S1 Fig. Proportion of pollen by mass in mixtures vs. proportion of sequence reads for using the liberal threshold. Gray dotted line represents expected one-to-one relationship. For ITS2, *Potentilla gracilis* = *Potentilla* sp.; *Thermopsis montana* = *Thermopsis rhombifolia*. For *rbcL*, *Onopordum acanthium* = *Hieracium* sp.; *Sidalcea oregana* = *Malva neglecta*; *Potentilla gracilis* = *Potentilla* sp.; *Thermopsis montana* = *Thermopsis* sp.
(TIF)

S2 Fig. Actual proportion of pollen by mass vs. proportion of sequencing reads using the liberal threshold. The first column of each graph shows the actual proportion of pollen by mass for each plant species in the mixture and the following six columns show the proportion of sequencing reads produced for each species for each of three replicates per mixture obtained with DNA metabarcoding for ITS2 and *rbcL* regions, respectively. For ITS2, *Potentilla gracilis* = *Potentilla* sp.; *Thermopsis montana* = *Thermopsis rhombifolia*. For *rbcL*, *Onopordum acanthium* = *Hieracium* sp.; *Sidalcea oregana* = *Malva neglecta*; *Potentilla gracilis* = *Potentilla* sp.; *Thermopsis montana* = *Thermopsis* sp.
(TIF)

S1 Table. Mass (proportion) of each plant species in the mixture stocks used to create the mixture replicates. Mass is measured in milligrams.
(DOCX)

S2 Table. Number (proportion) of ITS2 and *rbcL* sequencing reads for each plant species in single-species samples (three replicates/sample) using liberal and conservative thresholds. See table in [S4 Table](#) for a list of additional taxa identified in pollen samples. Additional taxa are defined as plant taxa detected in the samples that were not used to create the laboratory-prepared pollen mixtures. See table in [S9 Table](#) for taxonomic assignments using ITS2 and *rbcL* metabarcoding.
(DOCX)

S3 Table. Proportion by mass and number (proportion) of ITS2 and *rbcL* sequencing reads for each plant species in laboratory-prepared mixtures (three replicates/mixture) using liberal and conservative thresholds. Shaded boxes represent instances in which the conservative threshold failed to detect species in mixture, resulting in false negatives. See table in [S4 Table](#) for a list of additional taxa identified in mixture samples. Additional taxa are defined as plant taxa detected in the samples that were not used to create the laboratory-prepared pollen mixtures. See table in [S9 Table](#) for taxonomic assignments using ITS2 and *rbcL* metabarcoding.

(DOCX)

S4 Table. Plant taxa that were not used to create the laboratory-prepared mixtures but were detected in mixtures and single-species samples. Laboratory-prepared mixtures were created from the same pollen stocks used to create the single-species samples. “M” denotes a plant taxon that was detected in a mixture sample, “S” denotes a plant taxon that was detected in a single-species sample, and “MS” denotes a plant taxon that was detected in both a mixture sample and a single-species sample.

(DOCX)

S5 Table. Complete list of bee species included in [Fig 4](#).

(DOCX)

S6 Table. Complete list of plant taxa included in [Fig 4a](#).

(DOCX)

S7 Table. Complete list of plant taxa included in [Fig 4b](#).

(DOCX)

S8 Table. Mass and DNA yield for each pollen sample.

(DOCX)

S9 Table. Plant taxonomic assignments from ITS2 and *rbcL* metabarcoding.

(DOCX)

S10 Table. Plant species and corresponding number of sequencing reads found in each of the negative controls. Extraction, PCR and Plate refers to negative controls that were collected during DNA isolation, PCR, and plating respectively.

(DOCX)

S1 File. Complete lists of plant species detected in each individual bee pollen load. Worksheets of plant taxonomic assignments retained after using a liberal or conservative sequence count removal threshold for each individual bee pollen load.

(XLSX)

Acknowledgments

The authors thank the United States Forest Service for access to study sites and foraging bees; Victoria Skillman for assisting with flower collection and laboratory sample preparation; and Lauren Smith for field assistance.

Author Contributions

Conceptualization: Katherine A. Arstingstall, Sandra J. DeBano, David E. Wooster, Mary M. Rowland, Kenneth Frost.

Data curation: Katherine A. Arstingstall, Xiaoping Li, Skyler Burrows.

Formal analysis: Katherine A. Arstingstall, Xiaoping Li.

Funding acquisition: Katherine A. Arstingstall, Sandra J. DeBano, Kenneth Frost.

Investigation: Katherine A. Arstingstall, Xiaoping Li.

Methodology: Katherine A. Arstingstall, Xiaoping Li, David E. Wooster, Kenneth Frost.

Project administration: Katherine A. Arstingstall, Sandra J. DeBano, Mary M. Rowland, Kenneth Frost.

Resources: Sandra J. DeBano, Xiaoping Li, David E. Wooster, Mary M. Rowland, Kenneth Frost.

Software: Xiaoping Li.

Supervision: Sandra J. DeBano, David E. Wooster, Mary M. Rowland, Kenneth Frost.

Validation: Sandra J. DeBano, David E. Wooster, Mary M. Rowland, Kenneth Frost.

Visualization: Katherine A. Arstingstall.

Writing – original draft: Katherine A. Arstingstall.

Writing – review & editing: Sandra J. DeBano, Xiaoping Li, David E. Wooster, Mary M. Rowland, Skyler Burrows, Kenneth Frost.

References

1. Goulson D, Nicholls E, Botías C, Rotheray EL. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science*. 2015 Mar 27; 347(6229):1255957. <https://doi.org/10.1126/science.1255957> PMID: 25721506
2. Cameron SA, Sadd BM. Global trends in bumble bee health. *Annual Review of Entomology*. 2020; 65(1):209–32. <https://doi.org/10.1146/annurev-ento-011118-111847> PMID: 31610137
3. Durant JL, Otto CRV. Feeling the sting? Addressing land-use changes can mitigate bee declines. *Land Use Policy*. 2019 Sep; 87:104005.
4. Free JB. The flower constancy of bumblebees. *Journal of Animal Ecology*. 1970; 39(2):395–402.
5. Rahl M. Microscopic identification and purity determination of pollen grains. In: Jones MG, Lympany P, editors. *Allergy Methods and Protocols* [Internet]. Totowa, NJ: Humana Press; 2008 [cited 2019 Dec 10]. p. 263–9. (Methods in Molecular Medicine). https://doi.org/10.1007/978-1-59745-366-0_22
6. Cornman RS, Otto CRV, Iwanowicz D, Pettis JS. Taxonomic characterization of honey bee (*Apis mellifera*) pollen foraging based on non-overlapping paired-end sequencing of nuclear ribosomal loci. *Jan E*, editor. *PLoS ONE*. 2015 Dec 23; 10(12):e0145365. <https://doi.org/10.1371/journal.pone.0145365> PMID: 26700168
7. Keller A, Danner N, Grimmer G, Ankenbrand M, von der Ohe K, von der Ohe W, et al. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. *Plant Biology*. 2015; 17(2):558–66. <https://doi.org/10.1111/plb.12251> PMID: 25270225
8. Richardson RT, Lin C-H, Sponsler DB, Quijia JO, Goodell K, Johnson RM. Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. *Applications in Plant Sciences*. 2015; 3(1):1400066.
9. Richardson RT, Lin C-H, Quijia JO, Riusech NS, Goodell K, Johnson RM. Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. *Applications in Plant Sciences*. 2015; 3(11):1500043. <https://doi.org/10.3732/apps.1500043> PMID: 26649264
10. Pornon A, Escaravage N, Burrus M, Holota H, Khimoun A, Mariette J, et al. Using metabarcoding to reveal and quantify plant-pollinator interactions. *Sci Rep*. 2016 Jun; 6(1):27282. <https://doi.org/10.1038/srep27282> PMID: 27255732
11. Smart MD, Cornman RS, Iwanowicz DD, McDermott-Kubeczko M, Pettis JS, Spivak MS, et al. A comparison of honey bee-collected pollen from working agricultural lands using light microscopy and ITS metabarcoding. *Environ Entomol*. 2017 Feb 1; 46(1):38–49. <https://doi.org/10.1093/ee/nvw159> PMID: 28062536

12. Potter C, de Vere N, Jones LE, Ford CR, Hegarty MJ, Hodder KH, et al. Pollen metabarcoding reveals broad and species-specific resource use by urban bees. *PeerJ*. 2019 Feb 19; 7:e5999. <https://doi.org/10.7717/peerj.5999> PMID: 30809427
13. Arstingstall KA, DeBano SJ, Li X, Wooster DE, Rowland MM, Burrows S, et al. Capabilities and limitations of using DNA metabarcoding to study plant–pollinator interactions. *Mol Ecol*. 2021 Oct; 30(20):5266–97. <https://doi.org/10.1111/mec.16112> PMID: 34390062
14. Campbell BC, Al Kouba J, Timbrell V, Noor MJ, Massel K, Gilding EK, et al. Tracking seasonal changes in diversity of pollen allergen exposure: Targeted metabarcoding of a subtropical aerobiome. *Science of The Total Environment*. 2020 Dec 10; 747:141189. <https://doi.org/10.1016/j.scitotenv.2020.141189> PMID: 32799020
15. Leontidou K, Vokou D, Sandionigi A, Bruno A, Lazarina M, De Groeve J, et al. Plant biodiversity assessment through pollen DNA metabarcoding in Natura 2000 habitats (Italian Alps). *Sci Rep*. 2021 Sep 14; 11(1):18226. <https://doi.org/10.1038/s41598-021-97619-3> PMID: 34521917
16. Tremblay ÉD, Duceppe MO, Thurston GB, Gagnon MC, Côté MJ, Bilodeau GJ. High-resolution biomonitoring of plant pathogens and plant species using metabarcoding of pollen pellet contents collected from a honey bee hive. *Environmental DNA*. 2019; 1(2):155–75.
17. Kraaijeveld K, de Weger LA, García MV, Buermans H, Frank J, Hiemstra PS, et al. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Molecular Ecology Resources*. 2015; 15(1):8–16. <https://doi.org/10.1111/1755-0998.12288> PMID: 24893805
18. Bell KL, Burgess KS, Botsch JC, Dobbs EK, Read TD, Brosi BJ. Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. *Molecular Ecology*. 2019; 28(2):431–55. <https://doi.org/10.1111/mec.14840> PMID: 30118180
19. Richardson RT, Curtis HR, Matcham EG, Lin C-H, Suresh S, Sponsler DB, et al. Quantitative multi-locus metabarcoding and waggle dance interpretation reveal honey bee spring foraging patterns in Mid-west agroecosystems. *Molecular Ecology*. 2019; 28(3):686–97. <https://doi.org/10.1111/mec.14975> PMID: 30549365
20. Baksay S, Pornon A, Burrus M, Mariette J, Andalo C, Escaravage N. Experimental quantification of pollen with DNA metabarcoding using ITS1 and trnL. *Scientific Reports*. 2020 Mar 6; 10(1):1–9.
21. Polling M. DNA metabarcoding using nrITS2 provides highly qualitative and quantitative results for airborne pollen monitoring. *Science of The Total Environment*. 2022; 806(1). <https://doi.org/10.1016/j.scitotenv.2021.150468> PMID: 34583071
22. Álvarez I, Wendel JF. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*. 2003 Dec 1; 29(3):417–34. [https://doi.org/10.1016/s1055-7903\(03\)00208-2](https://doi.org/10.1016/s1055-7903(03)00208-2) PMID: 14615184
23. Kembel SW, Wu M, Eisen JA, Green JL. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput Biol* [Internet]. 2012 Oct 25 [cited 2020 Mar 9]; 8(10). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3486904/> PMID: 23133348
24. Brooks JP, Edwards DJ, Harwich MD, Rivera MC, Fettweis JM, Serrano MG, et al. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiology*. 2015 Mar 21; 15(1):66. <https://doi.org/10.1186/s12866-015-0351-6> PMID: 25880246
25. Bell KL, Fowler J, Burgess KS, Dobbs EK, Gruenewald D, Lawley B, et al. Applying pollen DNA metabarcoding to the study of plant–pollinator interactions. *Applications in Plant Sciences*. 2017; 5(6):1600124.
26. Pawluczyk M, Weiss J, Links MG, Egaña Aranguren M, Wilkinson MD, Egea-Cortines M. Quantitative evaluation of bias in PCR amplification and next-generation sequencing derived from metabarcoding samples. *Anal Bioanal Chem*. 2015 Mar; 407(7):1841–8. <https://doi.org/10.1007/s00216-014-8435-y> PMID: 25577362
27. Pompanon F, Deagle BE, Symondson WOC, Brown DS, Jarman SN, Taberlet P. Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology*. 2012; 21(8):1931–50. <https://doi.org/10.1111/j.1365-294X.2011.05403.x> PMID: 22171763
28. Pornon A, Andalo C, Burrus M, Escaravage N. DNA metabarcoding data unveils invisible pollination networks. *Scientific Reports*. 2017 Dec 4; 7(1):1–11.
29. Macgregor CJ, Kitson JJN, Fox R, Hahn C, Lunt DH, Pocock MJO, et al. Construction, validation, and application of nocturnal pollen transport networks in an agro-ecosystem: a comparison using light microscopy and DNA metabarcoding. *Ecological Entomology*. 2019; 44(1):17–29.
30. Janzen DH. Seed and pollen dispersal by animals: convergence in the ecology of contamination and sloppy harvest. *Biol J Linn Soc*. 1983 Aug 1; 20(1):103–13.

31. Whitehead DR. Wind pollination: some ecological and evolutionary perspectives. In: Real L, editor. *Pollination Biology*. Cambridge: Academic Press, Inc.; 1983. p. 97–107.
32. Latta RG, Linhart YB, Fleck D, Elliot M. Direct and indirect estimates of seed versus pollen movement within a population of ponderosa pine. *Evolution*. 1998; 52(1):61–7. <https://doi.org/10.1111/j.1558-5646.1998.tb05138.x> PMID: 28568141
33. Rowland MM, Bryant LD, Johnson BK, Noyes JH, Wisdom MJ, Thomas JW. The Starkey project: history, facilities, and data collection methods for ungulate research. Portland, Oregon, USA: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station; 1997. 62 p. General Technical Report PNW-GTR-396.
34. Kuhlman M, Burrows S. Checklist of bees (Apoidea) from a private conservation property in west-central Montana. *Biodivers Data J*. 2017 Mar 30;(5):e11506. <https://doi.org/10.3897/BDJ.5.e11506> PMID: 28765718
35. Sickel W, Ankenbrand MJ, Grimmer G, Holzschuh A, Härtel S, Lanzen J, et al. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecol*. 2015 Dec; 15(1):20. <https://doi.org/10.1186/s12898-015-0051-y> PMID: 26194794
36. Palmieri L, Bozza E, Giongo L. Soft fruit traceability in food matrices using real-time PCR. *Nutrients*. 2009 Dec; 1(2):316–28. <https://doi.org/10.3390/nu1020316> PMID: 22253987
37. Kress WJ, Erickson DL. A Two-Locus Global DNA Barcode for Land Plants: The coding *rbcl* gene complements the non-coding *trnH-psba* spacer region. *PLOS ONE*. 2007 Jun 6; 2(6):e508. <https://doi.org/10.1371/journal.pone.0000508> PMID: 17551588
38. Klindworth A, Priesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*. 2013 Jan 1; 41(1):e1–e1. <https://doi.org/10.1093/nar/gks808> PMID: 22933715
39. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019 Aug; 37(8):852–7. <https://doi.org/10.1038/s41587-019-0209-9> PMID: 31341288
40. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016 Jul; 13(7):581–3. <https://doi.org/10.1038/nmeth.3869> PMID: 27214047
41. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*. 2018 Dec; 6(1):90. <https://doi.org/10.1186/s40168-018-0470-z> PMID: 29773078
42. Armbruster DA, Pry T. Limit of blank, limit of detection and limit of quantitation. *Clin Biochem Rev*. 2008 Aug; 29(Suppl 1):S49–52. PMID: 18852857
43. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
44. Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
45. Dormann CF, Gruber B, Fründ J. The bipartite package. Version 0.73. R Project for Statistical Computing. 2008.
46. Song J, Shi L, Li D, Sun Y, Niu Y, Chen Z, et al. Extensive Pyrosequencing Reveals Frequent Intra-Genomic Variations of Internal Transcribed Spacer Regions of Nuclear Ribosomal DNA. *PLOS ONE*. 2012 Aug 30; 7(8):e43971. <https://doi.org/10.1371/journal.pone.0043971> PMID: 22952830
47. Marinucci M, Romi R, Mancini P, Luca MD, Severini C. Phylogenetic relationships of seven palearctic members of the *maculipennis* complex inferred from ITS2 sequence analysis. *Insect Molecular Biology*. 1999; 12. <https://doi.org/10.1046/j.1365-2583.1999.00140.x> PMID: 10634971
48. Reboud X, Zeyl C. Organelle inheritance in plants. *Heredity*. 1994 Feb; 72(2):132–40.
49. Thomas AC, Deagle BE, Eveson JP, Harsch CH, Trites AW. Quantitative DNA metabarcoding: improved estimates of species proportional biomass using correction factors derived from control material. *Molecular Ecology Resources*. 2016; 16(3):714–26. <https://doi.org/10.1111/1755-0998.12490> PMID: 26602877
50. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology* [Internet]. 2017 [cited 2023 Jan 16]; 8. Available from: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.02224> PMID: 29187837
51. Lang D. Genome-skimming provides accurate quantification for pollen mixtures. *Molecular Ecology Resources*. 2019; 19(6). <https://doi.org/10.1111/1755-0998.13061> PMID: 31325909
52. Bell KL. Comparing whole-genome shotgun sequencing and DNA metabarcoding approaches for species identification and quantification of pollen species mixtures. *Ecology and Evolution*. 2021; 11(22). <https://doi.org/10.1002/ece3.8281> PMID: 34824813