Quantitative Amplicon Sequencing for Meta-DNA Analysis Reveals Patterns in Pollen Use by Bees

Aubrie James[1], Monica Geber[2], and David Toews[3]

[1]The University of Queensland
[2]Cornell University
[3]Pennsylvania State University

April 28, 2020

Abstract

An underdeveloped but potentially valuable molecular method in ecology is the ability to quantify the frequency with which foraging pollinators carry different plant pollens. Thus far, DNA metabarcoding has only reliably identified the presence/absence of a plant species in a pollen sample, but not its relative abundance in a mixed sample. Here we use a system of four congeneric, co-flowering plants in the genus Clarkia and their bee pollinators to (1) develop a molecular method to quantify different Clarkia pollens found on foraging bees; and (2) determine if bee pollinators carry Clarkia pollens in predictable ways, based on knowledge of their foraging behaviors. We develop a molecular method we call quantitative amplicon sequencing (qAMPseq) which varies cycling number (20, 25, 30, and 35 cycles) in polymerase chain reaction (PCR), individually indexing the same samples in different cycle treatments, and sequencing the resulting amplicons. These values are used to approximate an amplification curve for each Clarkia species in each sample, similar to the approach of quantitative PCR, which can then be used to estimate the relative abundance of the different Clarkia species in the sample. Using this method, we determine that bee visitation behaviors are generally predictive of the pollens that bees carry while foraging. We also show that some bees carry multiple species of Clarkia at the same time, indicating that Clarkia likely compete via interspecific pollen transfer. In addition to adding a 'missing link' between bee visitation behavior and actual pollen transfer, we suggest qAMPseq as another molecular method to add to the developing molecular ecology and pollination biology toolbox.

INTRODUCTION

Discovering how the interactions of plants and pollinators play out is critical to understanding how they mutually rely on each other, and molecular methods are an increasingly common toolkit in this endeavor. In particular, identifying the source of DNA in mixed-DNA samples has become increasingly popular with the development of a range of metabarcoding methods. These approaches rely on single-locus PCR amplicons, leveraging the power of high-throughput DNA sequencing to compare reads to a database of putative DNA sources. The applications of metabarcoding are manifold, including analyses of microbiomes and diets (Sousa et al. 2019), as well as environmental DNA analysis to quantify community composition (Deiner et al. 2017, Sousa et al. 2019). Such molecular approaches have also recently been applied to plant identification from the pollen found on bees to determine which species of flowers they have visited (Wilson et al. 2010; Galimberti et al. 2014; Sickel et al. 2015; Bell et al. 2016, Bell et al. 2017). Historically, pollen has been identified using microscopic morphological differences to distinguish different plant species (Martin and Harvey 2017). However, for some plant taxa, morphological similarity of pollen between distinct species makes this method impossible. In these situations, pollen grains taken from wild bees can be molecularly interrogated to identify the species of plant they are from (Mitchell et al. 2009, Galliot et al. 2017, Lucas et al. 2018).

Molecular methods for identifying pollen origin have been generally restricted to making inferences about the presence or absence of a source plant species from pollen (via metabarcoding techniques; Bell et al. 2017).

This is in part because previous methods have been unable to reliably quantify relative abundance of mixed pollen samples for several reasons, discussed in detail in Bell et al. (2019). One reason is that when using plastid DNA in metabarcoding (e.g. Galimberti et al. 2014), it is unclear how the abundance of chloroplast DNA (cpDNA) is related to overall pollen abundance; if the ratio of cpDNA and pollen grains is not one-to-one, this could bias estimates of relative abundance (Richardson et al. 2015, Bell 2019). Another bias of pollen metabarcoding approaches has to do with how polymerase chain reaction (PCR) amplifies target markers (Bell et al. 2017). In PCR, the final concentration of amplicon DNA after a full PCR protocol is not necessarily directly correlated to input DNA concentration. This is because most PCR protocols will take the amplification process into the "plateau phase", usually after approximately 30 thermal cycles. At the plateau stage, amplicon concentration may be a function of exhausted reaction reagents rather than original concentration of input DNA - the desired information. Because of this, studies using the "plateau abundance" of amplicons could be subject to PCR bias, especially in cases of low abundance (rarity) in the sample.

Though the potential weakness of using PCR to quantify amplicons has been raised by previous authors (e.g. Richardson et al. 2015; Bell et al 2017; Bell et al. 2019), it is still of interest to use molecular methods to estimate the relative abundance of different pollen sources found in bees' scopae or corbiculae (that is, in their pollen-carrying morphological structures). This is in part because understanding resource use among mutualists such as plants and pollinators is fundamental to determining the extent to which they may rely on each other for population persistence (Roulsten and Gooddell 2011), support ecosystem functioning (Lucas et al. 2018), and mediate interspecific competition and coexistence (Johnson 2019). In fact, as Bell et al. (2019) noted, the ability to molecularly determine species abundances in mixed-pollen samples may be 'groundbreaking' for understanding plant-pollinator communities, because successful pollen transport between plants determines plant seed production.

Plant-pollinator interactions can provide a predictive framework for how pollinators introduce reproductive interference and/or fitness benefits to co-flowering plants (James 2020). The effects of pollinator visitation on plant seed set are determined by the extent to which pollinators (1) carry a mix of pollen on their bodies while foraging (constancy) and (2) carry rare versus common species in their pollen balls (preference). During a single pollen collection bout, bees can visit multiple flower species, a behavior known as inconstancy, or visit the same flower species, known as constancy (Kunin and Iwasa 1996). Pollinator inconstancy exposes plants to reproductive interference via heterospecific pollen transfer, which has been shown to drive lower seed production and fitness in plants (Mitchell et al. 2009; Carvalheiro et al. 2014; Arceo-Gómez et al. 2019). Pollinator preference is a measure of flower choice by pollinators. If a strongly competitive plant is preferred by pollinators, pollinator preference might exacerbate competitive exclusion, but if a weakly competitive plant is preferred by pollinators, preference could mitigate competitive exclusion. Though pollinator-mediated plant interactions are most often understood through the lens of pollinator behavior, ample evidence suggests that linking pollinator visitation to plant seed set is problematic: not all plant-pollinator contacts result in pollen transfer (Mayfield 2001; Popic et al. 2013; Ballantyne et al. 2015, 2017; Barrios et al. 2016). Because of this, the relative abundances of pollen on bees could provide valuable information about how pollinators mediate plant interactions.

Here, we develop and use a novel high throughput amplicon sequencing method to quantify the relative abundance of different pollen sources on bee pollinators visiting a group of sympatric winter annual plants in the genus *Clarkia* (Onagraceae). This group of plants – *C. cylindrica* ssp. *clavicarpa* (Jeps.) Lewis & Lewis, *C. speciosa* ssp. *polyantha* Lewis & Lewis, *C. unguiculata*Lindl., *C. xantiana* ssp. *xantiana* A. Gray — are sympatric in the woodland-chaparral areas of the southern foothills of the Sierra Nevada mountain range (from here, we do not use their subspecies epithets). The four species of *Clarkia* rely on a small group of bee pollinators specialized on the genus *Clarkia* rather than any one species (MacSwain et al. 1973; Moeller 2005). Though these*Clarkia* have distinct adult phenotypes, their pollen grains are morphologically indistinguishable. The *Clarkia* also co-occur with each other more often than they occur alone in plant communities in their range of sympatry, and assemblages can contain one to four species of *Clarkia* (Eisen and Geber 2018). Finally, *Clarkia* bloom much later in the growing season than the vast majority of co-

occurring flowering annual plants, and as such are often the only flowering plants where they occur.

A previous study of the bee visitors to one of the four species of interest to this study, *C. xantiana* , showed that though 49 species of bee visit *C. xantiana,* there were only 12 likely "core pollinators" of the species, nine of which carry almost exclusively *Clarkia* pollen (Moeller 2005). Studies including the other three species of *Clarkia* in their range of sympatry with *C. xantiana* have found that there are three consistently common pollinator taxa in multi-species *Clarkia* assemblages (Singh 2013, James 2020). The most common pollinator, *Hesperapis regularis* (Melittidae) has been shown to preferentially visit *C. xantiana* . Preferences of bees in the *Lasioglossum* genus (Halictidae), the second most common pollinator taxon, is unclear: they have been shown to visit all *Clarkia* species at relatively the same rates (Singh 2014) or preferentially visit *C. xantiana* and *C. cylindrica* (James, 2020). The unresolved nature of *Lasioglossum* preferences are at least partially explained by the fact that it is difficult to identify different *Lasioglossum* species when observing them on the wing, as in (James, 2020). Despite preferences, *Hesperapis regularis* and *Lasioglossum* species visit *C. cylindrica, C. unguiculata,* and *C. xantiana* regularly, and are inconstant when foraging in diverse arrays and thus likely to transfer incompatible pollen between plants (James 2020). The final most-common bee pollinator in the system, *Diadasia angusticeps* (Apidae), is behaviorally more specialized on one *Clarkia* species, *C. speciosa* , and rarely visits the other species of *Clarkia* (Singh 2013, James 2020).

Critically, experimental evidence in this system has linked the behavioral inconstancy and preference of pollinators with *Clarkia* seed production (James 2020). *Clarkia speciosa* exhibits low pollen limitation to reproduction, which may be explained by the constancy and preference behavior of *Diadasia angusticeps.* The other three species exhibit higher pollen limitation to reproduction, which may be due to the inconstancy of *Hesperapis regularis* and *Lasioglossum sp* . However, because pollinator visitation does not equate with pollen transfer, it remains unknown if pollinator preference and constancy in the *Clarkia* system in fact determine plant interactions.

The wealth of natural history knowledge and the morphological similarity of *Clarkia* pollens make the *Clarkia* system ideal for developing a method that can both identify and quantify different species in pollen samples. In this paper, we develop a method that we call "quantitative amplicon sequencing" or "qAMPseq" to quantify the relative abundance of *Clarkia* pollen in pollen balls from wild bees. Quantitative amplicon sequencing uses the amplification curve of PCR as a backbone for quantification: it targets single nucleotide polymorphisms private to each species, and then uses PCR to amplify these regions with the goal of post-amplification sequencing, as in metabarcoding. Critically, PCR amplification is stopped before saturation at four different times so one can estimate when each species' amplification curve crosses a critical threshold (as in quantitative PCR or qPCR; Figure 2). The estimate of when the curve crosses the critical threshold point is then used to estimate the relative abundance of each species in each sample. If previously-observed trends in bee behavior match what they carry in their pollen balls (James 2020), then we predict the inconstant bees, *Hesperapis regularis* and *Lasioglossum sp.* , will carry multiple species of pollen at once. In addition, because these particular bee taxa have an established behavioral preference for *C. xantiana* , we expect that they will carry more *C. xantiana* than other *Clarkia* species. We also predict that the pollen on *Diadasia angusticeps* will contain only *C. speciosa* pollen. Finally, because *Hesperapis* and *Lasioglossum sp.* are the most common taxa in this region of *Clarkia* sympatry, we predict an overall pollinator preference for *C. xantiana* pollen.

MATERIALS AND METHODS

Study Species and Field Sampling

The four species of *Clarkia* in this study are sympatric in the Kern River Canyon in Kern County, California. To determine if pollinator foraging behaviors are reflected in the contents of their pollen balls, we collected bees visiting *Clarkia* from May-June of 2014. We sampled bees in 14 *Clarkia* communities throughout the four species' range of sympatry (Figure 1). *Clarkia* communities varied in *Clarkia* species richness, and contained either one, two, or four species of *Clarkia* (Table 1). In each community, we placed four, 20m transects through patches of *Clarkia* . We sampled all transects in all communities between 15 May and 15

June. Each community was sampled twice on different days: once in the morning (between 8AM and 12PM) and once in the afternoon (between 1PM and 3:30PM). Sampling entailed walking along each transect for 20 minutes and catching bees using a sweep net when they landed on *Clarkia*. Bees were sacrificed using ammonium carbonate, and we noted the location, date, and *Clarkia* species bees were visiting when we caught them. We stored, pinned, and identified bee samples to species (or in the absence of species-level resolution, to genus) using Michener et al. (1994). We then scraped the pollen contents off of all collected bees and stored each pollen ball in 90% ethanol in centrifuge tubes at -20°C.

At the end of every netting period, we surveyed *Clarkia* floral abundance to estimate the relative abundance of each species. To do so, we placed $\frac{1}{2}m^2$ quadrats every four meters on either side of transects and counted all open flowers inside the quadrats. Relative abundance of each species was calculated as the proportion of the number of flowers that were open, divided by the total number of flowers we counted at the survey time.

In the summers of 2015 and 2016, we also collected pollen from each species of *Clarkia* for use in testing our methodological design. To do so, we collected mature anthers from all four species of *Clarkia* in various communities throughout their range of overlap. We removed pollen from the anthers and stored them in the same manner that we stored pollen ball samples.

Transcriptome sequencing from greenhouse plants

To identify genetic markers that could distinguish the four *Clarkia* species, we performed transcriptome sequencing from tissue obtained from *Clarkia* plants we grew in the greenhouse (five individuals per species; $n = 20$). We chose to use transcriptome sequencing because it was a reduced-representation genomic approach that produced long, contiguous sequences—as compared to ddRAD sequencing—that was necessary for us to develop subsequent amplicon probes. To grow plants in the greenhouse, we first cold-stratified and germinated seeds of each species of *Clarkia* in February 2016 in Ithaca, NY. Germinated seedlings were transferred into D40L conetainers (Stuewe & Sons, Tangent, 208 Oregon, USA) with a mix of 50% potting soil and 50% perlite. Once in the conetainers, seedlings were bottom-watered and grown in common conditions in a greenhouse for two months. We harvested seedlings for RNA extraction when seedlings had more than four true leaves, but before they had started flowering (March-April 2016).

The leaf tissue of seedlings was harvested and flash frozen in liquid nitrogen. To extract total RNA, we first mechanically homogenized approximately 100mg of leaf tissue with a nitrogen-chilled mortar and pestle. We then mixed the homogenized tissue with 1mL of TRIzol and 200uL chloroform, following the manufacturers guidelines for RNA isolation. We then used 200uL of the isolation to a RNeasy Mini Elute silica column (Qiagen). We added 5 ul of DNAase (NEB) to the final 45 uL of the final elution, and aliquoted 20 uL of NEBNext Oligo $d(t)_{25}$ beads to isolate mRNA from the total RNA pool. We then followed the protocol for NEBNext Ultra Directional RNA Library prep kit (NEB #E7429L). Due to low mRNA yield, we modified the protocol such that PCR enrichment included 30 cycles. We individually indexed the 20 samples and ran these on a single lane of Illumina HighSeq, using single end 100 bp sequencing chemistry.

We combined data from the five *C. speciosa* individuals to generate a reference draft transcriptome assembly using Trinity (Haas et al. 2013). We note here that many of our sequence reads derived from likely chloroplast DNA (cpDNA). Our goal was to obtain DNA markers within the nuclear genome as the presence and amount of choloroplast organelles within each pollen grain is unknown. We note that future studies might instead sample non-leafy tissue where possible, which will reduce the number of choloroplast reads. However, for the present study, we removed these reads by initially aligning the total read pools from all individuals to the chloroplast genome of a related species, *Oenothera picensis* (NCBI accession number KX118607). From the pool of reads that *did not* align to the *O. picensis* cpDNA genome, we aligned these reads from each individual to the draft *C. speciosa* transcriptome from above. We then called SNPs using the GATK pipeline, using the same set of presets as in Toews et al. (2016). We allowed for filtered out SNPs with more than 50% missing data and a minor allele frequency of less than 5%.

Ideally, our goal was to identify a single genomic region that (1) we could PCR amplify and (2) included derived, fixed SNPs for each species. To do this, we estimated Weir and Cockerham per-SNP F_{ST} estimates

from VCFTools (Danecek et al. 2011). We generated F_{ST} estimates for one species compared to the other 15 individuals, and replicated this across all four species. We then determined which transcriptome contig contained multiple SNPs that had $F_{ST} = 1$ for each of the four species. We used BLAST to compare our top contig in our assembly to the nucleotide database at NCBI Genbank. The top hit was a *Clarkia unguiculata* sequence (NCBI accession number EF017402), and our contig aligned to a region that spans the *5.8S*rRNA gene, the *internal transcribed spacer 2* gene, and the *26S* rRNA gene. To amplify this region across additional samples, we used the forward primer sequence [TCGTCGGCAGCGTC]GTGCCTCGGAGATCATCTGT and reverse primer sequence

[GTCTCGTGGGCTCG]GCCGTGAACCATCGAGTCTTT, with the brackets indicating the portion of the sequence (P5 and P7, respectively) that would align to our dual-indexed (i5 and i7) adaptors.

A note on TaqMAN Probe Sensitivity

Our method of quantitative amplicon sequencing (below) applies the general theory of quantitative PCR. Before introducing this method, a natural question regarding our approach is why we did not use fluorescence-based quantitative PCR, such as TaqMan probes (Thermo Fischer Scientific, Waltham MA, USA). We note here that we did develop and test TaqMan probes for the small region that distinguished the four *Clarkia* species within the manufacturers recommended design specifications. However, these probes, with 1-4 species-specific SNPs, were not sensitive to fluoresce exclusively enough in the target species, and therefore did not allow us to distinguish among any of the four *Clarkia* species. Thus, the TaqMan chemistry was not sensitive enough to generate reliable relative abundance information. We are unaware of other published reports discussing this sensitivity, which was also not clear to the manufacturer, and thus raise this point here for researchers interested in applying fluorescent probes to quantify relative abundance using a small number of SNPs.

Quantitative amplicon sequencing—An Overview

Given the lack of sensitivity of the attempted fluorescence-based method, we developed a sequenced-based amplicon assay. We used two methods to quantify relative input DNA from the four *Clarkia* . First, we used a common approach to quantify relative abundance of input DNA, which simply uses the relative read abundance following the full PCR. We refer to the traditional sequencing approach—using the relative read abundance of amplicons at-or-near the PCR plateau phase—as "RRA-plateau" or "RRA". Second, we used an approach that was specifically designed as an attempt to control for some of the biases introduced by PCR. We contrast the RRA method with our method that utilizes a PCR cycle treatment, which we refer to as "quantitative amplicon sequencing" (qAMPseq).

The premise of qAMPseq applies the theory of quantitative PCR (qPCR, A.K.A. real-time PCR) with the ability to individually index, multiplex, and sequence hundreds of metabarcoded samples (Figure 2). Quantitative PCR analysis uses a pre-determined threshold when the PCR reaction is in an exponential phase of amplification, because the PCR cycle where a reaction product moves into the exponential phase is directly related to the starting DNA concentration, unlike the plateau stage (Kubista, 2005). Realtime PCR uses fluorescence (e.g. TaqMan chemistry) quantified throughout thermocycling to determine the 'cycle number' where the product fluorescence is higher than a background level, as the product is in the exponential amplification phase. The estimated number of PCR cycles when the product hits this threshold is known as threshold cycle (Ct). This Ct value can then be compared across samples to compare starting DNA concentrations.

In qAMPseq, we generate the same PCR amplicon in quadruplicate, with the same starting conditions, but across different PCR cycling numbers (e.g. 20, 25, 30, and 35 cycles; Figure 2B). Subsequent cleanup and indexing steps preserve the relative DNA amounts in each of these reactions, which are then individually indexed (i.e. each original sample has four unique indexes, which correspond to the different cycle 'treatments') and then pooled and sequenced with all other samples (Figure 2C). Samples can then be de-multiplexed (Figure 2D) and, within each sample and treatment, reads are assigned to predicted taxonomic units ("OTUs"; in this case, the four *Clarkia* species). The read abundance across each sample and OTU can then be used

to calculate Ct (Figure 2E), and a more robust value the relative contribution of input DNA (Figure 2F).

Pollen DNA extraction and amplicon library preparation

Pollen DNA was extracted from sample pollen balls (2015) and anther pollen (2015 and 2016) using a CTAB-Chloroform DNA preparation protocol (as in Agrawal et al. 2013), and stored at -20°C until amplification and quantification. We first quantified the DNA concentration in each sample using a Qubit fluorometer, and diluted each DNA sample to ˜2 ng/uL. We also created standard dilutions from 1:10 to 1:10000 in triplicate from a single sample of known origin. We assayed 152 unknown origin pollen ball DNA samples split between two sets. Each set included pollen DNA from 76 unknown samples, as well as the same 8 DNA samples of known origin (two from each species), and 12 samples from the standard dilution in triplicate.

Each set of 96 was then transferred to four identical 96-well plates, where we ran a PCR amplification. We conducted 10uL reaction volumes, including: 6.4 uL of ddH_20, 1 uL of MgCl2, 1 uL of dNTPs, 0.2 uL of each forward and reverse primers (above), 0.1 uL (0.25 units) of JumpStart Taq (Sigma-Aldrich), and 1 uL of template (at 2 ng/uL). For each set of four plates, we then used four identical thermal cyclers to run the following protocol simultaneously: 94°C for 3 minutes, and then for plates 1, 2, 3 and 4 we had 20, 25, 30, or 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 minute, respectively. We then used a final extension time of 5 minutes.

This resulted in eight 96-well plates—four for each set—representing the different cycle treatments. We then cleaned up each reaction with 1.8X volume SeraPure beads: 10uL of sample with 18uL of beads, and performed two 70% etOH washes. We eluted in 20 uL of resuspension buffer (Illumina).

We then ran an individual indexing reaction for each sample within each set (i.e. 384 randomly chosen, unique indexes for each set). The 20 uL indexing reaction included 4 uL of ddH_20, 10 uL of HiFi Master Mix (KAPA Biosystems), 1 uL of each the forward and reverse i5 or i7 indexes, and 4 uL of template DNA from the amplification step. This was run with the following thermal cycling conditions: 95°C for 3 minutes, 98°C for 30 seconds, followed by 8 cycles of 98°C for 30 seconds, 63°C for 30 seconds, and 72°C for 30 seconds. We had a final extension time of 3 minutes.

Within each sample set, we pooled 5uL of each indexed sample from across the four-cycle treatments, resulting in one plate for each of the two sample sets. As before, we used a 1.8X SeraPure bead cleanup for the 20uL pooled samples, and completed two 70% etOH washes. We eluted samples into 20uL of resuspension buffer. An equal volume of each sample was then pooled—within each set—into the final library. We sequenced each of the two final libraries separately across two lanes of an Illumina MiSeq, with 2x150 paired end sequencing chemistry.

Bioinformatics and analysis

Demultiplexing resulted in 1,536 individual fastq files (192 samples across four cycle treatments with forward and reverse reads). We used zgrep in bash to identify sequence motifs unique to each of the four species, combining forward and reverse read counts (Supplemental information).

We generated a standard curve by combining results from across the two sets (Figure S2). As discussed, the critical number to determine relative abundance in qPCR is the Ct value. Because qAMPseq data do not directly yield amplicon counts at the end of every cycle, we did not have direct knowledge of the exact shape of the PCR curve – the important step in determining relative abundance. To determine the cycle when samples crossed a Ct value required using a different approach: first, we log-transformed read counts associated with each of the four-cycle points for which we quantified amplicons. Log-transformation of a PCR sinusoidal curve theoretically results in a linear relationship of cycle and amplicon number. We took advantage of this by determining the slope of the amplification line, i.e. $\frac{log(amplicon\ count)}{\text{cycle number}}$. We set our Ct number as log(10,000 reads), and used it in a simple equation to determine the cycle that corresponded to Ct for each amplification curve of each species in each sample. We henceforth call this number the cycle count. If after 35 cycles a species in a sample had fewer than 10,000 reads, it was assigned a cycle count

value of zero. We calculated cycle numbers and performed all of the following analyses in R version 3.5.2 (R Core Team, 2018).

Ecologically speaking, our goals were to quantify the amount of each *Clarkia* species' pollen in each pollen ball to determine if bees (1) were inconstant pollen foragers (2) exhibited preference for certain species of *Clarkia* and (3) used *Clarkia* species pollen in ways that were not apparent from observations of floral visitation. We also wanted to compare results yielded by the new approach, qAMPseq, to the results from relative read abundance from the plateau phase of PCR.

To determine if bees were inconstant while foraging for pollen, we asked which *Clarkia* species were present or absent in pollen balls using both the qAMPseq method as well as the RRA method. Inconstant pollinators will have more than one species in their pollen balls. In qAMPseq, we determined presence/absence of *Clarkia* in our samples by asking simply if the cycle values were nonzero (present) or zero (absent) for each species of *Clarkia* . In contrast, RRA does not yield a cycle count but instead yields relative reads after full amplification (35 cycles). To determine presence/absence of *Clarkia* with RRA, we used three different sample proportion cutoffs to determine the presence of a species in samples: 0%, where presence was defined as any nonzero read count; and 5% and 10 % cutoffs, where presence was defined as anything above 5% or 10%, respectively. We counted the proportion of pollen balls that had more than one species of *Clarkia* using our four different metrics (raw relative read abundance (RRA), and relative read abundance with 5% and 10% sample proportion cutoffs (RRA5 and RRA10)).

To further understand constancy, we compared the number of species in bee pollen balls to the number of flowering species where they were captured. When we sampled bees in *Clarkia* communities, the communities contained one to four species of flowering *Clarkia* . If bees are completely inconstant, then we expect their pollen balls to contain the same number of species as the communities they were captured in. To test this, we tallied the number of bees caught in communities with one, two, three, and four species of flowering, as well as the number of *Clarkia* in each of their pollen balls. We ran a Pearson's Chi-squared test to determine if the proportion of samples containing one to four species of *Clarkia* pollen matched the proportion of bees caught in communities with one to four species of flowering *Clarkia.* If bees are inconstant when pollen foraging, these proportions would be the same, and the test would return a non-significant result.

Preference for different *Clarkia* species was estimated as the difference between the relative amount of a species' pollen in a sample and the relative amount of that species' floral abundance in a surveyed *Clarkia* community where the bee was captured (as in James 2020). This measure of preference can only be calculated for communities with more than one *Clarkia* species, because there is not an available 'choice' to make between plants in single-species communities; as such, we only calculate preference using the pollinator visits/pollen balls from communities with more than one *Clarkia* species. The calculation yields a value between -1 and 1 for each *Clarkia*species in each pollen ball. Negative values indicate avoidance, positive values indicate preference, and values of zero indicate that bees do not preferentially forage for any species.

We calculated preference using values generated by both RRA with a 0% sample proportion cutoff and qAMPseq. To calculate preference using qAMPseq values, we used the cycle count for a given species divided by the sum of cycle counts in a sample to estimate relative abundance of each species in each sample. We used a paired t-test to determine if there was a significant difference in estimates of preference using qAMPseq versus RRA. We then ran an ANOVA, using Tukey's honest significant difference test to determine if pollinator preference for *Clarkia* species were significantly different, and t-tests to determine if pollinator preferences were significantly different from zero.

For a complete picture of pollen use by pollinators, we also incorporated *Clarkia* abundance in communities and in pollen balls into our analyses. First, we compared the average flowering abundance of *Clarkia* species to compare the amount of floral resources the different *Clarkia* provided to pollinators when flowering. To do so, we log-transformed all non-zero values of flowering abundance and ran an ANOVA with Tukey's honest significant difference test to *Clarkia* species. There was one 'zero' value that we retained in the analysis: the four-*Clarkia* community Kingsnake only had three flowering *Clarkia* species during one sampling period, and

as such, the species with no open flowers (*C. xantiana*) was assigned a zero. The second way we incorporated abundance into our analyses was to use the data we generated with the qAMPseq method. For each bee taxon X *Clarkia* species combination, we summed the number of bees carrying pollen from that *Clarkia* species, weighted by the proportion the *Clarkia* species was represented in the pollen ball. This weighted value tells us not only the presence/absence of *Clarkia* pollen on the bee, but the extent to which the bee species used that particular pollen resource.

Finally, we compared pollinator visitation to *Clarkia* and *Clarkia* pollen use by constructing two networks of plants and pollinators: one network with observations of the *Clarkia* species bees were caught on, and the other with the *Clarkia* pollen that we identified in bees' pollen balls. In the case of the visitation network, the dataset consists of the number of times each pollinator was caught visiting each of the *Clarkia* species. The pollen-use dataset consists of proportions of *Clarkia* pollens in each sample, rather than a single plant-pollinator connection or the presence/absence of *Clarkia* species in a pollen ball. To address this and build the dataset for the pollen network, we multiplied the proportion of each *Clarkia* species in each pollen sample by 100, and rounded to the nearest whole number.

We measure and compare networks' network-level specialization, H2', to understand if pollinators use pollen in ways similar to their floral visitation. We use H2' because it is robust to differences in the number of interactions (Blüthgen, Menzel & Blüthgen 2006). Values of network specialization, H2', are between 0 and 1, where higher H2' values indicate that a network is comprised of more specialized relationships between plants and pollinators, and lower values indicate the network has more generalized relationships. Network specialization should be the same between the two networks if pollinators carry *Clarkia* pollen at the same rates that they visit *Clarkia* . Bipartite networks were built and H2' was calculated using the package bipartite (Dorrman et al. 2020).

RESULTS

In total, we used 192 pollen samples, 40 of which were samples of known pollen contents composed of the pollen from field-collected *Clarkia* anthers, and 152 of which were pollen balls of unknown composition harvested from bees in 2014. Sequencing resulted in 45,847,334 reads, 94% of which aligned to one of the four *Clarkia* species reference sequences. All reads in known samples post-amplification were consistent with the known composition of *Clarkia* in the sample (Figure S1), barring one sample with a small number of reads. We attempted to analyze the contents of all 152 pollen balls, but two contained pollen in such low amounts they were excluded.

Constancy

Of the 150 pollen balls we analyzed, 21 out of 150 bees (14%) were caught on *Clarkia* flowers that were different from the majority of the *Clarkia* pollen found in their pollen balls, indicating at least some amount of pollinator inconstancy (Figure 3). The contents of the pollen balls, however, indicate that pollinators were often constant. This includes a caveat: the final measurement of RRA with no cutoff returned all four species of *Clarkia* in 100% of the samples, indicating the unlikely result that all pollinators are not only inconstant, but visited all four *Clarkia* species - even when collected in communities containing fewer than four *Clarkia* species. Not only is this result unlikely based on the biology of the system, but it is exceedingly rare that any quantitative analysis with relative read abundance would use raw read count in the analysis. As such, the rest of our results will compare RRA with 5% (RRA5) and 10% (RRA10) cutoffs with qAMPseq.

In the large majority of samples, pollen balls contained only one species of *Clarkia* pollen. Estimates of single-species pollen balls varied among methods, with 66% (RRA5), 74% (RRA10), or 76% (qAMPseq) of samples containing only one *Clarkia* species. This indicated a striking level of constancy, emphasized by the fact that 70% of bees we sampled were captured in multi-species *Clarkia* communities. Furthermore, the Pearson's Chi-squared test comparing the number of bees from communities with one to four flowering *Clarkia* species versus the number of pollen samples with one to four flowering *Clarkia* species was significant (X^2 (3)=77.05, p<0.001), confirming that even in diverse *Clarkia* communities, bees were constant (Figure 4, Panel B; Table 2).

Despite overall pollinator constancy, a quarter (qAMPseq, 24%; RRA10, 26%) to a third (RRA5, 33%) of the bees were inconstant, often carrying two species of *Clarkia* pollen. The most common multi-species combination in pollen balls was *C. cylindrica* and *C. unguiculata* (Figure 3). No pollen balls contained all four species of *Clarkia* . Bees carrying *C. speciosa* pollen - the most behaviorally specialized of the *Clarkia* (James 2020) - tended to only carry *C. speciosa* pollen: when present, *C. speciosa* was the only species of pollen in the pollen ball in 12/14 cases. Bees carrying the other three species of *Clarkia* often carried mixtures of the three (Figure 3).

Relative abundance measurements of *Clarkia* in pollen balls were largely the same between qAMPseq, RRA5, and RRA10, and most similar between qAMPseq and RRA10 measurement methods (Table 2 and Figure 4; Panel B).

Preference

The two methods we used to determine pollinator preference, RRA (with 0% cutoff) and qAMPseq, did not differ in their estimates of preference (t(599)=-6.8e-10, p=1). Preferences for *Clarkia* species were significantly different from each other (F(4, 596)=6.210, p<0.001). Bees preferentially carried *C. xantiana* pollen (preference= 0.06±0.04 95% CI) and avoided *C. cylindrica* pollen (preference = -0.09±0.04 95% CI). Bees carried *C. speciosa* and *C. unguiculata* at roughly the same frequency these two species occurred in sampled communities, indicating neither preference nor avoidance (*C. speciosa* t(596)=0.42, p=0.67; *C. unguiculata* t(596)=0.74, p=0.45; Figure 4, Panel A).

Floral abundance of Clarkia *species*

Abundances of the four *Clarkia* species were different (F(3, 47)=6.16; p=0.001). *Clarkia cylindrica* exhibited significantly higher floral abundances than *C. speciosa* and *C. xantiana* . *Clarkia unguiculata* floral abundances were also slightly higher than those of *C. speciosa* and *C. xantiana* , but were not statistically different from any other species. Finally, *C. speciosa* and *C. xantiana* exhibited similar floral abundances (Figure S3).

Pollen use and network comparison

Weighted estimates of pollen use show that pollinator species as a whole carried markedly different proportions of each *Clarkia* pollen. We were able to distinguish two different *Lasioglossum* taxa in our study, and found that the two identifiable taxa of *Lasioglossum* exhibited different rates of carrying each species of *Clarkia* . A putative *Clarkia* specialist, *Lasioglossum pullilabre* , carried all four species but was most associated with *C. cylindrica* , whereas *L. (Dialictus) sp.,* a likely generalist, carried *C. xantiana* at higher rates (Figure 5). Furthermore, the *Clarkia* specialist *Hesperapis regularis* carried the three most pollinator-sharing *Clarkia* species, *C. cylindrica, C. unguiculata* , and *C. xantiana* at almost equivalent rates, while *Diadasia angusticeps* used *C. speciosa* almost to the exclusion of all other *Clarkia* (Figure 5).

Pollinator *Clarkia* visitation and *Clarkia* pollen use networks were similar (Figure 6). Overall network specialization, H2', was 0.38 in the *Clarkia* visitation network, and 0.29 in the *Clarkia* pollen-use network. Differences between networks were most apparent with the less-abundant pollinators: *Apis mellifera* (Apidae), *Bombus sp.* (Apidae), and *Megachile sp.* (Megachilidae). Each of these species were only captured on a subset of *Clarkia* , but carried multiple species of *Clarkia* . The honeybee, *Apis mellifera* , was only caught on *C. xantiana* but carried both *C. cylindrica* and *C. xantiana* ; *Bombus sp.* was only caught on *C. unguiculata* but carried both *C. unguiculata* and *C. xantiana* ; and *Megachile sp.* was caught on *C. cylindrica* and *C. speciosa* , but in fact carried all four *Clarkia* species (Figures 5 and 6).

DISCUSSION

In community ecology, the relationship between plants and pollinators is critical to understanding how pollinators contribute to plant community functioning. Though observations of pollinator visitation can be used to infer how plants interact with each other through shared pollinators, visitation does not necessarily correspond to pollen transfer (Mayfield 2001; Popic et al. 2013; Ballantyne et al. 2015, 2017; Barrios et al. 2016). Metabarcoding has been critical for understanding if and when bees use certain pollen resources in

plant communities (Galliot et al. 2017; Lucas et al. 2018), but to date has only reliably shown if bees use pollen resources from specific species in flowering plant communities, not to what extent (Bell et al. 2017, Bell et al. 2019). This is because relative read abundance from PCR can be unreliable for accurate use in amplicon quantitation in a sample.

Using our new approach, we have shown that shared pollinators among *Clarkia* species have preferences for different species of *Clarkia* , and are also inconstant foragers, often carrying more than one species of pollen in their pollen balls at a time. Though the trends of pollen use were similar to what we had expected from pollinator visitation observations alone, our molecular analysis added nuance to how pollinators used *Clarkia* resources. We have additionally shown that our new method, qAMPseq, estimates the relative amounts of pollen species in a sample.

Pollen and Pollinators

With this study, we have contributed to the growing body of literature that uses molecular methods to understand pollen use by bees (Mitchell et al 2009, Galimberti 2014, Bell et al. 2017, Galliot et al 2017, Lucas et al. 2018, Bell et al. 2019), and in the process provided some evidence for heterospecific pollen transfer in flowering plant communities (Mitchell et al. 2009, Arceo-Gomez et al. 2019a, b). That said, pollen balls in our study were typically comprised of only one species of *Clarkia* , which indicates that heterospecific pollen transfer may not be as common in the *Clarkia* system as in other flowering plant systems. The bees that did carry heterospecific pollen, *Hesperapis regularis* and *Lasioglossum sp.* , were the two bee taxa commonly shared between *C. cylindrica, C. unguiculata* and *C. xantiana* , which suggests that pollinator inconstancy introduces heterospecific pollen between these three species of *Clarkia* (Figures 3 and 4, panel B).

There were two important results revealed by our pollen analysis that would not have been available using metabarcoding or visitation observations alone. First, though *C. xantiana* was preferred by pollinators, *C. cylindrica* was carried with greater total representation in *Hesperapis* and *Lasioglossum(Dialictus) sp.* pollen balls (Figure 5). This result is likely due to the number of *C. cylindrica* in *Clarkia* communities- *C. cylindrica* has the highest average floral abundance of all the species (Figure S3). Therefore, while its pollen was found in high abundance on many bees (as in Figure 5), it is not carried in proportion to its relative abundance in *Clarkia* communities (shown in Figure 4). Consequently, *C. cylindrica* is avoided relative to its abundance, suggesting the pollinator behaviors in this system might limit its competitive dominance. Second, we were able to resolve differences in preference between *Lasioglossum* taxa. Because we collected and sacrificed the bee specimens in this study, we could identify *Lasioglossum* with higher resolution than observing visitation without sampling, and show that the taxa exhibit differences in their relationships to *Clarkia* : the putative specialist on the *Clarkia* genus, *Lasioglossum pullilabre* (Moeller 2004, Eckhart et al. 2006), carries *C. cylindrica* with higher frequency, whereas the likely generalist, *L. (Dialictus) sp.* , carries *C. xantiana* with higher frequency (Figure 5).

With this study, we were also able to better delineate the *Clarkia* use by rare pollinators in our dataset. The rare pollinators, *Apis mellifera* , *Bombus sp.* , and *Megachile sp.* (rare to the dataset, but not rare in the ecosystem; Moeller 2005; Eckhart 2006, Singh 2013) all carried more species of *Clarkia* pollen in their pollen balls than they had been observed visiting. The specialization of the pollinator *Clarkia* visitation network was higher than that of the pollen-use network, indicating that pollinators use more diverse resources than the plants we caught them on. Given that sampling effort is a perennial issue in network analyses, we suggest that pollen networks like this one and others (for example: Alarcón 2010, Galliot et al. 2017, Lucas et al. 2018) are a potential means to understand plant-pollinator relationships when sampling effort is constrained. We say this with caution, however: pollen analysis data should complement, not supplant, well-designed sampling methods. For example, it is highly likely that *Apis mellifera* and *Bombus sp.* use all four *Clarkia* pollens, rather than just two per species (Singh 2013), but our sample size of a few bees per species makes that impossible to say with certainty.

Other patterns in pollen use were similar to what we expected based on previous observations of pollinator behavior. The two *Clarkia* species most often found in multi-species pollen balls, *C. cylindrica* and *C. un-*

guiculata , have been shown to occur together with higher frequency than any other *Clarkia* species pair in this system, and exhibit pollinator-mediated character displacement in floral traits (Eisen and Geber 2018). Given the frequency with which they occur together in pollen balls, it's possible that character displacement in the floral traits of these two species could be driven by the competitive effects of heterospecific pollen transfer. Furthermore,*Diadasia angusticeps* bees carried pollen balls of single-species composition (*C. speciosa*), which corresponds to previous observations that the species is behaviorally more specialized on*C. speciosa* (Singh 2013, James 2020).

Quantitative Amplicon Sequencing

This study represents a new method of meta-DNA relative abundance analysis via sequencing. Given the dramatic decline in sequencing costs as well as the sensitivity of current sequencing methods to detecting SNPs, this method may represent a lower-cost alternative to florescence-based qPCR. We also note that qAMPseq offers a higher-accuracy alternative to relative read abundance that is not subject to the issues that can arise when trying to quantify PCR products from metabarcoding (Bell et al. 2019). That said, the protocol here could be streamlined. One example is that it may be acceptable to perform fewer bead cleaning steps to reduce cost and time at the bench. Our protocol also necessitated four thermocyclers, running simultaneously. However, similar to how gradient thermocyclers vary annealing temperatures across reaction wells, we can envision a modification of thermocycler heating blocks that might allow for variation in the number of reaction cycles, and perhaps allow qAMPseq reactions to be run on the same machine, also lowering costs (e.g. Schicke and Hofmann, 2007).

One important difference between our study and studies that use markers such as ITS2 and rcbL is that we had specific target species we quantified in our samples. As such, we designed primers to amplify regions where we knew there were SNPs that differentiated our target species from each other, rather than relying on variation in ITS2 and rcbL to distinguish *Clarkia* from each other. It is also important to note that, because we were using closely related species within the*Clarkia* genus, our application of the meta-barcoding approach was likely not subject to many of quantitative biases identified by Bell et al. (2019). These biases include copy number variation of the amplified gene, differences in DNA isolation efficacy among samples, and variation in primer amplification efficiency. If others are to use this method with primers that target a broader range of possibly more divergent species, these additional biases need to be carefully considered in experimental design.

Finally, the difference between estimates of relative abundance from qAMPseq versus relative read abundance depended on the tolerance with which we filtered raw RRA values. In our study, the 10% cutoff of RRA best matched the results from qAMPseq. A benefit to using qAMPseq, rather than relative read abundance, is that it did not require an arbitrary cutoff for the proportion of reads at the end of PCR that we needed to filter out. However, qAMPseq still required that we define an arbitrary number of reads as a threshold for amplification; this value is analogous to the critical Ct value in the qPCR method of quantitation. Importantly, it is likely that the similarity in relative abundances between our qAMPseq approach and RRA was due to the fact that our samples were diluted to a similar starting concentration of DNA (2 ng/uL). In studies wishing to use RRA in lieu of qAMPseq or qPCR, it should be noted that RRA may not yield accurate estimates of relative abundance if DNA concentrations are highly variable (Bell et al. 2019). We also highlight that because all of our samples were diluted to the same initial concentration, our analysis does not incorporate information about the size of the sampled pollen ball, so we cannot draw conclusions about the amount of pollen that different pollinators transport in their pollen balls. Irrespective of approach, care must be taken in determining the concentration of pre-amplification samples with any particular primer/target combination, as well the interpretation of the resulting data.

Conclusions

Our novel approach to molecular identification of bee pollen allowed us to determine the relative abundance of different pollens from mixed samples taken from foraging bees, which is critically important for more precise estimates of pollen use networks (Bell et al. 2017). Our method allowed us to understand ecologically

relevant and nuanced characteristics of plant-pollinator interactions, such as how many species of *Clarkia* bees were interacting with in a given foraging bout, and to what extent (Roslin et al. 2019). This, in turn, revealed how *Clarkia* interacted via their pollinators, and provided insights into more general patterns of species interactions. As the theories of plant species interactions continue to incorporate mutualistic, complex community dynamics (Morales-Castilla et al. 2015; Vázquez et al. 2015), we believe these kinds of molecular approaches can expand the broader toolkit of community ecologists and facilitate a more robust understanding of species interactions.

Acknowledgements

The authors acknowledge Steve Bogdanowicz, the Cornell Evolutionary Genetics Core Facility, the Cornell Lab of Ornithology, Bronwyn Butcher, and the Lovette Lab for help in method development and for providing lab space and time. Dave Moeller and Indrani Singh helped with bee identification and provided helpful natural history knowledge, and Amy Hastings helped in DNA extraction methods. We thank Julia Brokaw, who was instrumental in the field collections of bees and pollen, as well as Alyssa Anderson, who aided in data preparation. This work was supported by the National Science Foundation Doctoral Dissertation Improvement Grant program, grant number 1701675 awarded to ARMJ and MAG, and NSF DEB-1754299 awarded to MAG.

REFERENCES

Agrawal, A. A., Johnson, M. T. J., Hastings, A. P.and Maron, J. L. 2013. A field experiment demonstrating plant life-history evolution and its eco-evolutionary feedback to seed predator populations. American Naturalist 181: S35-S45.

Arceo-Gómez, G., Kaczorowski, R. L., Patel, C., and Ashman, T. L. 2019*a* . Interactive effects between donor and recipient species mediate fitness costs of heterospecific pollen receipt in a co-flowering community. Oecologia 189: 1041-1047.

Arceo-Gómez, G., Schroeder, A., Albor, C., Ashman, T. L., Knight, T. M., Bennett, J. M., and Suarez, B. et al. 2019*b* . Global geographic patterns of heterospecific pollen receipt help uncover potential ecological and evolutionary impacts across plant communities worldwide. Scientific Reports 9.

Alarcón, Ruben. 2010. Congruence between Visitation and Pollen-Transport Networks in a California Plant-Pollinator Community. Oikos 119: 35–44.

Ballantyne, G., K. C. R. Baldock, L. Rendell, and P. G. Willmer. 2017. Pollinator importance networks illustrate the crucial value of bees in a highly speciose plant community. Scientific Reports.

Ballantyne, G., Baldock K. C. R., and Willmer, P. G. 2015. Constructing more informative plant-pollinator networks: Visitation and pollen deposition networks in a heathland plant community. Proceedings of the Royal Society B: Biological Sciences.

Barrios, B., Pena, S. R., Salas, A. and Koptur, S. 2016. Butterflies visit more frequently, but bees are better pollinators: The importance of mouthpart dimensions in effective pollen removal and deposition. AoB PLANTS.

Bell, K. L., De Vere, N., Keller, A., Richardson, R.T., Annemarie Gous, Kevin S. Burgess, and Berry J. Brosi. 2016. Pollen DNA Barcoding: Current Applications and Future Prospects. *Genome* , 59:629–40.

Bell, Karen L., Fowler, J., Burgess, K.S., Dobbs, E.K., Gruenewald, D., Lawley, B., Morozumi, C. and Brosi, B.J. 2017. Applying Pollen DNA Metabarcoding to the Study of Plant–Pollinator Interactions. Applications in Plant Sciences 5.

Bell, K. L., Burgess, K.S., Botsch, J.C., Dobbs, E.K, Read, T.D., and Brosi, B.J. 2019. Quantitative and Qualitative Assessment of Pollen DNA Metabarcoding Using Constructed Species Mixtures. Molecular Ecology 28: 431–55.

Cane, J. H., and Sipes, S.S. 2006. Characterizing floral specialization by bees: Analytical methods and a revised lexicon for oligolecty. Plant–pollinator interactions: From specialization to generalization.

Carvalheiro, L. G., Biesmeijer, J. C., Benadi, G., Fründ, J., Stang, M., Bartomeus, I., Kaiser-Bunbury, C. N., et al. 2014. The potential for indirect effects between co-flowering plants via shared pollinators depends on resource abundance, accessibility and relatedness. Ecology Letters 17:1389–1399.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. and McVean, G., 2011. The variant call format and VCFtools. Bioinformatics, 27: 2156-2158.

Danforth, B. N., Minckley, R. L., and Neff, J. L. 2019. The Solitary Bees. The Solitary Bees. Princeton University Press, Princeton, New Jersey.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., et al. 2017. "Environmental DNA Metabarcoding: Transforming How We Survey Animal and Plant Communities." *Molecular Ecology* 26 (21): 5872–95.

Eckhart, V.M., Rushing, N.S., Hart, G.M., and Hansen, J.D. 2006. Frequency-Dependent Pollinator Foraging in Polymorphic *Clarkia xantiana ssp. xantiana* Populations: Implications for Flower Colour Evolution and Pollinator Interactions. Oikos 112: 412–21.

Eisen, K. E., and Geber, M.A. 2018. Ecological sorting and character displacement contribute to the structure of communities of Clarkia species. Journal of Evolutionary Biology.

Mitchell, R. J., Irwin, R. E., Flanagan, R. J., and Karron, J. D. 2009. Ecology and Evolution of Plant-Pollinator Interactions.M Annals of Botany 103:1355-1363.

Galliot, J. N., Brunel, D., Berard, A., Chauveau, A., Blanchetete, A., Lanore, L., and Farruggia, A. 2017. Investigating a flower-insect forager network in a mountain grassland community using pollen DNA barcoding. Journal of Insect Conservation 21:827–837.

Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., Macmanes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C.N., Henschel R., Leduc R.D., Friedman N., Regev A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. 2013. Nature Protocols. 8:1494-512.

James, A.R.M. 2020. Pollinator Sharing and Spatial Partitioning Affect Flowering Plant Coexistence. Cornell University.

James, A.R.M., Geber, M.A, and Toews, D.P.L. 2020. Quantitative amplicon sequencing for meta-DNA analysis reveals patterns in pollen use by bees, Dryad, Dataset, https://doi.org/10.5061/dryad.0p2ngf1x0

Johnson, C. A., and Bronstein, J.L. 2019. Coexistence and competitive exclusion in mutualism. ecology 100:e02708.

Kubista, M., Andrade, J.M., Bengtsson, M., Forootan, A., Jonak, J., Lind, K., Sindelka, R., Sjoback, R., Sjogreen, B., Strombom, L., Stahlberg, A., Zoric, N. 20016. The real-time polymerase chain reaction. Molecular Aspects of Medicine 27: 95-125.

Lucas, A., Bodger, O., Brosi, B. J., Ford, C. R., Forman, D. W., Greig, C., Hegarty, M., et al. 2018. Generalisation and specialisation in hoverfly (Syrphidae) grassland pollen transport networks revealed by DNA metabarcoding. Journal of Animal Ecology 87:1008-1021.

MacSwain, J. W., Raven, P. H., and Thorp, R. W. 1973. Comparative behavior of bees and Onagraceae. IV. Clarkia bees of the Western United States. University of California Publications in Entomology 70:1–80.

Martin, A. C., and Harvey, W. J. 2017. The Global Pollen Project: a new tool for pollen identification and the dissemination of physical reference collections. Methods in Ecology and Evolution 8:892–897.

Mayfield, M. M., Waser, N. M., and Price, M. V. 2001. Exploring the "most effective pollinator principle" with complex flowers: Bumblebees and Ipomopsis aggregata. Annals of Botany 88: 591-596.

Mitchell, R. J., Irwin, R. E., Flanagan, R. J., and Karron, J. D. 2009. Ecology and evolution of plant-pollinator interactions. Annals of Botany 103: 1355-1363.

Michener, C. D., McGinley, R. J., and Danforth, B. N. 1994. The Bee Genera of North and Central America. Edited by Rosemary Sheffield. Washington: Smithsonian Institution Press.

Moeller, D. A. 2005. Pollinator community structure and sources of spatial variation in plant-pollinator interactions in *Clarkia xantiana* ssp. *xantiana* . Oecologia 142: 28-37.

Morales-Castilla, I., Matias, M. G., Gravel, D. and Araujo, M. B. 2015. Inferring biotic interactions from proxies. Trends in Ecology and Evolution 30: 347–356.

Popic, T. J., Wardle, G. M., and Davila, Y. C. 2013. Flower-visitor networks only partially predict the function of pollen transport by bees. Austral Ecology 38: 76-86.

Richardson, D. M., Allsopp, N., D'antonio, C. M., Milton, S. J., and Rejmanek, M. 2000. Plant Invasions — the Role of Mutualisms. Biological Reviews 75: 65–93.

Roslin, T., Traugott, M., Jonsson, M., Stone, G. N., Creer, S., and Symondson, W.O.C. 2019. Introduction: Special Issue on Species Interactions, Ecological Networks and Community Dynamics – Untangling the Entangled Bank Using Molecular Techniques. *Molecular Ecology*28: 157–64.

Sickel, W., Ankenbrand, M. J., Grimmer, G., Holzschuh, A., Hartel, S., Lanzen, J., Keller, A. et al. 2015. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC Ecology 15: 20.

Singh, I. 2013. Master's Thesis. Cornell University.

Schicke, K. and Hofmann, C. 2007. Thermocycler with a temperature control block driven in cycle. US Patent number US 2007/0212774 A1.

Sousa, L.L., Silva, S. M., and Xavier, R. 2019. DNA Metabarcoding in Diet Studies: Unveiling Ecological Aspects in Aquatic and Terrestrial Ecosystems." *Environmental DNA* 1: 199–214.

Vazquez, D. P., Ramos-Jiliberto, R., Urbani, P., and Valdovinos, F. S. 2015. A conceptual framework for studying the strength of plant – animal mutualistic interactions. Ecology Letters 18: 385-400.

Data Accessibility

Pollen sample composition data will be archived in the Dryad digital repository.

Author Contributions

ARMJ, MAG, and DPLT designed the study. ARMJ and DPLT performed research, analyzed data, and wrote the paper. All authors contributed to editing and revising the manuscript.

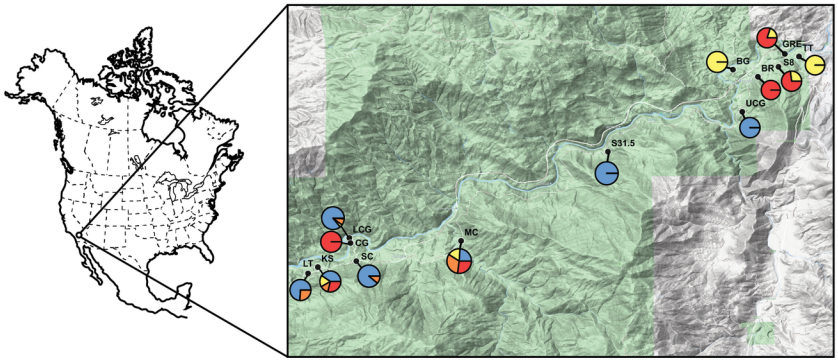

Figure 1. Plant and pollinator sampling locations in the Kern River Canyon in Kern County, California, USA. Pie charts show the *Clarkia* species community composition in the first round of sampling. Colors: red, *C. xantiana* ; blue, *C. cylindrica* ; orange, *C. unguiculata* ; and yellow, *C. speciosa* (see Table 1 for details).

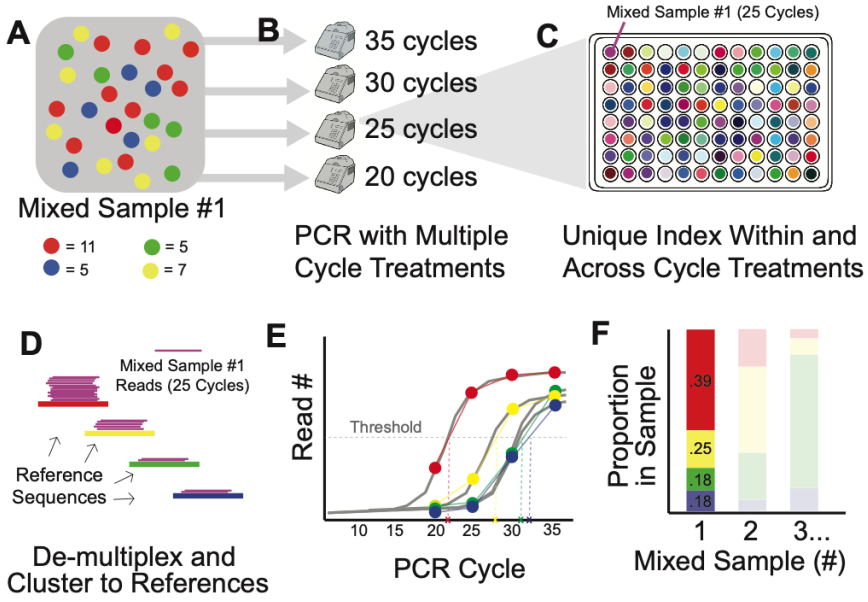

Figure 2. Quantitative amplicon sequencing schematic. **A** A mixed-composition sample. Each circle represents a single pollen grain from four different species, in this example, indicated by the four colors. The numbers of each pollen grain in the mixed sample are shown below (i.e. mixed sample #1 had 11 grains from red, 7 grains from yellow, etc.). **B** This mixed sample is amplified via PCR across four different thermal cyclers, each with different cycle number "treatments". **C** Each of these samples are then uniquely indexed to keep track of sample identify and PCR cycle treatment. All samples are pooled, post-PCR, and run on a single Illumina MiSeq lane, where they are sequenced. **D** After sequencing, samples are de-multiplexed. **E** Read abundance for each species in each sample is used to calculate the Ct value from a simplified read abundance 'curve'. **F** These Ct values are used to calculate the proportion of each species in each sample

(i.e. mixed sample #1 had more red pollen grains, a lower Ct, and subsequently a higher relative proportion in the sample).

Hosted file

`image3.emf` available at https://authorea.com/users/312481/articles/443084-quantitative-amplicon-sequencing-for-meta-dna-analysis-reveals-patterns-in-pollen-use-by-bees

Figure 3. Proportion plot for each sample of pollen harvested by bees, represented by bars. Small circles above each bar indicate the species of *Clarkia* the bee was caught on, while larger pie charts indicate the *Clarkia* species composition of the community each bee came from. White bars indicate that the amount of pollen found on the bee was too low for compositional analysis.

Hosted file

`image4.emf` available at https://authorea.com/users/312481/articles/443084-quantitative-amplicon-sequencing-for-meta-dna-analysis-reveals-patterns-in-pollen-use-by-bees

Figure 4. **A** Preference for different *Clarkia* pollen based on samples taken from foraging bees. Preference was calculated as the difference between frequency of a *Clarkia* species blooming in the community and frequency it was represented in bees' pollen balls. Mean preference is plotted with a 95% confidence interval, and was estimated using quantitative amplicon sequencing. Estimates of preference did not significantly differ between quantitation methods (not shown). **B** Bar plots showing pollinator constancy. The far-left bar shows what we would expect if bees are inconstant: for example, the 34 bees caught in communities with four co-flowering *Clarkia* species (black section of the far-left bar) should have four *Clarkia* species in their pollen balls. We estimated the number of *Clarkia* species found on bees in four ways: quantitative amplicon sequencing (qAMPseq), and relative read abundance (RRA) at different sample proportion cutoffs. Bees carried fewer species in their pollen balls than were flowering in their corresponding *Clarkia* communities. The far-right bar is relative read abundance with no sample proportion cutoff; with this quantitation method, all samples contained all four species.

Hosted file

`image5.emf` available at https://authorea.com/users/312481/articles/443084-quantitative-amplicon-sequencing-for-meta-dna-analysis-reveals-patterns-in-pollen-use-by-bees

Figure 5. Relative abundances of *Clarkia* pollen on pollinators. *Clarkia* present in a pollen ball were weighted by their relative abundance in the sample and then summed to yield a value for that species' relative pollen contents for each pollinator. Larger bars indicate higher pollinator counts, where *Lasioglossum pullilabre* was the most frequently caught pollinator, followed by *Hesperapis regularis, Lasioglossum (Dialictus) sp.,* and *Diadasia angusticeps.*

Hosted file

`image6.emf` available at https://authorea.com/users/312481/articles/443084-quantitative-amplicon-sequencing-for-meta-dna-analysis-reveals-patterns-in-pollen-use-by-bees

Figure 6. Bipartite networks of plants and pollinators from surveying done in 2014. The network on the left shows the relationships of pollinator taxa (bottom row) and the plant species they were captured on (top row; C, *Clarkia cylindrica;* S, *C. speciosa;* U, *C. unguiculata;* X, *C. xantiana*). The network on the right shows the relationships of pollinator taxa and the plant species in their pollen balls as identified and quantified using the qAMPseq method. The size of the purple bars in the top and bottom rows are proportional to the number of bees/plants in the data, and the size of the blue, connecting bars are proportional to the number of times a plant/pollinator combination occurred.

Table 1. Sites sampled in the study and the sampling schedule.

Species composition	Site name	Dates sampled (2014)	Degree Decimal Latitude / Longitude	Degree Decimal Latitude / Longitude
C. cylindrica C. unguiculata	Little Tree (LT)	15 May 19 May	15 May 19 May	35.527197 -118.66324
	Summer Camp (SC)	17 May 21 May	17 May 21 May	35.531156 -118.64774
	Site 31.5 (S31.5)	24 May	24 May	35.566167 -118.56695
C. speciosa C. xantiana	Green Rock East (GRE)	23 May 29 May	23 May 29 May	35.597436 -118.51016
	Site 8 (S8)	24 May 30 May	24 May 30 May	35.593325 -118.51219
	Black Gulch (BG)	29 May 30 May	29 May 30 May	35.592425 -118.52681
C. cylindrica C. unguiculata C. speciosa C. xantiana	Lower China Gardens (LCG)	16 May 26 May	16 May 26 May	35.538564 -118.64998
	Kingsnake (KS)	16 May 26 May	16 May 26 May	35.529239 -118.6601
	Mill Creek (MC)	17 May 23 May	17 May 23 May	35.537633 -118.61411
C. xantiana	Borel (BR)	28 May 30 May	28 May 30 May	35.590131 -118.5188
	Upper China Gardens (UCG)	18 May 31 May	18 May 31 May	35.578939 -118.52383
C. cylindrica	Coyote Gulch (CG)	24 May 25 May	24 May 25 May	35.536933 -118.64963
C. speciosa	Tip Top (TT)	29 May 31 May	29 May 31 May	35.596622 -118.50561

Table 2. The number of pollen samples collected in communities, and the number of flowering species in the pollen balls estimated with different methods – quantitative amplicon sequencing (qAMPseq) and relative read abundance (RRA) at different sample proportion cutoffs.

	Number of flowering species in community	Number of flowering spe
	1	**2**
Number of pollen samples collected	44	63
qAMPseq	114	31
RRA	0	0
RRA, 5%	100	42
RRA, 10%	111	35