



**HAL**  
open science

## Honey bee-collected pollen richness and protein content across an agricultural land-use gradient

Michael P. Simanonok, Clint R. V. Otto, Deborah D. Iwanowicz, R. Scott  
Cornman

► **To cite this version:**

Michael P. Simanonok, Clint R. V. Otto, Deborah D. Iwanowicz, R. Scott Cornman. Honey bee-collected pollen richness and protein content across an agricultural land-use gradient. *Apidologie*, 2021, 52 (6), pp.1291-1304. 10.1007/s13592-021-00902-5 . hal-03912837

**HAL Id: hal-03912837**

**<https://hal.science/hal-03912837v1>**

Submitted on 26 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Honey bee-collected pollen richness and protein content across an agricultural land-use gradient

Michael P. SIMANONOK<sup>1</sup> , Clint R. V. OTTO<sup>1</sup>, Deborah D. IWANOWICZ<sup>2</sup>, and R. Scott CORNMAN<sup>3</sup>

<sup>1</sup> Northern Prairie Wildlife Research Center, U.S. Geological Survey, 8711 37th St. SE, Jamestown, ND 58401, US

<sup>2</sup> Eastern Ecological Science Center, U.S. Geological Survey, Kearneysville, WV, US

<sup>3</sup> Fort Collins Science Center, U.S. Geological Survey, Fort Collins, CO, US

Received 15 March 2021 – Revised 28 September 2021 – Accepted 1 November 2021

**Abstract** – Higher protein content and greater taxonomic richness of pollen have been noted as beneficial for bee health, but the relationship between pollen protein content and richness as well as how pollen protein content varies across landscape gradients remain understudied. We assessed correlations between pollen protein content and plant species richness and species composition in pollen collected from honey bee colonies across three states over 2 years. We selected sites along an agricultural-grassland gradient to investigate how pollen protein content and species composition varied with land-use. We observed a weak, negative relationship between pollen protein content and taxonomic richness, although we found no evidence of relationships between land-use and either pollen richness or species composition. Results suggest that the richness of pollen collected by honey bees can, albeit weakly, correlate with the pollen protein content in agroecosystems of the Upper Midwest, and that in some systems, land use may have minimal influence on the species collected by honey bees.

**Pollinators / Mass-flowering crops / Foraging / Floral resources / Pollen DNA / Metabarcoding**

## 1. INTRODUCTION

Ample quantities of pollen as well as high-protein pollen contents are essential for honey bee (*Apis mellifera*) development (Di Pasquale et al. 2013), while limited quantity or protein content of pollen can have detrimental health effects (Di Pasquale et al. 2016; Rotheray et al. 2017). Despite the importance of pollen for bees, evidence suggests honey bee foragers may not demonstrate preferences for varying protein content in pollen (Pernal and Currie 2001; Cook et al. 2003, Corby-Harris et al. 2018). However,

honey bees still work to maximize the number of species of pollen collected at the colony level (Nürnberg et al. 2019), suggesting some degree of selection for diverse pollen. Greater pollen richness has demonstrable health benefits for honey bees, such as buffering against the effects of parasitism (Di Pasquale et al. 2013) or improving baseline immunocompetence (Alaux et al. 2010). Most studies investigating relationships between honey bee health and pollen richness have been performed in controlled settings, and thus it is yet unclear whether pollen richness can improve bee health in the field, particularly at landscape scales (Alaux et al. 2017).

The relationship between the landscape-level floral resources and the pollen richness collected

Corresponding author: M. P. Simanonok, msimanonok@usgs.gov

Manuscript Editor: Sara Diana Leonhardt

by bees is unclear. The nutritional content of pollen is often defined by its protein content which is phylogenetically conserved among related plants: i.e., the protein content of a plant species is similar across individuals of the same species (Roulston et al. 2000; Vaudo et al. 2020), but at least some plant species may exhibit environmental variability in their pollen nutrition (Descamps et al. 2021). Thus, the taxonomic composition and richness of floral species may directly affect the pollen protein content available to foraging bees as well as the protein content of pollen they collect (Simanonok and Burkle 2020). Both the richness of flowering species and the nutritional content of available floral resources can be mediated by land-use (Goulson et al. 2015); however, while bees may increase foraging effort in depauperate floral landscapes, evidence suggests neither the quantity nor richness of pollen collected by honey bees change with landscape richness (Danner et al. 2016; Simanonok et al. 2020). This leaves an open question as to how the richness of pollen collected by bees in complex landscapes may relate to the protein content of pollen.

The Prairie Pothole Region (PPR), a region defined by small “pothole” wetlands (Euliss et al. 2004), has experienced dramatic land-use change through the last two decades, primarily via the conversion of historical grassland to row-crop production (Lark et al. 2015). The PPR provides flowers for up to 40% of commercial honey bee colonies in the USA, yet researchers have identified that land-use change in this region is negatively impacting honey bees as well as other bee species (Smart et al. 2016a, b, 2018; Otto et al. 2016) likely via limiting the area of and reducing floral resources across land-use (Smart et al. 2018). The PPR’s recent landscape conversion, its regional importance for the beekeeping industry, as well as the documented negative effects of this land-use change on bees makes this an ideal ecoregion for studying the effects of land-use on floral resources. While previous work has identified changing land-use (Otto et al. 2016), reduction of floral resources (Smart et al. 2018), as well as effects on the nutritional value of available flora (Simanonok et al. 2020), no

study has yet linked how floral resources (e.g., floral richness) and pollen protein content may be interrelated across land-use in this region.

Honey bees have been previously identified as model organisms for quantifying the nutritional content and taxonomic richness of flowers in altered landscapes (Beil et al. 2008; Dimou and Thrasyvoulou 2009; Simanonok et al. 2020), though most studies have focused on broad measures of pollen nutrition or on microscopic identification of pollen (Smart et al. 2017). The advent of DNA metabarcoding has allowed researchers to study bee pollinator foraging and nutrition at scales not previously realized (Bell et al. 2016). The earliest papers on the application of metabarcoding to understanding honey bee foraging were strictly methodological and often lacked repeated sampling across a robust set of sampling sites across space and time (Keller et al. 2015, Richardson et al. 2015a; Cornman et al. 2015). Researchers have recently applied metabarcoding techniques to determine how honey bee pollen collection varies across land cover types (Richardson et al. 2015b, 2021; Smart et al. 2017). However, most of these previous studies inferred honey bee response to land cover by treating individual colonies within an apiary as statistically independent units or did not have substantial replication of study sites across land cover types. For example, Richardson et al. (2021) used metabarcoding to study honey bee pollen foraging across an urban to agricultural gradient, but had only four study sites and lacked replication among agricultural treatment groups. Previous work from our laboratory investigated honey bee pollen collection across six apiaries, but our research revealed the need for replication across additional apiaries to capture the existing land cover gradient more accurately (Smart et al. 2017). Although these studies provided a preliminary demonstration of the use of metabarcoding to honey bee landscape research, additional research is needed across a broader suite of land cover conditions at independent apiaries. Here, we apply metabarcoding to investigate honey bee foraging among 38 apiaries that were distributed across a grassland to row crop gradient and cover an area of approximately 40,000 km<sup>2</sup>.

We previously assessed relationships between land-use (i.e., grasslands, wetlands, and bee forage crops), pollen protein content, and the amount of pollen collected; finding no evidence of linear relationships among those parameters with this dataset, and only a weak non-linear trend with pollen protein content and some land-uses across the growing season (Simanonok et al. 2020). In this manuscript, we expand on prior work by using metabarcoding of pollen samples to assess relationships between pollen richness and pollen protein content, as well as how those relationships may be mediated by land-use change in the PPR. Specifically, we asked (1) does the protein content of pollen or the amount of pollen collected by honey bees change with the taxonomic richness or taxonomic composition of that pollen and (2) does the taxonomic richness or taxonomic composition of pollen collected by honey bees change with land-use?

## 2. METHODS

### 2.1. Site selection

Site selection has been previously described (Smart et al. 2018; Simanonok et al. 2020). Briefly, we intended to select study apiaries which represented a gradient of land-use from primarily grasslands to mostly row-crops. We partnered with commercial beekeepers in Minnesota, North Dakota, and South Dakota. We quantified the land-use in a 4 km radius around each apiary following the Cropland Data Layer (CDL, USDA National Agricultural Statistics Service Cropland Data Layer 2014). We categorized land-use following Smart et al. (2018): grasslands (CDL: grassland, conservation lands, pasture, fallow land, wildflowers, and hay land), bee forage crops (CDL: alfalfa, canola, and sunflower), and wetlands (CDL: herbaceous and woody wetlands). We grouped apiaries together based on whether they had low, medium, or high surrounding floral resources:  $> 1$  s.d.,  $\pm 1$  s.d., and  $< 1$  s.d., respectively from the overall mean (Smart et al. 2018). We then randomly assigned 30 apiary points split evenly across each state

while maintaining the land-use gradient. Six additional apiaries were added to the study which were part of a parallel long-term bee health dataset (Smart et al. 2016b). In addition, we selected an additional two apiaries in ND surrounded by less than 1,000 ha grassland cover to help account for beekeeper site-selection bias and enhance our total land-use gradient. In total our study design encompassed 38 apiaries studied in both 2015 and 2016.

### 2.2. Pollen collection

Honey bee colonies were owned and operated by commercial beekeepers who employed standard best practices for keeping honey bees. Two colonies per apiary were fit with a 10-frame Superior Pollen Trap (Mann Lake, Hackensack, MN, USA). We ensured each colony fitted with a pollen trap was queenright and free from symptoms associated with observable diseases (e.g., fungal and bacterial infections, common viral symptoms). Pollen traps were opened for 72 h every 2 weeks beginning in early June and ending in mid-September. Pollen was collected from pollen traps into sealed plastic bags and kept on ice during transportation to the laboratory. Pollen was stored at  $-20$  °C until analysis. We sorted pollen samples in the laboratory to remove non-pollen material, and the resulting sorted pollen measurement was used as our measure of pollen quantity. For pollen protein content analysis and pollen metabarcoding we collected two 5 g sub-samples of pollen to homogenize with a mortar and pestle. Pollen samples were then dried at 60 °C for 60 h. One of each sub-sample were then shipped to Midwest Labs in Omaha, NE for pollen protein content analysis (AOAC 990.03) to determine pollen crude protein content and to the U.S. Geological Survey's Lee-town Science Center, Kearneysville, WV, for metabarcoding analysis. Pollen protein content is reported as percent protein content per 5 g sub-sample. While using different sub-samples for protein content and genetic analyses has the possibility to influence our results if each sub-sample contained different species compositions

of pollen, it is methodologically impossible to assess this possibility as both sub-samples are destroyed in the analysis. Pollen protein content data are available in Simanonok and Otto (2019) and have been previously analyzed in Simanonok et al. (2020).

### 2.3. DNA sequencing and bioinformatics

Analyses of relationships between pollen protein content, pollen quantity, and land-use for some of these samples were previously described in Smart et al. (2018) and Simanonok et al. (2020) while detailed molecular methods are available in Cornman et al. (2015). Briefly, DNA was extracted using a modified Doyle's method (Doyle 1991; Cornman et al. 2015). We amplified the internal transcribed spacer (ITS) of the nuclear ribosomal locus and sequenced it on the Illumina MiSeq to generate non-overlapping paired-end reads of 300 bases in length. Within each sample, *de novo* operational taxonomic units (OTU) were selected by clustering the first read of each pair at 97% identity with *vsearch* (v. 2.14.1; Rognes et al. 2016) with "iddef" set to 1. Cluster representatives were denoised and checked for chimerism using the "unnoise" and "uchime3" commands of *vsearch* with default settings. Cluster representatives and their mates were then searched against the Nucleotide database of the National Center for Biotechnology Information (NCBI) using BLAST+ v. 2.9.0 (Camacho et al. 2009), using default settings for *blastn* but with low-complexity filtering disabled. The representative sequence pair was assigned a taxonomy using the lowest common ancestor method (Huson et al. 2007) as implemented for paired ITS reads by Cornman et al. (2015). Specifically, the common ancestor of all species scoring within 3% of the best bit score, summed across both reads, was assigned to the read pair at one of five standard ranks (species, genus, family, order, or class) or it was left unassigned. Species-level assignments that did not have an average percent identity to database references of at least 96% and a best bit score of at least 550 were demoted to genus. Genus assignments

that did not have an average percent identity to database references of at least 92% and a best bit score of at least 450 were demoted to family. The abundance of the cluster was not taken from the *vsearch* cluster size for the first read, but instead determined by remapping read pairs to OTUs with *bowtie2* (Langmead and Salzberg 2012) in "local" mode with the scoring threshold set to "G,80,8" (see Langmead and Salzberg 2012 for details). The resulting alignment file was filtered to retain pairs that matched at 97% identity and with a maximum of five indel positions (alignment skips of up to five position at OTU ends were counted as single gap positions, Cornman et al. 2015). If reads of a pair mapped to two different OTUs (e.g., due to chimeras or levels of differentiation at ITS1 versus ITS2), the count was tabulated at the taxonomic level in common to the two OTUs. Counts valid by this scheme were then tabulated per sample. We opted to use an OTU approach vs. other "exact sequence" methods for several reasons: briefly, we used OTU representative sequences to identify candidate taxa *de novo*, but actual abundances are obtained by a direct mapping approach to reduce artifacts associated with clustering by a percent identity method. The lowest common ancestor method considers taxa scoring within 3% of the best match as equally plausible candidates, and that best match itself is usually less than perfect—it is therefore moot in practice to further partition clusters if the assignment method is not correspondingly more precise. Furthermore, recent research has suggested that selecting OTU vs. "exact sequence" methods does not alter the results of ecological community analyses (e.g., Glassman and Martiny 2018).

We removed taxa assigned from OTUs which did not occur in our study area based on the USDA Plants Database (USDA NRCS 2020). We performed this post hoc filtering as opposed to pre-analyses curation because our previous attempts to use USDA PLANTS, field observations, and other checklists have shown that unsatisfactory gaps persist (e.g., Simanonok et al. 2021a). Furthermore, we did not want to omit possible plant species given that the scale of our study covered several US states and honey

bees have an expansive foraging range and niche breadth. OTU counts were converted to proportions of each pollen sample to scale the range of values. In addition, to reduce the effects of contamination and crosstalk we removed all taxa that comprised less than 2% of counts for each pollen sample. Using a low percentage cut-off is relatively standard in molecular analyses of this nature (e.g., Deagle et al. 2019) and has been previously used for similar pollen studies (Simanonok et al. 2021a). While this threshold may limit our detection of rare taxa or reduce the richness detected, it helps prevent against identifying non-target taxa in the samples (e.g., some wind-pollinated species) as well as hedge against contamination or erroneous reads. Furthermore, given the dominance of highly abundant, weedy species in our data (see “3”) we do not believe a moderately low 2% threshold would influence our results. Our metabarcoding methods do not always resolve to species (e.g., Simanonok et al. 2021a), instead they provide results at the species, genera, and family levels. Thus, we perform analyses on “taxonomic” richness and composition which include species-, genus-, and family-level identifications from metabarcoding.

## 2.4. Statistical analyses

With the genetic data, we wanted to test whether pollen protein content or the amount of pollen honey bees collected were related to the taxonomic richness detected; in addition, we wanted to investigate if such patterns were related to land-use. We tested these questions with a linear mixed-effects model comparing pollen protein content against taxonomic richness, area of grassland, bee forage crops, and wetland as well as interactive terms for each land-use  $\times$  richness, while again holding sampling date nested within colony nested within apiary as a random effect to account for repeated samples from the same colonies. This same model structure was used for testing if pollen quantity varied with pollen richness, except pollen quantity was log-transformed due to skewness. We expected that pollen protein

content may increase with greater taxonomic richness, in that bees would be more likely to have collected high-protein content pollen with more species collected. Likewise, we expected that the quantity of pollen collected may increase with greater pollen taxonomic richness as foragers may collect greater amounts of pollen with more nutritious value. Furthermore, we tested if the protein content of pollen increased with the amount of total Fabaceae pollen by comparing pollen protein content against the log-ratio transformed proportion of Fabaceae pollen in each sample with a linear mixed-effects model with sampling date nested within colony nested within apiary as a random effect. Log-ratio transformation was used to standardize across samples, as the proportions are sample-specific and may be biased dependent on the DNA amplification of different species of pollen.

To test if pollen protein content varied with pollen taxonomic composition, we performed a principle coordinate analysis (PCO) of the pollen community matrix from a square root-transformed Bray–Curtis distance matrix. We then overlaid pollen protein content as a ‘surface’ across the PCO using the `ordisurf` function from the `vegan` (Oksanen et al. 2018) R package, which effectively builds a generalized additive model (GAM) to assess how pollen protein content changes with pollen community composition. We expected pollen protein content to vary with pollen community composition, as has been previously observed at landscape scales (e.g., Simanonok and Burkle 2020).

We assessed how the taxonomic richness of pollen collected by honey bees changed with land-use by using a linear mixed effects model comparing pollen taxonomic richness across hectares of grassland, bee forage crops, and wetlands with sampling date nested within colony nested within apiary as a random effect to account for repeated samples. We predicted pollen richness would be positively correlated with the area of grassland and wetlands yet decline with bee forage crop acreage.

We wanted to assess the extent to which the taxonomic composition of pollen would vary

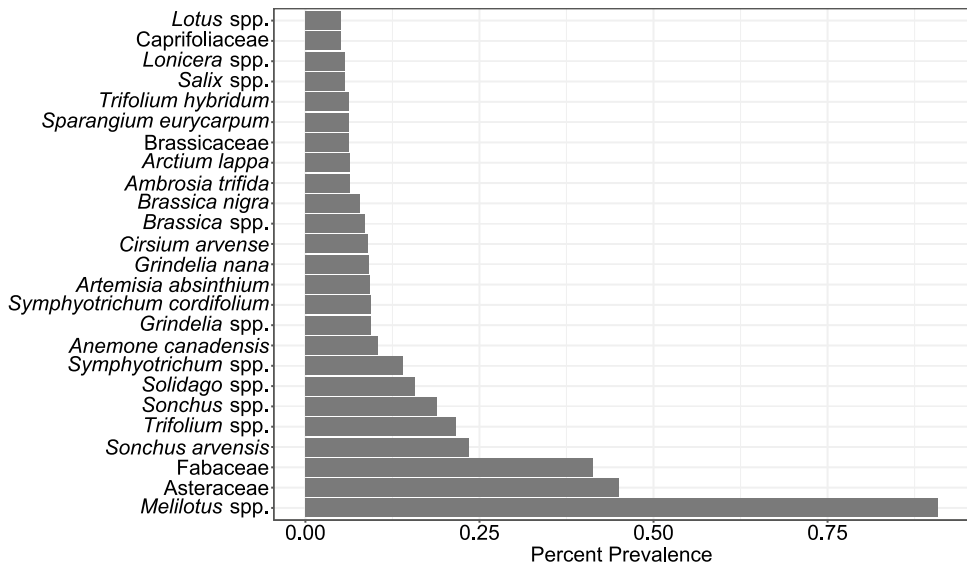
with land-use. To explicitly consider ecological gradients we selected a constrained ordination method, using canonical correspondence analysis (CCA) to assess if the taxonomic composition of pollen varied with land-use (i.e., area grasslands, bee forage crops, or wetlands).

For all linear mixed-effects models (LMMs), we calculated goodness of fit using the `r.squaredGLMM` from the R package `MuMIn` (Barton 2020); for all LMMs, we report both a marginal  $R^2$  ( $mR^2$ ), which represents the variance explained by just the fixed effects, as well as the conditional  $R^2$  ( $cR^2$ ), which can be interpreted as the variance explained by the entire model, considering both fixed and random effects. All analyses were performed in R version 3.5.1 (R Core Team 2018), using packages `lme4` (Bates et al. 2015), `vegan` (Oksanen et al. 2018), and `lmerTest` (Kuznetsova et al. 2017). Pollen protein content data are available in Simanonok and Otto (2019) while pollen taxonomic data are available in Otto et al. (2019).

### 3. RESULTS

#### 3.1. Floral resources

We detected 155 different taxa with the ten most prevalent taxa detected among samples being *Melilotus* spp. (91.3%), Asteraceae (45.1%), unclassified Fabaceae (41.3%), *Sonchus arvensis* (23.5%), *Trifolium* spp. (21.7%), *Sonchus* spp. (19.0%), *Solidago* spp. (15.7%), *Symphyotrichum* spp. (14.0%), *Anemone canadensis* (10.4%), and *Grindelia* spp. (9.5%) (Figure 1). Richness among samples varied from 1 to 14 with a mean richness of  $5.15 \pm 0.08$  SE, and pollen sample quantity had a mean weight of  $90.67 \text{ g} \pm 3.02$  SE. *Melilotus* spp. was consistently the most proportionally dominant taxa detected in all states in both years (Supplemental Figs. S1–S6), being the most dominant taxa detected in 702 of 1034 samples. Asteraceae was the next most proportionally dominant taxa in 44 samples while *Sonchus arvensis* and *Symphyotrichum cordifolium* were the most proportionally present species in 23 samples each.

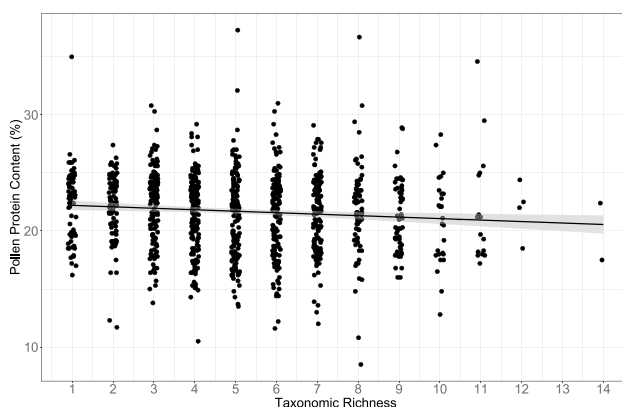


**Figure 1.** Percent prevalence of the 25 most prevalent taxa detected across all pollen samples. Note that taxa detection methods do not always resolve to species-level, instead results are at the species, genera, and family levels. Taxonomic assignment is performed with a lowest-common-ancestor approach to within 3%. This figure thus shows the prevalence of the detection of different taxa assigned from operational taxonomic units, rather than an absolute percent prevalence of detected taxa at higher levels. See “2” and “3” for more detail.

**Table 1**

Model outputs of linear mixed effects models comparing pollen protein content and pollen quantity of each pollen sample against pollen sample richness, area of grassland (ha), area of bee forage crops (ha), area of wetland (ha), and interactive terms of richness and each land-use. *P* values less than 0.05 are in bold. For pollen protein content model  $mR^2=0.03$  and  $cR^2=0.65$ . For pollen quantity model  $mR^2=0.01$  and  $cR^2=0.33$ . Marginal  $R^2$  ( $mR^2$ ) represents the variance explained by just the fixed effects, conditional  $R^2$  ( $cR^2$ ) represents the variance explained by the entire model considering both fixed and random effects

	Estimate	S.E	<i>t</i>	<i>P</i>
Pollen protein content				
Richness	-0.35	0.13	-2.70	<b>0.01</b>
Grassland	$5.06e^{-4}$	$3.25e^{-4}$	-1.56	0.12
Bee forage crops	$3.48e^{-3}$	$3.84e^{-3}$	0.91	0.37
Wetland	$-4.61e^{-3}$	$1.90e^{-4}$	-2.44	<b>0.02</b>
Richness × grassland	$2.00e^{-5}$	$4.30e^{-5}$	0.47	0.64
Richness × bee forage crops	$5.29e^{-5}$	$5.35e^{-4}$	0.10	0.92
Richness × wetland	$7.81e^{-4}$	$2.49e^{-4}$	3.14	<b>&lt; 0.01</b>
Pollen quantity				
Richness	0.02	0.05	0.41	0.69
Grassland	$6.33e^{-5}$	$9.47e^{-5}$	0.67	0.5
Bee forage crops	$5.39e^{-4}$	$1.20e^{-3}$	0.45	0.65
Wetland	$2.34e^{-4}$	$5.70e^{-4}$	0.41	0.68
Richness × grassland	$4.62e^{-6}$	$1.52e^{-5}$	0.30	0.76
Richness × bee forage crops	$-3.88e^{-5}$	$1.90e^{-4}$	-0.20	0.84
Richness × wetland	$-4.90e^{-6}$	$8.92e^{-5}$	-0.06	0.96



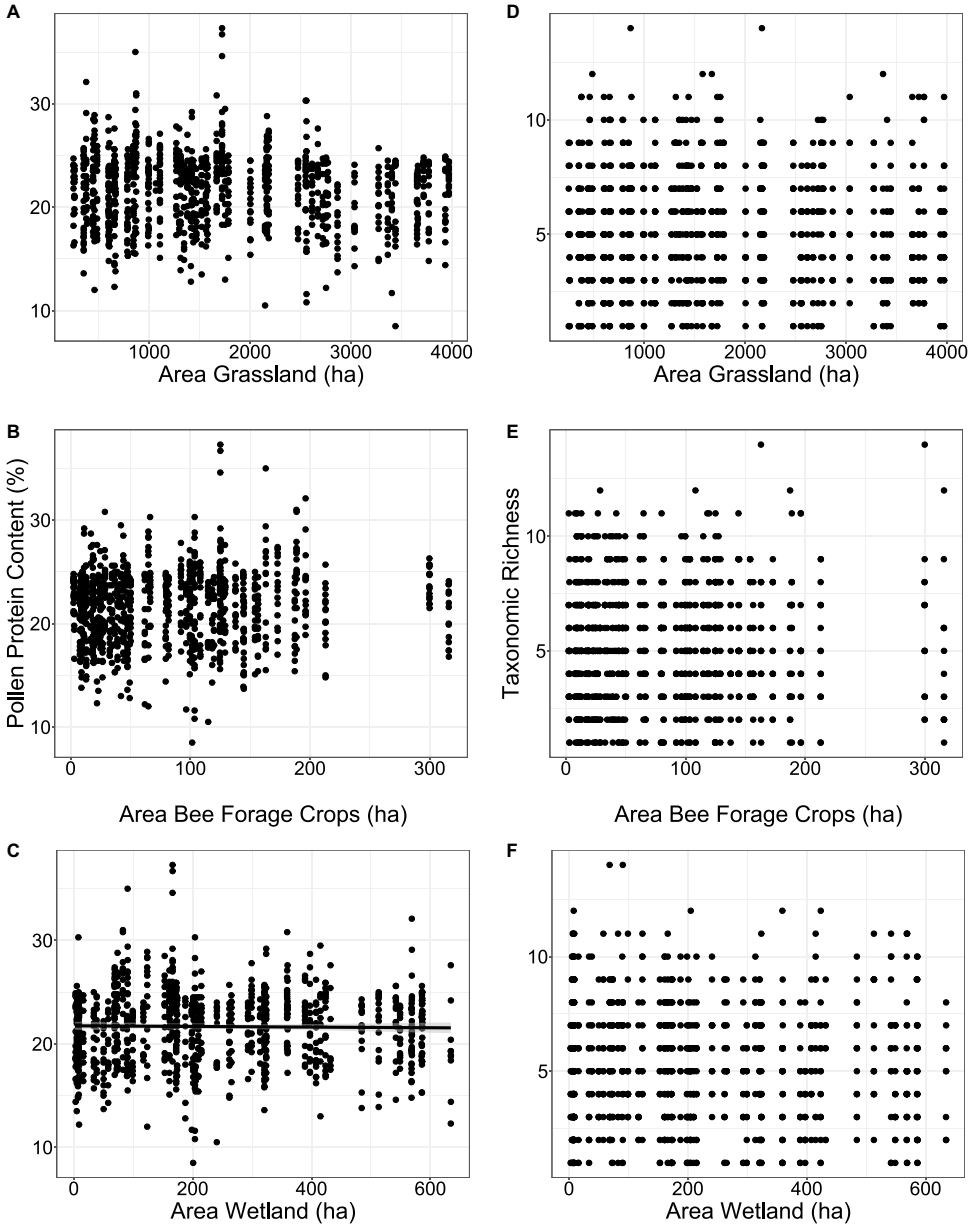
**Figure 2.** Pollen protein content of each pollen sample plotted against that sample's taxonomic richness. Points are jittered slightly along the *x*-axis. Presence of a regression line indicates a statistically significant relationship (i.e.,  $P < 0.05$ ), and shaded area around the line is a 95% confidence interval. See “3” for model output.



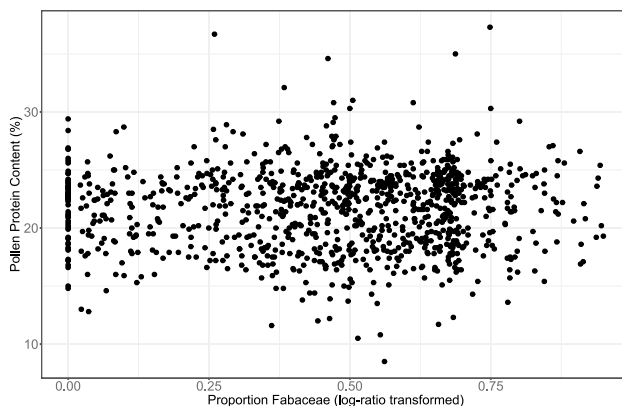
### 3.2. Pollen richness, protein content, quantity, and land use

Pollen protein content declined with increasing pollen richness (Table I, Figure 2), as well

as with increasing area of wetlands (Table I, Figure 3C). Although both of these relationships were statistically significant the strengths of these relationships were extremely weak (Table I). Pollen protein content did not vary



**Figure 3.** Pollen protein content of each sample plotted against area of grasslands (A), bee forage crops (B), and wetlands (C), as well as the taxonomic richness of each sample plotted against area of grasslands (D), bee forage crops (E), and wetlands (F). Points are jittered slightly along the x-axis. Presence of a regression line in C indicates a statistically significant relationship (i.e.,  $P < 0.05$ ), and shaded area around any line is a 95% confidence interval. See Table I for model output.



**Figure 4.** Pollen protein content of each pollen sample plotted against the log-ratio transformed proportion of pollen from all taxa in the family Fabaceae within each sample.

with area of grasslands or bee forage crops (Table I, Figure 3A, B). Likewise, pollen protein content did not change with the log-ratio transformed proportion of total Fabaceae pollen present in pollen samples (estimate: 0.48,  $t = 1.07$ ,  $P = 0.28$ ,  $mR^2 < 0.01$ ,  $cR^2 = 0.10$ , Figure 4). Pollen protein content varied from 8.5 to 37.3%, with a mean value of  $21.68\% \pm 0.11$  SE. Pollen quantity did not vary with the richness of pollen collected nor across any land-use (Table I).

Pollen protein content, when applied as a GAM surface to a PCO of the pollen community composition, varied slightly with pollen taxonomic composition ( $F = 0.60$ ,  $P = 0.03$ ) yet model fit was particularly poor (model  $R^2 < 0.01$ , Deviance explained = 0.66%, Figure 5). Pollen quantity did not vary with pollen taxonomic composition ( $F < 0.01$ ,  $P = 0.38$ , model  $R^2 < 0.01$ , Deviance explained < 0.02%).

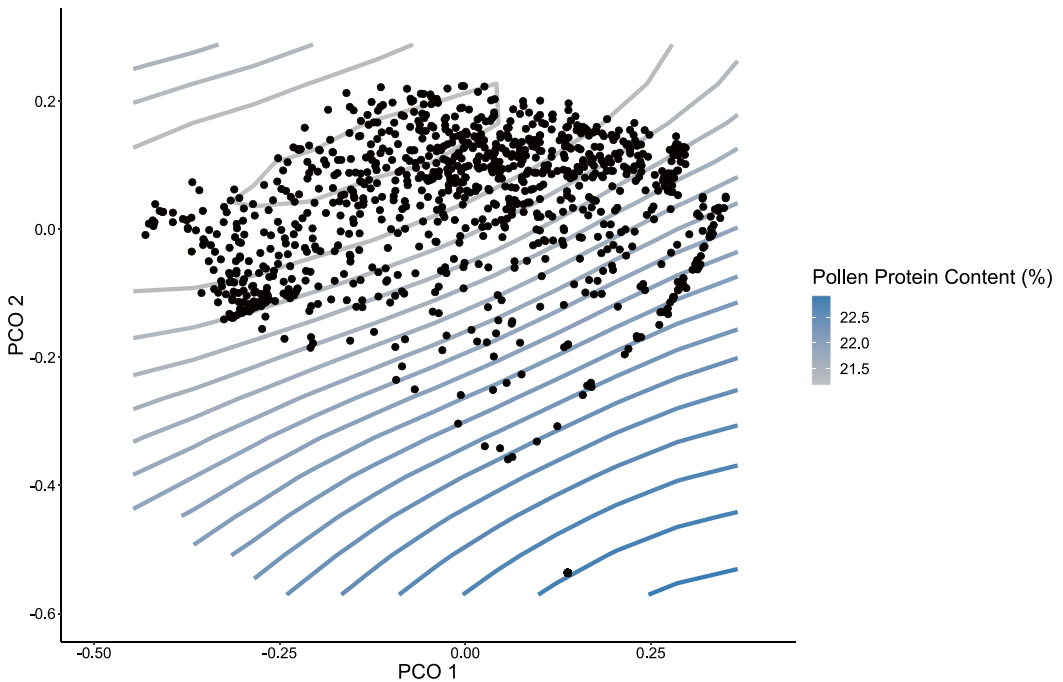
Pollen taxonomic richness collected by each honey bee colony did not vary with surrounding area of grassland ( $t = -1.27$ ,  $P = 0.21$ , Figure 3D), area of bee forage crops ( $t = 0.89$ ,  $P = 0.38$ , Figure 3E), or area of wetlands ( $t = -0.90$ ,  $P = 0.37$ , Figure 3F,  $mR^2 < 0.01$ ,  $cR^2 = 0.15$ ).

CCA results provided no evidence that pollen taxonomic composition varied across any studied land-use gradient (Figure S7). Total inertia, analogous to the dispersion among samples, was 21.59 and constrained axes accounted for only

0.08 of that total inertia leaving most taxonomic composition dissimilarity unexplained (99.6%); eigenvalues for the constrained axes, referring to hectares grassland, bee forage crops, and wetlands, were 0.033, 0.025, and 0.025, respectively.

#### 4. DISCUSSION

Our study is the first multi-state field investigation of honey bee-collected pollen taxonomic richness and protein content across a grassland to row-crop gradient. Counter to our initial predictions, we found little support that the protein content or taxonomic composition of honey bee-collected pollen varied across our land-use gradient. Although we detected a statistically significant trend between pollen protein content and wetland area, when viewed graphically (Figure 3) and considering the magnitude of the estimate of change (Table I), this relationship was weak and most likely driven by a high number of sites with nearly no wetland cover. Furthermore, our data did not demonstrate a strong relationship between pollen protein content and plant richness. Although there was a statistically significant relationship between pollen protein content and taxonomic richness, the total change in pollen protein content was < 2% across pollen samples that varied in richness between 1 and 14 plant taxa, and most samples had relatively



**Figure 5.** Principle coordinate analysis (PCO) of pollen taxonomic composition with surface contours of pollen protein content shaded gray to blue.

few taxa (mean  $5.15 \pm 0.08$  SE). Collectively, we conclude that (1) land-use is not related to the richness or taxonomic composition of pollen collected by honey bees across the gradient we sampled, (2) taxonomic richness of bee pollen may be weakly correlated with the protein content of the pollen, and (3) honey bee-collected pollen is dominated by highly abundant floral species, evidenced by the dominance of *Melilotus* spp. and Asteraceae in our samples.

The weak, negative relationship we observed between pollen richness and pollen protein content is counter-intuitive when considering previous studies highlighting the health benefits and prioritization of pollen richness for honey bees (e.g., Di Pasquale et al. 2013; Nürnberger et al. 2019). However, since the protein content of pollen for different floral species is phylogenetically conserved (Roulston et al. 2000; Vaudo et al. 2020), pollen richness per se may not correlate with pollen protein content, i.e., some work suggests pollen protein content may increase with greater pollen richness if those additional species are high-protein pollen content taxa (e.g.,

Simanonok and Burkle 2020). In addition, our study is limited and thus lacks inference on other important nutritional components of pollen (e.g., phytochemicals, amino acids, lipids, Geldert et al. 2020; Vaudo et al. 2020), which may correlate differently with pollen richness or land-use. We observed a weak, yet statistically significant, negative correlation between pollen protein content and area of wetland, dependent on an interactive effect with pollen richness. Previous research in our region showed that commercial beekeepers are more likely to select apiary locations that have larger area of wetlands in the surrounding landscape (Otto et al. 2016) and wetlands are also important for supporting native bee communities (Vickruck et al. 2019; Begosh et al. 2020), presumably because upland buffers around wetlands harbor floral resources for bees. While these areas may support other aspects of bee health such as refugia from pesticides or by providing nectar plants, our study shows relatively little benefit of wetlands from the perspective of increasing the richness of pollen diets or pollen protein content in honey bees. Honey bees

located near reservoirs have previously been noted to collect pollen from fewer floral species (Guzman et al. 2019). Were a similar pattern to hold in our region with wetlands, it could be that wetlands, despite possibly functioning as habitat refugia for native bees (Evans et al. 2018), do not provide added value to honey bees with respect to pollen nutrition.

Despite previous evidence that pollen protein content and honey bee health can be impacted by variable amounts of bee-friendly land-use in the PPR (Smart et al. 2016a, b, 2018; Simanonok et al. 2020), we observed no strong relationships between the richness or taxonomic composition of pollen collected by bees and the area of grassland, bee forage crops, or wetland near each apiary. There are several potential explanations for this discrepancy. First, our use of a Geographic Information System to quantify these landscapes may inaccurately represent how pollen-foraging honey bees perceive their nutritional landscape. During study design, we hypothesized apiaries surrounded by more grassland would support a greater richness of flowers; however, this design may need further scrutiny. An ideal study design would include intensive field surveys of the area surrounding potential research apiaries so that apiaries can be selected across a gradient of floral richness, as opposed to a gradient of land-use as we have done here. Given the > 4 km foraging radius of honey bees, implementing a design such as this would likely be cost-prohibitive, particularly for a landscape-scale study, unless it could be done in a landscape where the growth of flowering plants can be rigorously controlled. Second, it may be the case that grasslands, despite having avoided conversion to row-crop production, have been sufficiently degraded to where even bee-friendly land-use in the PPR lack floral resources, and thus, a gradient of grassland-cover does not have a gradient of floral richness in our system. Indeed, many of the grasslands in the PPR are used as rangeland or hayed multiple times during the growing season (Lark et al. 2015; Otto et al. 2016). While pasture and hayland may provide refugia from pesticides for bees, grasslands that are heavily used have the lowest flower density in our region,

relative to other bee-friendly land covers (Smart et al. 2021). Third, our inability to ascribe proportional abundance of detected flora to specific land-use with this method may have limited our ability to detect predicted relationships. Thus, while we may not detect a variable number of species across different land-use within honey bee-collected pollen, there may still be considerable variability in floral resources and especially in the quantity of preferred floral species (Simanonok et al. 2021b) that we are unable to assess with this methodology. Such variability may have profound effects on native bee species, many of which are much more specialized in their pollen resources and exhibit different pollen foraging preferences or suitability compared to generalist honey bees (Leonhardt and Blüthgen 2012). Taken together, because of the extreme generalist foraging behavior shown by honey bees coupled with land-use change and degradation in the PPR, we may not observe differences in pollen collection by honey bees regardless of the considered land-use gradient.

Future studies focusing on the variable presence of preferred floral species across different land-uses or more detailed measures of pollen nutritional content such as protein-lipid ratios may provide additional insights into how pollen collection by bees may vary across landscapes (Vaudo et al. 2020). Furthermore, with the substantial native bee richness noted for the PPR (e.g., Evans et al. 2018; Brendel et al. 2019; Vickruck et al. 2019), studying relationships among land-use, pollen taxonomic composition, and pollen nutrition as well as pollen suitability for native bee species may clarify the role that different land-covers in this region provide foraging and nesting resources for both native and managed bee species.

## SUPPLEMENTARY INFORMATION

The online version contains supplementary material available at <https://doi.org/10.1007/s13592-021-00902-5>.

## ACKNOWLEDGEMENTS

We thank the collaborating beekeepers and USGS field technicians who supported this research. Autumn Smart

helped develop the sampling design and deployment of pollen traps. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

## AUTHOR CONTRIBUTION

CRVO, DDI, and RSC conceived of the study and all authors contributed to study design. DDI performed the DNA extraction and genetic analysis. RSC performed the bioinformatics work. MPS performed the statistical analysis and wrote the initial draft of the manuscript. All authors contributed to subsequent drafts of the manuscript.

## FUNDING

USDA-Farm Service Agency (17IAMRECRPHB1) and USDA-Natural Resources Conservation Service (673A7514178).

## AVAILABILITY OF DATA AND MATERIAL

Data used in this study are available in two USGS data releases: Pollen protein and land-use data are available in Simanonok and Otto (2019) while pollen taxonomic data are available in Otto et al. (2019).

## CODE AVAILABILITY

R script can be made available by contacting the corresponding author.

## DECLARATIONS

**Ethics approval** N/A.

**Consent for publication** N/A.

**Consent to participate** N/A.

**Conflict of interest** The authors declare no competing interests.

## REFERENCES

- Alaux, C, Ducloz, F, Crauser, D, Le Conte, Y (2010) Diet effects on honeybee immunocompetence. *Biology Letters* 6(4): 562-565. <https://doi.org/10.1098/rsbl.2009.0986>
- Alaux, C, Allier, F, Decourtye, A, Odoux, J-F, Tamic, T, Chabirand, M, Delestra, E, Decugis, F, Le Conte, Y, Henry, M (2017) A 'landscape physiology' approach for assessing bee health highlights the benefits of floral landscape enrichment and semi-natural habitats. *Scientific Reports* 7(1): 1-10. <https://doi.org/10.1038/2Fsrep40568>
- Barton, K (2020). MuMin: Multi-Model Inference. R package version 1.43.17. <https://CRAN.R-project.org/package=MuMin>
- Bates, D, Maechler, M, Bolker, B, Walker, S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1-48. <https://doi.org/10.18637/jss.v067.i01>
- Begosh, A, Smith, LM, McMurry, ST, Harris, JP (2020) Influence of the Conservation Reserve Program (CRP) and playa wetlands on pollinator communities in the Southern High Plains, USA. *Journal of Environmental Management* 256: 109910. <https://doi.org/10.1016/j.jenvman.2019.109910>
- Beil, M, Horn, H, Schwabe, A (2008) Analysis of pollen loads in a wild bee community (Hymenoptera: Apidae) – a method for elucidating habitat use and foraging distances. *Apidologie* 39: 456-467. <https://doi.org/10.1051/apido:2008021>
- Bell, KL, De Vere, N, Keller, A, Richardson, RT, Gous, A, Burgess, KS, Brosi, BJ (2016) Pollen DNA barcoding: current applications and future prospects. *Genome* 59(9): 629-640. <https://doi.org/10.1139/gen-2015-0200>
- Brendel, C.R., Kral-O'Brian K.C., Hovick, T.J., Limb, R.F., Harmon, J.P. (2019) Plant-pollinator networks in grassland working landscapes reveal seasonal shifts in network structure and composition. *Ecosphere*: e02569. <https://doi.org/10.1002/ecs2.2569>
- Camacho, C, Coulouris, G, Avagyan, V, Ma, N, Papadopoulos, J, Bealer, K, Madden TL (2009) BLAST+: Architecture and Applications. *BMC Bioinformatics* 10(1): 421. <https://doi.org/10.1186/1471-2105-10-421>
- Cook, SM, Awmack, CS, Murray, DA, Williams, IH (2003) Are honey bees' foraging preferences affected by pollen amino acid composition? *Ecological Entomology* 28(5): 622-627. <https://doi.org/10.1046/j.1365-2311.2003.00548.x>
- Corby-Harris, V, Bowsher, JH, Carr-Markell, M, Carroll, MJ, Centrella, M, Cook, SC, Couvillon, M, DeGrandi-Hoffman, G, Dolezal, A, Jones, JC, Mogren, CL (2019) Emerging Themes from the ESA Symposium Entitled "Pollinator Nutrition: Lessons from Bees at Individual to Landscape Levels". *Bee World* 96(1): 1-15. <https://doi.org/10.1080/0005772X.2018.1535951>
- Cornman, RS, Otto, CRV, Iwanowicz, D, Pettis, JS (2015) Taxonomic Characterization of Honey Bee (*Apis mellifera*) Pollen Foraging Based on Non-Overlapping Paired-End Sequencing of Nuclear Ribosomal Loci. *PLoS One* 10: e0145365. <https://doi.org/10.1371/journal.pone.0145365>
- Danner, N, Molitor, AM, Schiele, S, Härtel, S, Steffan-Dewenter, I (2016) Season and landscape composition affect pollen foraging distances and habitat use

- of honey bees. *Ecological Applications* 26(6): 1920-1929. <https://doi.org/10.1890/15-1840.1>
- Deagle, BE, Thomas, AC, McInnes, JC, Clarke, LJ, Vesterinen, EJ, Clare, EL, Kartzinel, TR, Eveson, JP (2019) Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular ecology* 28(2): 391-406. <https://doi.org/10.1111/mec.14734>
- Descamps, C, Boubnan, N, Jacquemart, A-L, Quinet, M. (2021) Growing and Flowering in a Changing Climate: Effects of Higher Temperatures and Drought Stress on the Bee-Pollinated Species *Impatiens glandulifera* Royle. *Plants* 10(5): 988. <https://doi.org/10.3390/plants10050988>
- Dimou, M and Thrasyloulou, A (2009) Pollen analysis of honey bee rectum as a method to record the bee pollen flora of an area. *Apidologie* 40(2): 124-133. <https://doi.org/10.1051/apido/2008066>
- Di Pasquale, G, Salignon, M, Le Conte, Y, Belzunces, LP, Decourtye, A, Kretzschmar, A, Suchail, S, Brunet, JL, Alaux, C (2013) Influence of pollen nutrition on honey bee health: do pollen quality and diversity matter? *PLoS One* 8(8): e72016. <https://doi.org/10.1371/journal.pone.0072016>
- Di Pasquale, G, Alaux, C, Le Conte, Y, Odoux, J-F, Pioz, M, Vaissiere, BE, Belzunces, LP, Decourtye, A (2016) Variations in the availability of pollen resources affect honey bee health. *PLoS One* 11(9): e0162818. <https://doi.org/10.1371/journal.pone.0162818>
- Doyle, J (1991) DNA Protocols for Plants. in Hewitt, GM, Johnston, AWB, Young, JPW (eds.), *Molecular Techniques in Taxonomy*. NATO ASI Series (Series H: Cell Biology), vol 57. Springer Berlin Heidelberg: Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-83962-7\\_18](https://doi.org/10.1007/978-3-642-83962-7_18)
- Euliss, NH, LaBaugh, JW, Fredrickson, LH, Mushet, DM, Laubhan, MK, Swanson, GA, Winter, TC, Rosenberry, DO, Nelson, RD (2004) The wetland continuum: a conceptual framework for interpreting biological studies. *Wetlands* 24(2): 448-458. [https://doi.org/10.1672/0277-5212\(2004\)024\[0448:TWACAF\]2.0.CO;2](https://doi.org/10.1672/0277-5212(2004)024[0448:TWACAF]2.0.CO;2)
- Evans, E, Smart, M, Cariveau, D, Spivak, M (2018) Wild, native bees and managed honey bees benefit from similar agricultural land uses. *Agriculture, Ecosystems & Environment* 268: 162-170. <https://doi.org/10.1016/j.agee.2018.09.014>
- Geldert, C, Abdo, Z, Stewart, JE, Arathi, HS (2020) Dietary supplementation with phytochemicals improves diversity and abundance of honey bee gut microbiota. *Journal of Applied Microbiology* 130(5): 1705-1720. <https://doi.org/10.1111/jam.14897>
- Glassman, SI, and Martiny, JBH (2018) Broad-scale ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. *mSphere* 3(4): e00148-18. <https://doi.org/10.1128/mSphere.00148-18>
- Goulson, D, Nicholls, E, Botías, C, Rotheray, EL (2015) Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* 347(6229): 1255957. <https://science.sciencemag.org/lookup/doi/10.1126/science.1255957>
- Guzman, A, Gaines-Day, HR, Lois, AN, Steffan, SA, Brunet, J, Zalapa, J, Guédot, C (2019) Surrounding landscape and spatial arrangement of honey bee hives affect pollen foraging and yield in cranberry. *Agriculture, Ecosystems & Environment* 286: 106624. <https://doi.org/10.1016/j.agee.2019.106624>
- Huson, DH, Auch, AF, Qi, J, Schuster, SC (2007) MEGAN Analysis of Metagenomic Data. *Genome Research* 17(3): 377-386. <https://doi.org/10.1101/gr.5969107>
- Keller, A, Danner, N, Grimmer, G, Ankenbrand, MVD, Von Der Ohe, K, Von Der Ohe, W, Rost, S, Härtel, S, Steffan-Dewenter, I (2015) Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. *Plant Biology* 17(2): 558-566. <https://doi.org/10.1111/plb.12251>
- Kuznetsova, A, Brockhoff, PB, Christensen, RHB (2017) lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* 82(13): 1-26. <https://doi.org/10.18637/jss.v082.i13>
- Lark, TJ, Salmon, JM, Gibbs, HK (2015) Cropland expansion outpaces agricultural and biofuel policies in the United States. *Environmental Research Letters* 10: 44003. <https://doi.org/10.1088/1748-9326/10/4/044003>
- Langmead, B, and Salzberg, SL (2012) Fast Gapped-Read Alignment with Bowtie 2. *Nature Methods* 9(4): 357-59. <https://doi.org/10.1038/nmeth.1923>
- Leonhardt, SD and Blüthgen, N (2012) The same, but different: pollen foraging in honeybee and bumblebee colonies. *Apidologie*, 43(4), 449-464. <https://doi.org/10.1007/s13592-011-0112-y>
- Nürnberg, F, Keller, A, Härtel, S, Steffan-Dewenter, I (2019) Honey bee waggle dance communication increases diversity of pollen diets in intensively managed agricultural landscapes. *Molecular Ecology* 28(15): 3602-3611. <https://doi.org/10.1111/mec.15156>
- Oksanen J, Blanchet G, Friendly M, Kindt R, Legendre P, McGlenn D, Minchin R, O'Hara B, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2018) vegan: community ecology package. R package version 2.5-3
- Otto, CR, Roth, CL, Carlson, BL, Smart, MD (2016) Land-use change reduces habitat suitability for supporting managed honey bee colonies in the Northern Great Plains. *Proceedings of the National Academy of Sciences* 113(37): 10430-10435. <https://doi.org/10.1073/pnas.1603481113>
- Otto, C, Cornman, R, Iwanowicz, D (2019) Dataset: Molecular identification of honey bee collected pollen in the Northern Great Plains, North America, 2015-2016: U.S. Geological Survey, <https://doi.org/10.5066/P9Z7DYY4>
- Pernal, SF and Currie, RW (2001) The influence of pollen quality on foraging behavior in honeybees (*Apis mellifera* L.). *Behavioral Ecology and Sociobiology* 51(1): 53-68. <https://doi.org/10.1007/s002650100412>

- R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Richardson, RT, Lin, CH, Sponsler, DB, Quijia, JO, Goodell, K, Johnson, RM (2015a) Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. *Applications in Plant Sciences* 3(1): 1400066. <https://doi.org/10.3732/apps.1400066>
- Richardson, RT, Lin, CH, Quijia, JO, Riusech, NS, Goodell, K, Johnson, RM (2015b) Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. *Applications in plant sciences* 3(11): 1500043. <https://doi.org/10.3732/apps.1500043>
- Richardson, RT, Eaton, TD, Lin, CH, Cherry, G, Johnson, RM, Sponsler, DB (2021) Application of plant metabarcoding to identify diverse honeybee pollen forage along an urban-agricultural gradient. *Molecular Ecology* 30(1): 310-323. <https://doi.org/10.1111/mec.15704>
- Rognes, T, Flouri, T, Nichols, B, Quince, C, Mahé, F (2016) VSEARCH: A Versatile Open Source Tool for Metagenomics. *PeerJ* 4:e2584. <https://doi.org/10.7717/peerj.2584>
- Rotheray, EL, Osborne, JL, Goulson, D (2017) Quantifying the food requirements and effects of food stress on bumble bee colony development. *Journal of Apicultural Research* 56(3): 288-299. <https://doi.org/10.1080/00218839.2017.1307712>
- Roulston, TH, Cane, JH, Buchmann, SL (2000) What governs protein content of pollen: pollinator preferences, pollen-pistil interactions, or phylogeny? *Ecological Monographs* 70(4), 617-643. [https://doi.org/10.1890/0012-9615\(2000\)070\[0617:WGPCOP\]2.0.CO;2](https://doi.org/10.1890/0012-9615(2000)070[0617:WGPCOP]2.0.CO;2)
- Simanonok, MP and Burkle, LA (2020) High-severity wildfire limits available floral pollen quality and bumble bee nutrition compared to mixed-severity burns. *Oecologia* 192 (2): 489-499. <https://doi.org/10.1007/s00442-019-04577-9>
- Simanonok, MP, and Otto, CRV (2019) Data release for: 'Do the quality and quantity of honey bee-collected pollen vary across an agricultural land-use gradient?': U.S. Geological Survey <https://doi.org/10.5066/P9DKMY4L>
- Simanonok, MP, Otto, CRV, Smart, MD (2020) Do the quality and quantity of honey bee-collected pollen vary across an agricultural land-use gradient? *Environmental Entomology* 49(1): 189-196. <https://doi.org/10.1093/ee/nvz139>
- Simanonok, MP, Otto, CR, Cornman, RS, Iwanowicz, DD, Strange, JP, Smith, TA (2021a) A century of pollen foraging by the endangered rusty patched bumble bee (*Bombus affinis*): inferences from molecular sequencing of museum specimens. *Biodiversity and Conservation* 30(1): 123-137. <https://doi.org/10.1007/s10531-020-02081-8>
- Simanonok, SC, Otto, CRV, Buhl, DA (2021b) Floral resource selection by wild bees and honey bees in the Midwest U.S.A: Implications for designing pollinator habitat. *Restoration Ecology* <https://doi.org/10.1111/rec.13456>
- Smart, AH, Otto, CRV, Gallant, AL, Simanonok, MP (2021) Landscape characterization of floral resources for pollinators in the Prairie Pothole Region of the United States. *Biodiversity and Conservation* 30: 1991-2015. <https://doi.org/10.1007/s10531-021-02177-9>
- Smart, M, Pettis, J, Rice, N, Browning, Z, Spivak, M (2016a). Linking measures of colony and individual honey bee health to survival among apiaries exposed to varying agricultural land use. *PLoS One* 11(3): e0152685. <https://doi.org/10.1371/journal.pone.0152685>
- Smart, M, Pettis, J, Euliss, N, Spivak, M (2016b) Land use in the Northern Great Plains region of the US influences the survival and productivity of honey bee colonies. *Agriculture, Ecosystems & Environment* 230: 139-149. <https://doi.org/10.1016/j.agee.2016.05.030>
- Smart, M, Otto, C, Cornman, R, Iwanowicz, D (2017) Using colony monitoring devices to evaluate the impacts of land use and nutritional value of forage on honey bee health. *Agriculture* 8(1): 2. <https://doi.org/10.3390/agriculture8010002>
- Smart, MD, Otto, CR, Carlson, BL, Roth, CL (2018) The influence of spatiotemporally decoupled land use on honey bee colony health and pollination service delivery. *Environmental Research Letters* 13(8): 084016. <https://doi.org/10.1088/1748-9326/aad4eb>
- USDA National Agricultural Statistics Service Crop-land Data Layer (2014) Published crop-specific data layer. Available at <https://nassgeodata.gmu.edu/CropScape/> USDA-NASS, Washington, DC.
- USDA, NRCS (2020) The PLANTS Database (<http://plants.usda.gov>, 24 April 2020). National Plant Data Team, Greensboro, NC 27401-4901 USA
- Vaudo, AD, Tooker, JF, Patch, HM, Biddinger, DJ, Coccia, M, Crone, MK, Fiely, M, Francis, JS, Hines, HM, Hodges, M, Jackson, SW, Michez, D, Mu, J, Russo, L, Safari, M, Treanore, ED, Vanderplanck, M, Yip, E, Leonard, AS, Grozinger, CM (2020) Pollen Protein: Lipid Macronutrient Ratios May Guide Broad Patterns of Bee Species Floral Preferences. *Insects* 11(2): 132. <https://doi.org/10.3390/insects11020132>
- Vickruck, JL, Best, LR, Gavin, MP, Devries, JH, Galpern, P (2019) Pothole wetlands provide reservoir habitat for native bees in prairie croplands. *Biological Conservation* 232: 43-50. <https://doi.org/10.1016/j.biocon.2019.01.015>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.