

## **Application of ITS2 Metabarcoding to Determine the Provenance of Pollen Collected by Honey Bees in an Agroecosystem**

Authors: Richardson, Rodney T., Lin, Chia-Hua, Sponsler, Douglas B., Quijia, Juan O., Goodell, Karen, et al.

Source: Applications in Plant Sciences, 3(1)

Published By: Botanical Society of America

URL: <https://doi.org/10.3732/apps.1400066>

---

BioOne Complete ([complete.BioOne.org](https://complete.BioOne.org)) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at [www.bioone.org/terms-of-use](https://www.bioone.org/terms-of-use).

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

---

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

## APPLICATION OF ITS2 METABARCODING TO DETERMINE THE PROVENANCE OF POLLEN COLLECTED BY HONEY BEES IN AN AGROECOSYSTEM<sup>1</sup>

RODNEY T. RICHARDSON<sup>2</sup>, CHIA-HUA LIN<sup>2,4</sup>, DOUGLAS B. SPONSLER<sup>2</sup>, JUAN O. QUIJIA<sup>2</sup>,  
KAREN GOODELL<sup>3</sup>, AND REED M. JOHNSON<sup>2</sup>

<sup>2</sup>Department of Entomology, The Ohio State University–Ohio Agricultural Research and Development Center, 1680 Madison Ave., Wooster, Ohio 44691 USA; and <sup>3</sup>Department of Evolution, Ecology and Organismal Biology, The Ohio State University, 1179 University Dr., Newark, Ohio 43023 USA

- *Premise of the study:* Melissopalynology, the identification of bee-collected pollen, provides insight into the flowers exploited by foraging bees. Information provided by melissopalynology could guide floral enrichment efforts aimed at supporting pollinators, but it has rarely been used because traditional methods of pollen identification are laborious and require expert knowledge. We approach melissopalynology in a novel way, employing a molecular method to study the pollen foraging of honey bees (*Apis mellifera*) in a landscape dominated by field crops, and compare these results to those obtained by microscopic melissopalynology.
- *Methods:* Pollen was collected from honey bee colonies in Madison County, Ohio, USA, during a two-week period in mid-spring and identified using microscopic methods and ITS2 metabarcoding.
- *Results:* Metabarcoding identified 19 plant families and exhibited sensitivity for identifying the taxa present in large and diverse pollen samples relative to microscopy, which identified eight families. The bulk of pollen collected by honey bees was from trees (Sapindaceae, Oleaceae, and Rosaceae), although dandelion (*Taraxacum officinale*) and mustard (Brassicaceae) pollen were also abundant.
- *Discussion:* For quantitative analysis of pollen, using both metabarcoding and microscopic identification is superior to either individual method. For qualitative analysis, ITS2 metabarcoding is superior, providing heightened sensitivity and genus-level resolution.

**Key words:** *Apis mellifera*; bee foraging landscape; *Fraxinus*; palynology; pollen analysis; pollinator ecology.

Melissopalynology, the identification of pollen collected by bees, has proven to be an indispensable tool in fields such as pollination biology (Kearns and Inouye, 1993; Cusser and Goodell, 2013), pollinator foraging behavior (Louveaux, 1959; Wilson et al., 2010; Baum et al., 2011), sourcing and authentication of apicultural products (Louveaux et al., 1978; Jones and Bryant, 1992; Dimou et al., 2007), and honey bee (*Apis mellifera* L.) nutritional biology (Severson and Parry, 1981; Forcone et al., 2011; Girard et al., 2012). The lattermost has taken on particular urgency in recent years due to alarming and largely unexplained regional declines in honey bee populations (vanEngelsdorp et al., 2009). Pollen is the main source of amino acids, lipids,

sterols, vitamins, and minerals in the honey bee diet (Winston, 1987), and malnutrition, caused by the displacement of natural floral communities by agricultural development, may contribute to recent patterns of honey bee decline (Naug, 2009; vanEngelsdorp and Meixner, 2010; Huang, 2012). Melissopalynology allows for direct observation of the honey bee pollen diet, providing a unique means for identifying nutritional deficits and informing potential efforts to improve foraging habitat.

Traditional melissopalynology involves the careful preparation of pollen samples for microscopic inspection followed by the identification of individual pollen grains by comparison to a similarly prepared reference collection of pollen from local taxa (Erdtman, 1943; Kearns and Inouye, 1993). When pollen collected by honey bees is used, the steps above are normally preceded by the laborious sorting of bulk pollen samples into groups of pellets having similar color and texture (Dimou and Thrasyvoulou, 2007). To the extent that this sorting succeeds in discriminating true taxonomic groups, it provides a preliminary estimate of sample diversity and streamlines the process of microscopic examination for large volumes of pollen. These methods can provide reliable data when properly executed, but they suffer from being (1) highly dependent on human expertise and vulnerable to human error, (2) limited in taxonomic precision, as many taxa are only identifiable to family, and (3) prohibitively time-consuming for large-scale studies. A reliable and efficient

<sup>1</sup> Manuscript received 25 July 2014; revision accepted 4 December 2014.

The authors thank J. Wenger, E. E. Wilson, C. S. Sidhu, J. Wallace, B. Klips, A. D. Wolfe, and anonymous reviewers for discussion; N. Douridas for site access; G. Cobb for help with sample processing; M. E. Hernandez-Gonzalez and the Ohio Agricultural Research and Development Center Molecular and Cellular Imaging Center staff for technical support; and the Ohio Supercomputing Center for computing time. This study was funded by the Pollinator Partnership's Corn Dust Research Consortium grant to R.M.J. and an Ohio State University–Newark Scholarly Activity Grant to K.G.

<sup>4</sup> Author for correspondence: lin.724@osu.edu

doi:10.3732/apps.1400066

alternative would be of immense value for many melissopalynological applications, especially those that require the analysis of large and taxonomically diverse pollen samples, such as are commonly collected by honey bees (Wilson et al., 2010).

We present a novel approach to melissopalynology using DNA metabarcoding. We used previously validated primers specific for the second internal transcribed spacer (ITS2) in the plant ribosomal sequence (Chen et al., 2010) and performed amplicon sequencing on DNA isolated from four samples of pollen collected by honey bees over a two-week period in spring 2013. We chose the nuclear ITS2 region because prior investigation of ITS2 as a barcoding locus supports its suitability for differentiating plant taxa at the genus and, in some cases, the species level (Chen et al., 2010; Han et al., 2013; Pang et al., 2013; Tripathi et al., 2013). Wilson et al. (2010) successfully used the internal transcribed spacer (ITS), as well as other ribosomal sequences, as a genetic barcode to identify monospecific pollen collected by a specialist bee. Our primary goal was to obtain taxonomic identification at the genus level, which made the ITS2 region suitable for this barcoding project.

Here, we apply our method to determine the pollen taxa collected by honey bee colonies located in an agroecosystem devoted to corn and soybean cultivation, a context in which nutritional stress due to agricultural development would be plausible. We then compare the results of our metabarcoding approach to those obtained by traditional microscopic analysis of the same pollen samples, and discuss the relevance of our findings to both melissopalynological methodology and honey bee foraging in agricultural landscapes.

## MATERIALS AND METHODS

**Sample collection**—Samples from this study were obtained from an apiary managed by The Ohio State University located in Madison County, Ohio (39.96095, -83.43215), between April 23 and May 6, 2013. The apiary is located in an agricultural landscape dominated by corn and soybean fields (>60% land cover), with a small amount of forest, residential area, and uncultivated fields. Four healthy, actively foraging colonies in 8-frame Langstroth hives were fitted with Sundance I bottom-mounted pollen traps (Ross Rounds, Albany, New York, USA) to collect pollen samples. Pollen traps were emptied and turned on and off on a twice weekly cycle, alternating between colonies, so that pollen was always sampled from two colonies while the remaining colonies were allowed pollen for sustenance. Samples were pooled across colonies for each collection date for all analyses.

**Pollen identification by microscopy**—We sorted pollen pellets from 10% (15–68 g) of the total sample for each collection date into distinct color categories and weighed each category. A 10% subsample from each color category, or up to 10 pellets if a category contained fewer than 100 pellets, was then blended in a few drops of water to achieve a spreadable consistency, and four drops of the homogenized suspension were mounted separately in basic fuchsin jelly (Kearns and Inouye, 1993) on glass slides for microscopic examination at 400×–1000× magnification.

Pollen was identified to the lowest possible taxonomic level by comparison with reference pollens collected from fresh flowers found in the study area during April–May. Reference pollens were permanently mounted in basic fuchsin jelly on glass slides. Representative pollen grains on the reference slides were photographed at 400×–1000× magnification. Reference slides and digital images of the pollens were stored at the Ohio State University Bee Laboratory (1680 Madison Avenue, Wooster, Ohio, USA). Published pollen guides and image databases were used to identify pollens that did not match any of the reference collections. Reference pollen taxa and

published references used for pollen identification in this study are listed in Appendix S1.

The sum of weight for color categories belonging to the same taxon was recorded. If a color category contained pollens of multiple taxa, we examined 1000 pollen grains on slides prepared for that color category and estimated the proportional abundance by counting pollen grains for each taxon. We chose to exclude any plant taxa constituting less than 1% of the pollen sample. We then measured the grain size for each taxon and estimated the proportional abundance using the volumetric method described in O'Rourke and Buchmann (1991). The weight of pollen from each taxon was summed to calculate the final proportion in the total sample for each collection date.

Pollens from plants in the family Rosaceae often share similar pellet colors and morphological characteristics, making them difficult to distinguish at the generic level by visual sorting and microscopy (Moore et al., 1991). Genera within Brassicaceae were also difficult to identify using light microscopy. Improving the accuracy of pollen identification to the genus level in these families would require scanning electron microscopy (SEM) and was beyond the scope of this study. Therefore, the abundance of Rosaceae and Brassicaceae estimated by microscopy is only presented at the family level.

**Pollen identification by ITS2 metabarcoding**—For metabarcoding analysis, 5% (8–34 g) of the original pooled sample was mixed in a solution of 75% ethanol in deionized water at roughly a 4:1 ratio of solvent volume to sample volume. The volume of the pollen sample was estimated using a 50-mL graduated cylinder. The pollen and solvent were then mixed in a glass beaker and stirred on a magnetic stir plate for 25 min at room temperature. Additional solvent was added as necessary to ensure optimal sample mixing. Buchner funnel vacuum filtration was performed, and the resulting sample was transferred to a weigh boat to air dry in a flow hood.

We used a bead-beater pulverization method adapted from Simel et al. (1997) to free the genomic DNA from within the protective exine coat of the nongerminated pollen. Fifty milligrams of dry, homogenized pollen from each sample was placed in a 2.0-mL microcentrifuge tube with 600 µL of QIAGEN DNeasy Plant Mini Kit lysis buffer (QIAGEN, Venlo, The Netherlands). Zirconium/silica beads (0.5 mm diameter) were added until the total contents of each tube reached 1.5 mL. The pollen was then pulverized in a bead-beater (Mini-BeadBeater-1; BioSpec Products, Bartlesville, Oklahoma, USA) for 2 min. After pulverization, 300 µL of deionized water was transferred to each tube and mixed with the contents. A 300-µL portion of the resulting lysate was then transferred from each tube to a sterile 1.5-mL microcentrifuge tube. Pollen DNA was then isolated using the QIAGEN DNeasy Plant MiniKit.

After DNA isolation, the ITS2 region was amplified via PCR. We used two primers (forward: 5'-ATGCGATACTTGGTGTGAAT-3'; reverse: 5'-GAC-GCTTCTCCAGACTACAAT-3') designed and validated by Chen et al. (2010), but we used alternative PCR conditions to accommodate our use of the Phusion High-Fidelity PCR Kit (New England Biolabs, Ipswich, Massachusetts, USA). Parameters of the thermocycler (Mastercycler ep gradient; Eppendorf AG, Hamburg, Germany) were set as follows: initial denaturation at 98°C for 30 s, 30 cycles of 98°C denaturation for 10 s, annealing at 59°C for 30 s, and extension at 72°C for 30 s, and a final extension step at 72°C for 10 min. These PCR conditions were calculated with the use of the NEB Tm Calculator (New England Biolabs; <http://tmcalculator.neb.com/>).

Following PCR, ITS2 amplicons were purified using the PureLink PCR Purification Kit (Life Technologies, Carlsbad, California, USA). At this point, 500 ng of purified PCR product from each respective sample was indexed independently using the NEBNext Ultra DNA Library Prep Kit for Illumina and NEBNext Multiplex Oligos for Illumina (New England Biolabs). After multiplexing, samples were subjected to a final Agencourt AMPure XP purification step before being pooled (Beckman Coulter catalog no. A63880 [Brea, California, USA]). A final nine-cycle library amplification PCR step was performed and samples were then analyzed on a Qubit 2.0 Fluorometer (Life Technologies), as well as an Agilent 2100 Bioanalyzer (DNA 1000 kit; Agilent Technologies, Santa Clara, California, USA) to ensure sample quality before sequencing. Paired-end sequencing was performed with the Illumina MiSeq platform using the TruSeq LT assay (Illumina Inc., San Diego, California, USA). All sequence information has been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (accession code SRP044703).

To analyze the sequencing data for our samples, we adapted previously described metabarcoding analysis pipelines (Huson et al., 2011; Mitra et al., 2011). Illumina MiSeq reads were demultiplexed using CASAVA (version 1.9; Illumina Inc.) and preprocessed using the artifacts filtering tool from the

FASTX-Toolkit (version 0.0.13; [http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). Paired-end read merging was then accomplished using PEAR (version 0.9.1), a software package that calculates highly accurate paired-end mergers of forward and reverse reads (Zhang et al., 2014). PEAR was used with default parameter settings with the exception of the Phred scale quality score trimming threshold, which was set to 20. The resulting FASTQ files were dereplicated and converted to FASTA format using the FASTX-Toolkit Collapser tool. After generation of the FASTA files, the resulting sequences were compared against a reference library of 2628 plant ribosomal sequences downloaded from GenBank on 11 March 2014. This library represents approximately half of the 4918 plant species potentially present in Ohio and surrounding states (USDA PLANTS database; <http://plants.usda.gov/>). This library is available in FASTA format from the Dryad Digital Repository (<http://dx.doi.org/10.5061/dryad.64b5p>; Richardson et al., 2015). The BLASTN algorithm (blast-2.2.17) (Altschul et al., 1997) was used to align reads with the GenBank-derived sequence library using the following settings: *E*-value cutoff 1e-125, number of alignments 1, output format 0, number of descriptions 1, percent identity threshold 95%. Alignment outputs were then annotated in MEGAN (version 5.1.5) (Huson et al., 2011) using the lowest common ancestor (LCA) algorithm with the following parameter settings: min support 1, min support percent 0.0 (off), min score 50.0, max expected 1E-125, top percent 100.0, min complexity 0.00.

**Analysis of results**—We used two statistical tests to assess whether the metabarcoding approach could produce reliable quantitative measures of relative abundance for the plant taxa in our samples. The Spearman's  $\rho$  statistic, a rank-based measurement, was employed to test the association between the number of paired-end reads aligned and the amount of pollen in grams per plant taxon identified by microscopy. Additionally, average *R* coefficients, analogous to those described by Bryant and Jones (2001) in relation to honey authentication, characterizing the ratio of the number of aligned paired-end reads relative to the abundance of pollen determined by microscopy for taxa found in at least three of the four samples, were calculated using the equation below. This coefficient can provide information as to whether certain taxa are over-represented or under-represented in the metabarcoding results relative to the microscopic results, as well as the variance in this over- and under-representation across samples. With further development, the *R* coefficient may allow the quantitative

estimation of the taxa present within pollen samples by enabling the calculation of taxon-specific correction coefficients.

$$R \text{ coefficient} = \text{percent mapped reads} / \text{percent weight by microscopy}$$

## RESULTS

**Pollen identification by microscopy**—During the sampling period (April 23–May 6, 2013), a large proportion of bee-collected pollen originated from mass-flowering woody species, ranging from 57% to 95% of the total sample weight per collection date (Fig. 1; identified taxa are listed in Appendix S2). Pollens of trees in the genera *Acer* L. and *Fraxinus* L., both anemophilous, were the most abundant pollens on the three earliest sampling dates. On May 6, more than 90% of the total sample originated from entomophilous woody species, corresponding with the full bloom of rosaceous trees (e.g., *Malus* Mill., *Prunus* L., *Crataegus* L., and *Amelanchier* Medik.) and honeysuckles (*Lonicera morrowii* A. Gray and *L. maackii* (Rupr.) Maxim.) that were common in the surrounding landscape in early May. Pollen from herbaceous weeds, predominantly dandelion (*Taraxacum officinale* F. H. Wigg.) followed by crucifers (Brassicaceae), comprised 43% of the total sample on April 23 and declined progressively to 5% on May 6, when bees were collecting pollen from rosaceous trees and honeysuckles predominantly.

**Pollen identification by ITS2 metabarcoding**—After Illumina MiSeq sequencing and paired-end merging of the forward and reverse reads, we obtained between 593,478 and 1,062,208

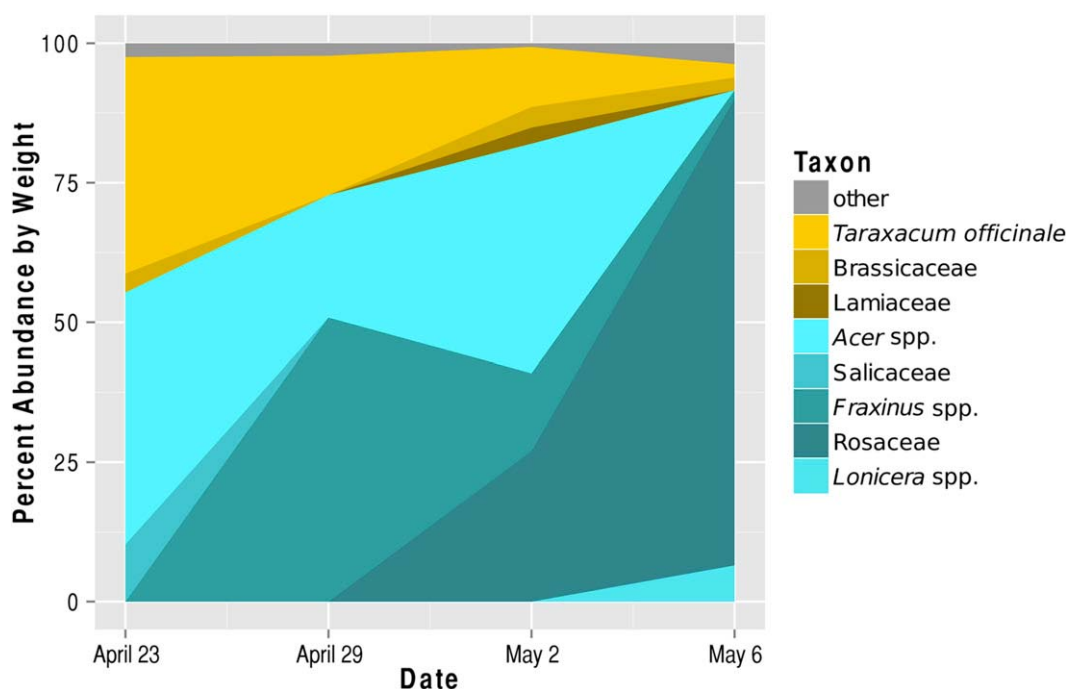


Fig. 1. Pollen origins identified by microscopy during the sampling period. Shades of yellow represent pollen of herbaceous plants whereas shades of turquoise represent pollen of woody plants.



paired-end reads from the samples. The mean paired-end read lengths for all sequencing runs were between 461 bp and 469 bp. Upon annotating the BLAST output files with MEGAN 5, between 71,125 and 340,606 reads were taxonomically assigned. In total, we detected 42 distinct plant genera across the four samples. However, upon cross-validating the list of genera with the known floral phenology calendar for Ohio (Ohio State University, Growing Degree Days and Phenology for Ohio; <http://www.oardc.ohio-state.edu/gdd/>) and herbarium records (Ohio State University Herbarium; <https://herbarium.osu.edu/>), we found alignments to two plant genera (*Lactuca* L. and *Glycine* Willd.) that were not likely to be flowering during our sampling period. For a complete list of the plant genera detected as well as their respective paired-end alignment coverage, see Appendix S3.

**Comparison of metabarcoding and microscopic techniques—**  
To compare the results of ITS2 barcoding with those of traditional microscopy, we calculated the percentage of mapped reads per plant taxon in comparison to the percentage by weight of plant taxa estimated using microscopic methods (Table 1). Using the barcoding approach, we were able to detect the majority of the plant taxa observed using the microscopic approach. Pollen from Salicaceae, Lamiaceae, and *Lonicera* were detected by microscopy but not by metabarcoding analysis.  
To test the ability of the metabarcoding method to accurately determine the relative abundance of different pollen types, we compared paired-end alignment coverage with grams of pollen collected as determined by color sorting and microscopic analysis (Spearman’s correlation test;  $P > 0.05$  for all samples; Table 2). We found no significant associations between the relative abundances of pollen types identified using microscopic-based and metabarcoding-based methods. Furthermore, upon calculating  $R$  coefficients across all taxa, specific taxa were overrepresented

TABLE 1. Comparison of percent weight by microscopy to percent paired-end read alignments per taxon for each sampling date.

Date	Taxon	Percent weight by microscopy (%)	Percent reads by sequencing (%)
April 23	<i>Taraxacum officinale</i>	38.81	12.77
	F. H. Wigg.		
	Brassicaceae Burnett	3.39	83.90
	<i>Acer</i> L.	45.08	2.86
	Salicaceae Mirb.	10.26	0
	Other	2.45	0.47
April 29	<i>T. officinale</i>	24.97	16.89
	<i>Acer</i>	21.96	1.25
	<i>Fraxinus</i> L.	50.84	0.18
	Other	2.23	81.68
May 2	<i>T. officinale</i>	10.75	21.70
	Brassicaceae	3.70	71.21
	Lamiaceae Martinov	2.89	0
	<i>Acer</i>	41.09	6.61
	<i>Fraxinus</i>	13.94	0.01
	Rosaceae Juss.	26.98	0.04
May 6	Other	0.65	0.43
	<i>T. officinale</i>	2.44	7.69
	Brassicaceae	2.26	91.22
	<i>Fraxinus</i>	1.54	0.01
	Rosaceae	83.54	0.01
	<i>Lonicera</i>	6.51	0
	Other	3.72	1.07

TABLE 2. Spearman’s rank-based correlation test between the number of paired-end alignments and the number of grams per plant taxon.

Sample	Spearman’s $\rho$ score	$df$	$P$ value
April 23	−0.4	14	0.75
April 29	0.5	6	1
May 2	0.05	53	0.91
May 6	−0.2	24	0.75

Note:  $df$  = degrees of freedom.

or underrepresented consistently in the metabarcoding results relative to the microscopic results, although the variation in the degree of misrepresentation was large, with relative standard deviations being between 39% and 84% RSD (Table 3). These results indicate that ITS2 sequencing data alone are not sufficient for quantitative measurement of relative abundance of pollens within bee-collected samples.

DISCUSSION

Metabarcoding proved useful for identifying the diversity of pollen collected by honey bees. However, metabarcoding also produced a relatively longer list of plant genera than the microscopic method. This may reflect the genuinely large diversity of plant taxa visited by many thousands of foraging honey bees over a large foraging range. Some genera identified through metabarcoding appear implausible given known floral phenology. The identification of these plant genera may be explained by (1) spurious false-positive BLAST alignments; (2) bees regurgitating honey stomach contents, which could contain stored honey or particles from bee bread collected and stored in the colony for months (Vasquez and Olofsson, 2009); or (3) contact between pollen foragers and stored bee bread within the hive prior to foraging. Additionally, in the case of spurious false-positive alignments, it is important to note that the GenBank-derived BLAST library used in this study lacked ITS2 reference sequences from approximately half of the 4918 plant species potentially present in Ohio and surrounding states (USDA PLANTS database; <http://plants.usda.gov/>). In the absence of a true reference sequence, the BLASTN algorithm could produce a misalignment, assuming quality control cutoffs were satisfied. Lastly, some unexpected grasses identified through metabarcoding, such as *Poa* L., *Anthoxanthum* L., and *Triticum* L. (family Poaceae) may reflect nonforaging contact with windborne pollen, which are usually in very low abundance when present in the bee-collected sample.  
The metabarcoding approach does not provide quantitative estimates of the relative abundance of pollen in polyfloral samples. The nature of ribosomal and plant genetics likely skew quantitative calculations in a species-specific or even intraspecies-specific manner. The number of ribosomal DNA cassette copies

TABLE 3. The average  $R$  coefficients and standard deviations per taxon across all samples.

Taxon	Average $\pm$ SD for ITS2 metabarcoding
Brassicaceae	28.119 $\pm$ 10.954
<i>Acer</i>	0.094 $\pm$ 0.058
<i>Fraxinus</i>	0.003 $\pm$ 0.002
<i>Taraxacum</i>	1.544 $\pm$ 1.296

varies widely within and between species (Long and Dawid, 1980) and could influence the number of reads generated for each taxon. Furthermore, variation in DNA extraction efficiency, primer annealing, and genome copy number between plant species may also affect the number of taxon-specific reads.

Metabarcoding failed to detect pollen from plants in the genus *Lonicera* and families Lamiaceae and Salicaceae, despite the fact that pollens from these taxa were identified using microscopy. This may be explained in several ways, such as sequence divergence in the PCR priming site or current sequencing length limitations for the Illumina MiSeq platform used (~300 bp). Given that the length of the ITS2 region ranges from 100 bp to 700 bp in plants (Yao et al., 2010), the failure of metabarcoding to detect these taxa suggests that improvements to this method are necessary to expand the scope of detection. This may be accomplished using different loci, multiple loci, or through the development of ITS2 primers specific to problematic plant clades. Recent work by Galimberti et al. (2014) applied *rbcL* and *trnH-psbA* amplicon cloning in conjunction with capillary sequencing to characterize the taxonomic composition of bee-collected pollen samples from Italian Alpine habitats. With the use of these loci, the investigators were able to detect *Lonicera* and Lamiaceae; however, given the absence of morphological validation of the amplicon clone-based method, it is difficult to infer whether the primers used for these loci were relatively more or less expansive than the primers used in this study.

Both methods of pollen identification were in agreement with the conclusion that honey bees collect pollen from a wide variety of flowers in the agricultural landscape and the taxa visited varied over the weeks of our study. What was apparent from our microscopic results, but less clear through metabarcoding, was a dramatic shift in the assemblage of bee-collected pollen (Fig. 1). Early samples (April 23 and 29) were dominated by nonrosaceous woody taxa (*Acer*, *Fraxinus*) and dandelion (*T. officinale*). The later samples (May 2 and 6), however, saw a sudden increase in the prevalence of rosaceous woody taxa and a corresponding decline in *Acer*, *Fraxinus*, and *T. officinale*. This shift in principal pollen sources represents not only a phenological-taxonomic transition but also a change in spatial foraging patterns between different floral habitats. *Acer* and *Fraxinus* were restricted to small tracts of forest and scattered residential areas within the field-crop matrix, while *T. officinale* occurred abundantly in residential areas, along roadsides, and in agricultural fields that had not been recently tilled or treated with herbicide. The rosaceous woody taxa that became important in the latter half of our samples consisted of a combination of cultivated trees (e.g., *Malus*) from residential properties and wild taxa such as *Crataegus* and *Prunus* that occurred mainly in forest edge and understory habitat.

Our data underscore the dependence of honey bees upon landscape features that are systematically marginalized by agricultural intensification, namely field edges, forest, and residential habitats. Furthermore, the prevalence of herbaceous weeds, such as *T. officinale*, is highly subject to weed control practices. Thus, intensive field crop agroecosystems are not necessarily nutritional deserts for honey bees in the springtime, but nutrition could become limiting if aggressive land conversion and weed control practices progress unchecked.

In addition to using this methodological pipeline for apicultural applications, our approach may provide a useful tool for many other areas of scientific research. Honey bees collect pollen

from a diverse array of plant taxa, possess a near-global geographic distribution, and a single colony can forage over an area of more than 100 km<sup>2</sup> (Beekman and Ratnieks, 2000; Moritz et al., 2005). As such, honey bees may be seen as a large-scale sampling tool. This trait of honey bees has already been explored for monitoring industrial pollution, airborne bacteria and viruses, and even explosives (Bromenshenk et al., 1985; Lighthart et al., 2000, 2005). Furthermore, given a reliable, high-throughput metabarcoding approach to melissopalynology, honey bees could be employed to sample a regional flora rapidly, inexpensively, and with an intensity unapproachable by human investigators. Moreover, honey bees can easily sample environments that are not amenable to conventional floral sampling, such as dense forests or urban centers. This unique utility of honey bees could be broadly applicable in areas of research such as biogeography, biodiversity assessment, population genetics, floral phenology tracking, and invasive species monitoring.

In this study, we applied high-throughput ITS2 metabarcoding in conjunction with traditional microscopic melissopalynology to determine the pollen foraging habits of honey bees in an agroecosystem dominated by field crops. We demonstrate that the metabarcoding approach exhibits strengths in terms of ease of implementation and its ability to detect pollens that are rare or difficult to identify by microscopy. This approach also exhibits weakness due to the nonquantitative nature of the method. As such, we suggest this method is useful for qualitative applications such as large-scale melissopalynology-based floral surveying and pollinator habitat preservation. Melissopalynology has previously been difficult to apply in such applications due to the time, expense, and expertise required for microscopic pollen identification. The development and further refinement of molecular melissopalynological techniques may be a suitable research avenue for overcoming these difficulties.

## LITERATURE CITED

- ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFER, J. ZHANG, Z. ZHANG, W. MILLER, AND D. J. LIPMAN. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25: 3389–3402.
- BAUM, K. A., W. L. RUBINK, R. N. COULSON, AND V. M. BRYANT JR. 2011. Diurnal patterns of pollen collection by feral honey bee colonies in southern Texas, USA. *Palynology* 35: 85–93.
- BEEKMAN, M., AND F. L. W. RATNIEKS. 2000. Long-range foraging by the honey-bee, *Apis mellifera* L. *Functional Ecology* 14: 490–496.
- BROMENSHENK, J. J., S. R. CARLSON, J. C. SIMPSON, AND J. M. THOMAS. 1985. Pollution monitoring of Puget Sound with honey bees. *Science* 227: 632–634.
- BRYANT, V. M. JR., AND G. D. JONES. 2001. The r-values of honey: Pollen coefficients. *Palynology* 25: 11–28.
- CHEN, S., H. YAO, J. HAN, C. LIU, J. SONG, L. SHI, AND Y. ZHU. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* 5: e8613.
- CUSSEY, S., AND K. GOODELL. 2013. Diversity and distribution of floral resources influence the restoration of plant–pollinator networks on a reclaimed strip mine. *Restoration Ecology* 21: 713–721.
- DIMOU, M. G., AND A. THRASYVOULOU. 2007. A comparison of three methods for assessing the relative abundance of pollen resources collected by honey bee colonies. *Journal of Apicultural Research and Bee World* 46: 144–148.
- DIMOU, M., G. GORAS, AND A. THRASYVOULOU. 2007. Pollen analysis as a means to determine the geographical origin of royal jelly. *Grana* 46: 118–122.
- ERDTMAN, G. 1943. An introduction to pollen analysis. *Chronica Botanica* Company, Waltham, Massachusetts, USA.

- FORCONE, A., P. V. ALOISI, S. RUPPEL, AND M. MUÑOZ. 2011. Botanical composition and protein content of pollen collected by *Apis mellifera* L. in the north-west of Santa Cruz (Argentinean Patagonia). *Grana* 90: 30–39.
- GALIMBERTI, A., F. DE MATTIA, I. BRUNI, D. SCACCABAROZZI, A. SANDIONIGI, M. BARBUTO, M. CASIRAGHI, AND M. LABRA. 2014. A DNA barcoding approach to characterize pollen collected by honeybees. *PLoS ONE* 9: e109363.
- GIRARD, M., M. CHAGNON, AND V. FOURNIER. 2012. Pollen diversity collected by honey bees in the vicinity of *Vaccinium* spp. crops and its importance for colony development. *Botany* 90: 545–555.
- HAN, J., Y. ZHU, X. CHEN, B. LIAO, H. YAO, J. SONG, S. CHEN, AND F. MENG. 2013. The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS. *Biomed Research International* 2013: 741476.
- HUANG, Z. 2012. Pollen nutrition affects honey bee stress resistance. *Terrestrial Arthropod Reviews* 5: 175–189.
- HUSON, D. H., S. MITRA, H.-J. RUSCHWEYH, N. WEBER, AND S. C. SCHUSTER. 2011. Integrative analysis of environmental sequences using MEGAN 4. *Genome Research* 21: 1552–1560.
- JONES, G. D., AND V. M. BRYANT JR. 1992. Melissopalynology in the United States: A review and critique. *Palynology* 16: 63–71.
- KEARNS, C. A., AND D. W. INOUE. 1993. Techniques for pollination biologists. University Press of Colorado, Boulder, Colorado, USA.
- LIGHTHART, B., K. PRIER, G. M. LOPER, AND J. BROMENSHENK. 2000. Bees scavenge airborne bacteria. *Microbial Ecology* 39: 314–321.
- LIGHTHART, B., K. R. S. PRIER, AND J. J. BROMENSHENK. 2005. Flying honey bees adsorb airborne viruses. *Aerobiologia* 21: 147–149.
- LONG, E. O., AND I. B. DAWID. 1980. Repeated genes in eukaryotes. *Annual Review of Biochemistry* 49: 727–764.
- LOUVEAUX, J. 1959. Recherches sur la récolte du pollen par les abeilles (*Apis mellifica* L.) (Fin). *Annales de L'Abeille* 2: 13–111.
- LOUVEAUX, J., A. MAURIZIO, AND G. VORWOHL. 1978. Methods of melissopalynology. *Bee World* 59: 139–153.
- MITRA, S., M. STAERK, AND D. H. HUSON. 2011. Analysis of 16S rRNA environmental sequences using MEGAN. *BMC Genomics* 12: S17.
- MOORE, P. D., J. A. WEBB, AND M. E. COLLINSON. 1991. Pollen analysis, 2nd ed. Blackwell Scientific Publications, Oxford, United Kingdom.
- MORITZ, R. F. A., S. HÄRTEL, AND P. NEUMANN. 2005. Global invasions of the western honeybee (*Apis mellifera*) and the consequences for biodiversity. *Ecoscience* 12: 289–301.
- NAUG, D. 2009. Nutritional stress due to habitat loss may explain recent honeybee colony collapses. *Biological Conservation* 142: 2369–2372.
- O'ROURKE, M. K., AND S. L. BUCHMANN. 1991. Standardized analytical techniques for bee-collected pollen. *Environmental Entomology* 20: 507–513.
- PANG, X., L. SHI, J. SONG, X. CHEN, AND S. CHEN. 2013. Use of the potential DNA barcode ITS2 to identify herbal materials. *Journal of Natural Medicines* 67: 571–575.
- RICHARDSON, R. T., C.-H. LIN, D. B. SPONSER, J. O. QUIJIA, K. GOODELL, AND R. M. JOHNSON. 2015. Data from: Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.64b5p>.
- SEVERSON, D. W., AND J. E. PARRY. 1981. A chronology of pollen collection by honeybees. *Journal of Apicultural Research* 20: 97–103.
- SIMEL, E. J., L. R. SAIDAK, AND G. A. TUSKAN. 1997. Method of extracting genomic DNA from non-germinated gymnosperm and angiosperm pollen. *BioTechniques* 22: 390–392, 394.
- TRIPATHI, A. M., A. TYAGI, A. KUMAR, A. SINGH, S. SINGH, L. B. CHAUDHARY, AND S. ROY. 2013. The internal transcribed spacer (ITS) region and *trnH-psbA* are suitable candidate loci for DNA barcoding of tropical tree species of India. *PLoS ONE* 8: e57934.
- VANENGELSDORP, D., J. D. EVANS, C. SAEGERMAN, C. MULLIN, E. HAUBRUGE, B. K. NGUYEN, AND M. FRAZIER. 2009. Colony collapse disorder: A descriptive study. *PLoS ONE* 4: e6481.
- VANENGELSDORP, D., AND M. D. MEIXNER. 2010. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *Journal of Invertebrate Pathology* 103: S80–S95.
- VASQUEZ, A., AND T. C. OLOFSSON. 2009. The lactic acid bacteria involved in the production of bee pollen and bee bread. *Journal of Apicultural Research* 48: 189–195.
- WILSON, E. E., C. S. SIDHU, K. E. LEVAN, AND D. A. HOLWAY. 2010. Pollen foraging behaviour of solitary Hawaiian bees revealed through molecular pollen analysis. *Molecular Ecology* 19: 4823–4829.
- WINSTON, M. L. 1987. The biology of the honey bee. Harvard University Press, Cambridge, Massachusetts, USA.
- YAO, H., J. SONG, C. LIU, K. LUO, J. HAN, Y. LI, X. PANG, ET AL. 2010. Use of ITS2 region as the universal barcode for plants and animals. *PLoS ONE* 5: e13102.
- ZHANG, J., K. KOBERT, T. FLOURI, AND A. STAMATAKIS. 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics (Oxford, England)* 30: 614–620.