




Preservation methods of honey bee-collected pollen are not a source of bias in ITS2 metabarcoding

Andreia Quaresma · Robert Brodschneider · Kristina Gratzter · Alison Gray · Alexander Keller · Ole Kilpinen · José Rufino · Jozef van der Steen · Flemming Vejsnæs · M. Alice Pinto 

Received: 26 July 2021 / Accepted: 26 October 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract Pollen metabarcoding is emerging as a powerful tool for ecological research and offers unprecedented scale in citizen science projects for environmental monitoring via honey bees. Biases in metabarcoding can be introduced at any stage of sample processing and preservation is at the forefront of the pipeline. While in metabarcoding studies pollen has been preserved at $-20\text{ }^{\circ}\text{C}$ (FRZ), this is not the best method for citizen scientists. Herein, we compared this method with ethanol (EtOH), silica gel (SG) and room temperature (RT) for preservation of pollen collected from hives in Austria and Denmark. After ~4 months of storage, DNAs were

extracted with a food kit, and their quality and concentration measured. Most DNA extracts exhibited 260/280 absorbance ratios close to the optimal 1.8, with RT samples from Austria performing slightly worse than FRZ and SG samples ($P < 0.027$). Statistical differences were also detected for DNA concentration, with EtOH samples producing lower yields than RT and FRZ samples in both countries and SG in Austria ($P < 0.042$). Yet, qualitative and quantitative assessments of floral composition obtained using high-throughput sequencing with the ITS2 barcode gave non-significant effects of preservation methods on richness, relative abundance and Shannon diversity, in both countries. While freezing and ethanol are commonly employed for archiving tissue for molecular applications, desiccation is cheaper and easier to

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10661-021-09563-4>.

A. Quaresma · M. A. Pinto (✉)
Centro de Investigação de Montanha, Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal
e-mail: apinto@ipb.pt

R. Brodschneider · K. Gratzter
Institute of Biology, University of Graz, Universitätsplatz 2, 8010 Graz, Austria

A. Gray
Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK

A. Keller
Center for Computational and Theoretical Biology, Hubland Nord, Würzburg, Germany

A. Keller
Department of Bioinformatics, University of Würzburg, Am Hubland, Biocenter Würzburg, Germany

O. Kilpinen · F. Vejsnæs
Danish Beekeepers Association, Sorø, Denmark

J. Rufino
Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Bragança, Portugal

J. van der Steen
Alveus AB Consultancy, Oisterwijk, Netherlands

use regarding both storage and transportation. Since SG is less dependent on ambient humidity and less prone to contamination than RT, we recommend SG for preserving pollen for metabarcoding. SG is straightforward for laymen to use and hence robust for widespread application in citizen science studies.

Keywords DNA metabarcoding · Pollen DNA barcoding · Preservation bias · Silica gel preservation · Citizen science

Introduction

Pollen collected by honey bees (*Apis mellifera* L.), and further sampled from hives equipped with pollen traps, is frequently investigated in order to understand the floral environment (Bilisik et al., 2008; Danner et al., 2017; Drummond et al., 2018; Jones et al., 2021; Requier et al., 2015; Tosi et al., 2018). The collected pollen typically contains a wealth of information on biodiversity of different landscapes and on important plant sources and their seasonal variation (Brodschneider et al., 2019; Coffey & Breen, 1997; Danner et al., 2017; Donkersley et al., 2014; Lau et al., 2019). A honey bee forager usually prefers to collect pollen within a 1- to 1.5-km radius (3 to 7 km²), although longer foraging distances by honey bees have also been recorded (Beekman & Ratnieks, 2000; Garbuzov et al., 2015). This vast territory that can be explored by hundreds of foragers makes honey bee colonies powerful environmental samplers not only of the floral resources but also of the substances accumulated on bee forage plants, like pesticides or pollutants (Tosi et al., 2018).

Forager bees collect pollen grains in specialised structures of their hind legs (corbiculae) and transport them to the hive. There, they are deposited in wax cells and the processed pollen (bee bread) remains the only source of proteins and lipids for the colony (Brodschneider & Crailsheim, 2010). Numerous studies have sought to gain a broader understanding on honey bee biology and health by examining botanical diversity of pollen loads transported by foragers into the colony (Avni et al., 2014; Danner et al., 2016; Di Pasquale et al., 2013; Smart et al., 2016). For decades, this goal has been addressed by identifying honey bee-collected pollen loads through visual inspection of the grains' exine under a light microscope. However, because this method is time consuming,

labour intensive, expert-knowledge dependent and frequently lacks taxonomic resolution, it is gradually being replaced by alternative approaches (Dunker et al., 2021) of which DNA metabarcoding is gaining increasing popularity (e.g. Bell et al., 2017; Cornman et al., 2015; Danner et al., 2017; Jones et al., 2021; Keller et al., 2015; Macgregor et al., 2019; Potter et al., 2019; Richardson et al., 2019; Smart et al., 2016).

Pollen DNA metabarcoding can achieve high taxonomic identification accuracy (Bell et al., 2018; Hawkins et al., 2015; Kraaijeveld et al., 2015; Richardson et al., 2015b) and, because it is based upon high-throughput sequencing, it allows simultaneous analysis of large numbers of samples at a much faster pace and lower cost than light microscopy (reviewed by Bell et al., 2016). While it is broadly acknowledged that metabarcoding can provide accurate lists of plant species (qualitative data) represented in bee-collected pollen mixtures, there is no clear consensus as to what extent it produces reliable estimates on their relative abundances (quantitative data) (Bell et al., 2018; Keller et al., 2015; Kraaijeveld et al., 2015; Pornon et al., 2016; Richardson et al., 2015b; Smart et al., 2017). Biases of varying sources can be introduced at any step of the sample processing pipeline, leading to inaccurate quantitative results (Bell et al., 2018). However, the major bias in DNA metabarcoding typically occurs in the polymerase chain reaction (PCR) step, when sequence variation in primer-binding sites or taxon-specific differences in amplicon length causes differential amplification rates between species (Bell et al., 2018; Pawluczyk et al., 2015; Piñol et al., 2019; Pompanon et al., 2012). In addition to PCR, biases can also be introduced downstream in the sample processing pipeline, during sequencing and taxonomic classification (Banchi et al., 2020; Richardson et al., 2017), and upstream, during DNA extraction step (Brooks et al., 2015; Pornon et al., 2016; Schiebelhut et al., 2017; Swenson & Gemeinholzer, 2021) or even further earlier during the tissue storage step (Delavaux et al., 2020; Feinstein et al., 2009; Rubin et al., 2013; Weißbecker et al., 2017). Poorly preserved tissue samples are more prone to DNA degradation and this may limit the length of fragments that can be successfully amplified by PCR, potentially influencing qualitative and quantitative outcomes of metabarcoding analysis (Pompanon et al., 2012). This type of bias is

particularly important if species in a mixed sample are differentially sensitive to degradation, which in the case of pollen may be dependent on the protection provided by the exine (Pacini & Hesse, 2005). To minimize preservation bias, samples should be collected and stored using protocols that prevent active DNA degradation (Liu et al., 2020). However, this can be a challenging endeavour in projects that rely on citizen scientists for collecting pollen samples throughout the honey bee season and storing them until molecular analysis.

Citizen science is becoming increasingly applied in ecological studies and also in bee research (Koffler et al., 2021; Miller-Rushing et al., 2012; Moro et al., 2021). The involvement of interested citizens has many advantages. In studies of pollen diversity in different landscapes, the most striking advantage is the extension of the available range of sampling sites, including also private land. Beekeepers have already participated in several studies involving pollen sampling via pollen traps attached to the hives of their colonies (Brodschneider et al., 2019; Drummond et al., 2018; Tosi et al., 2018). A spatially and temporally comprehensive study involving beekeepers collecting pollen has recently been undertaken in the framework of a research project aiming at developing a citizen science protocol for honey bee colonies as environmental bio-samplers. Between April and September of 2020, over 81 citizen scientists collected trapped pollen every two weeks in each of nine apiaries located in ten European countries (Brodschneider et al., 2021; Gratzler & Brodschneider, 2021). In this project, citizen scientists diligently collected and stored trapped pollen in $-20\text{ }^{\circ}\text{C}$ freezers at their own premises and later on shipped the samples to the laboratory for metabarcoding analysis, always maintaining a cold chain. To enhance and upscale citizen science, simple, safe and reliable methods for preserving pollen DNA are needed. Simple and safe in this context means that cost-effective preservation materials are used that can safely and easily be handled by laymen. For instance, cold chain preservation, handling of ethanol or buffers by citizen scientists should be avoided.

The main goal of tissue preservation is to avoid DNA denaturation and degradation, which is accomplished by freezing, desiccating or buffering. Freezing can be done at $-20\text{ }^{\circ}\text{C}$ or at $-80\text{ }^{\circ}\text{C}$ for long-term archiving. At $-20\text{ }^{\circ}\text{C}$, there is still enzymatic

activity and degradation, which can be minimised by adding 95% ethanol (Nagy, 2010). This is the most frequently used method for archiving animal tissue but not plant tissue, especially leaves or twigs (Alsos et al., 2020; Bressan et al., 2014; Chase & Hills, 1991; Doyle & Dickson, 1987; Murray & Pitas, 1996). Samples immersed in ethanol can also be kept at room temperature, although in this case preservation is only good for short-term periods (Nagy, 2010). Desiccation can be achieved by using simple (e.g. air drying, calcium sulphate, silica gel beads) or more complex physical processes (e.g. lyophilisation, cryodesiccation) or by using chemical desiccants (e.g. amyl acetate, xylene). Buffering is an alternative to freezing and desiccation and involves preservation in different buffers, e.g. EDTA, SDS and CTAB (see Nagy, 2010; Prendini et al., 2002, where all these methods are thoroughly reviewed). Pollen preservation for metabarcoding applications has typically been achieved by storing samples at $-20\text{ }^{\circ}\text{C}$ (e.g. Bell et al., 2017; Cornman et al., 2015; Danner et al., 2016; Smart et al., 2016). Whether this is the best method to keep preservation bias to a minimum or whether other methods of storing pollen are suited to metabarcoding applications was unknown until this study.

To facilitate storage at the citizen scientist's premises of a possibly large number of pollen samples collected throughout the honey bee season, while at the same time assuring sample integrity for downward metabarcoding analysis, here we compared four different methods for preserving pollen, namely (1) freezing at $-20\text{ }^{\circ}\text{C}$, (2) storage in ethanol, (3) desiccation at room temperature and (4) desiccation with silica gel beads. Poorly preserved samples may hinder PCR amplification due to various forms of DNA damage, which reduce the average length of intact template for polymerase action (Pompanon et al., 2012). This may introduce a bias during PCR, affecting floral spectra determined for pollen samples from sequence reads generated by high-throughput sequencing. Therefore, it is important to test how the method of preserving bee-collected fresh pollen mixtures influences qualitative and quantitative performance of metabarcoding. The results will inform citizen science projects on the simplest and cheapest method for preserving pollen without compromising the accuracy of downstream botanical identification via metabarcoding.

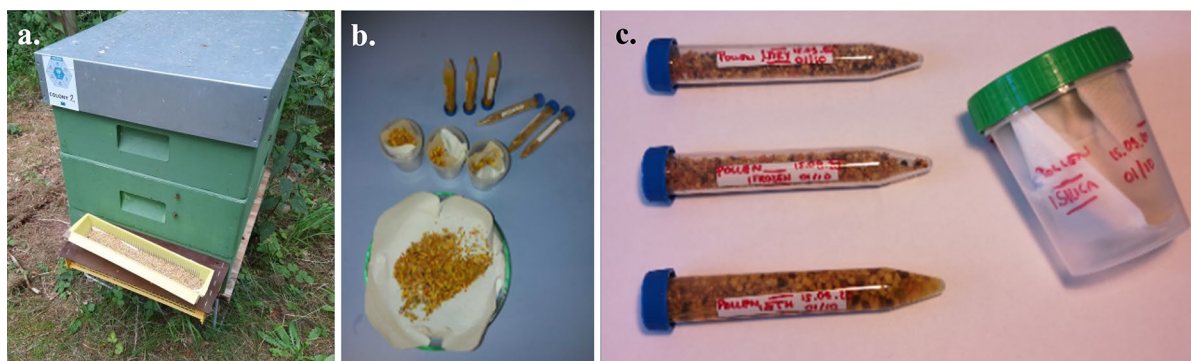


Fig. 1 Pollen sampling and storage. **a** Pollen trap on a bee hive. **b** Pollen drying at room temperature in a Petri dish, preparation for silica gel drying in three 125-mL vials, samples in ethanol and in 15-mL tubes for freezing. **c** Final appearance of

sample replicates for analysis (from top): pollen dried at room temperature, frozen pollen sample, pollen in ethanol and, on the right, silica-dried pollen sample

Materials and methods

Preliminary test: determining the amount of silica gel for efficient pollen desiccation

Prior to comparing performance of the four different pollen preservation methods, we tested the amount of silica gel (SG) required for efficient desiccation of 5 g of freshly collected pollen. Therefore, pollen was harvested on July 13th 2020 from traps placed at beehive entrances in an apiary in Denmark (Fig. 1). The pollen gathered for a ~12-h period was thoroughly mixed and split into 30 homogeneous replicates of 5 g each. Twenty-three replicates were transferred to individual porous paper ‘tea’ filters and placed into individual 125-mL capped vials together with a sachet containing 1 g ($N=7$), 5 g ($N=8$), or 10 g ($N=8$) of SG. Six replicates were exposed at room temperature (RT). Finally, one replicate was set aside for determining the baseline values for water activity and content on the sampling day (Table 1).

Water activity was measured in a total of 13 replicates (Table 1) on July 13th (day 0), 15th (day 1.5), 17th (day 3.5) and 20th (day 6.5) using a LabSwift-aw (Novasina, Lachen, Switzerland) water activity meter (non-destructive method). Water activity is a measure of the amount of free water (available for microbial growth), which ranges from zero to one (pure water). Water content was determined for the same samples as weight percentage of water by drying at 105 °C for 90 min (destructive method) in a KERN DAB 100–3 (Balingen, Germany) moisture

analyser. The remaining 17 samples were kept in a dark cabinet (to avoid DNA damage from UV light) for 10 days after sampling until DNA extraction, as described below.

Comparison of four different pollen preservation methods: 96% ethanol, freezing, silica gel and room temperature

Pollen traps were set up in August and September of 2020 in two apiaries (one in Denmark and another one in Austria) aiming to collect enough pollen for comparing the four different preservation methods (Fig. 1), namely (1) placement in a 15-mL tube filled with 96% ethanol and kept at room temperature (EtOH), (2) placement in a 15-mL tube kept at -20 °C (FRZ), (3) placement in a porous ‘tea’

Table 1 Sample sizes used in the pre-test experiment for determining the amount of silica gel (SG) for efficient pollen desiccation

| Method | Water activity and content | DNA |
|----------|----------------------------|-----|
| Baseline | 1 | 0 |
| RT | 3 | 3 |
| 1 g SG | 3 | 4 |
| 5 g SG | 3 | 5 |
| 10 g SG | 3 | 5 |
| Total | 13 | 17 |

One subset ($N=17$) was used for assessing DNA parameters (yield and quality) and the other ($N=13$) for water activity and content

filter inside a 125-mL capped vial supplied with 12 g of silica gel and kept at room temperature (SG), and (4) placement in a fine gauze/filter paper exposed to room temperature for 1 week followed by transfer to a 15-mL tube kept closed at room temperature (RT).

Pollen harvested over a ~12-h period was thoroughly mixed and split into 88 replicates of 5 g each. Sample size per preservation method varied between countries and sampling dates (Table 2). In Denmark, there was a single sampling event on August 11th ($N=40$). In Austria, due to pollen availability constraints, sampling took place on multiple dates in September (total $N=48$). All 88 samples were stored for ~4 months at room temperature in the dark, except those frozen at $-20\text{ }^{\circ}\text{C}$, until molecular analysis.

Sample preparation and DNA extraction

DNA was isolated from only 87 samples because one of them contained a *Lepidoptera* larva feeding on the pollen, presumably a wax moth (*Galleria mellonella* or *Achroia grisella*). Prior to DNA extraction, a homogenous pollen solution was prepared in a magnetic stirrer using 2 g (for RT, SG and FRZ samples) or 3 g (for EtOH) of pollen sample and 4 mL of sterile ultrapure water. A volume of 200 μL of this solution (~50 mg of pollen) was placed in a 1.5-mL tube and centrifuged at $21,206\times g$ for 3 min. After centrifugation, the supernatant was discarded and 1 mL of absolute ethanol was added to all 87 samples, which were then stored at $-20\text{ }^{\circ}\text{C}$ until DNA extraction.

Immediately before extraction with the Macherey–Nagel NucleoSpin Food Kit, the tubes containing ~50 mg of pollen were centrifuged at maximum speed for 3 min and the ethanol discarded. Pollen samples were then transferred to a 2.0-mL screwcap

tube containing a mix of zirconia beads of varying sizes to target different pollen grain sizes. A volume of 550 μL of the NucleoSpin lysis buffer was added to the 2.0-mL tube and the mixture was ground in a Precellys 24 tissue homogeniser (Bertin Instruments) three times at 6200 rpm for 5 s. The following steps of DNA extraction were implemented according to manufacturer’s instructions. After extraction, quality and yield were measured in a SPECTROstar Nano (BMG Labtech) and the DNA extracts were diluted to 10 $\text{ng}/\mu\text{L}$ before PCR.

DNA metabarcoding

DNA metabarcoding was performed using the internal transcribed spacer 2 (ITS2). PCR was carried out in triplicate for each sample using the primers ITS-S2F (Chen et al., 2010) and ITS-S4R (White et al., 1990) and a two-stage process. Stage one PCR was performed in a 10- μL total volume containing 5 μL of Q5 High-Fidelity 2X Master Mix (New England Biolabs), 0.5 μL of each primer at 10 μM , and 1 μL of DNA at 10 $\text{ng}/\mu\text{L}$. Thermal cycling conditions were 98 $^{\circ}\text{C}$ for 3 min, 35 cycles of 98 $^{\circ}\text{C}$ for 10 s, 52 $^{\circ}\text{C}$ for 30 s and 72 $^{\circ}\text{C}$ for 40 s, and a final extension of 72 $^{\circ}\text{C}$ for 2 min. The amplicons were purified using 0.8 \times reversible immobilization paramagnetic beads (Agencourt AMPure XP) per microlitre of PCR product and then subjected to the second-stage PCR for incorporation of the unique indexes. PCR was prepared in a 10- μL total volume containing 5 μL of KAPA HiFi HotStart ReadyMixPCR Kit (Kapa Biosystems), 0.5 μL of each oligonucleotide at 1 μM and 2 μL of 1:10 dilution of the purified amplicons. Thermal cycling conditions were 95 $^{\circ}\text{C}$ for 3 min, followed by 10 cycles of 95 $^{\circ}\text{C}$ for 30 s, 55 $^{\circ}\text{C}$ for 30 s, 72 $^{\circ}\text{C}$ for 30 s and a final extension

Table 2 Number of samples collected for each preservation method (EtOH—ethanol, FRZ—frozen at $-20\text{ }^{\circ}\text{C}$, RT—room temperature, SG—silica gel), country, and sampling date

| Country | Sampling date | EtOH | SG | RT | FRZ | Total |
|---------|---------------|------|----|----------------|-----|-------|
| Denmark | 11/08/2020 | 10 | 10 | 10 | 10 | 40 |
| Austria | 05/09/2020 | 3 | 3 | 3 | 3 | 12 |
| | 10/09/2020 | 3 | 4 | 4 | 3 | 14 |
| | 15/09/2020 | 5 | 6 | 6 ^a | 5 | 22 |
| Total | | 21 | 23 | 23 | 21 | 88 |

^aOne sample contained a *Lepidoptera* larva, presumably wax moth

of 72 °C for 5 min. Indexed amplicons were purified with the paramagnetic beads, as before, quantified in the Epoch Microplate Spectrophotometer (Bio Tek Instruments), normalised to a final concentration of 10 nM, and then pooled (one pool per 96-well plate). The amplicon size distribution was determined for each pool on a TapeStation 2200 using the HS D1000 kit (Agilent Technologies). Pools were quantified by a SYBR green quantitative PCR assay using the KAPA Library Quantification kit (Kapa Biosystems) and then combined equimolarly into one single sequencing library containing all samples. The sequencing library was diluted to 2 nM, spiked with 10% Illumina-generated PhiX control library and then sequenced on the Illumina MiSeq using the 2×250 cycles v2 chemistry, according to manufacturer's instructions.

Bioinformatics

Pools were de-multiplexed in BaseSpace Sequence Hub based on their unique indexes incorporated in the stage two PCR. Raw sequence reads (Fastq files) were processed using VSEARCH v2.15.2 (Rognes et al., 2016). Reads R1 and R2 were merged using the *fastq_mergepairs*. Low-quality reads (length < 200 bp and > 500 bp, ambiguous base pairs) and chimeras were discarded using *uchime3_denovo*. After filtering, sequence reads were classified directly to the species level using *usearch_global*, with sequence similarity set at 97%. Unclassified reads were subjected to hierarchical classification, with a similarity of minimum level of bootstrap support for the taxonomic rank of 90%. The ITS2 reference database used in the taxonomic classification was updated from Sickel et al. (2015), with sequences retrieved from the National Center for Biotechnology Information (NCBI) platform (<https://www.ncbi.nlm.nih.gov>). A community matrix format table, with columns representing samples and rows species, and a file with the taxonomic

lineage of each species were created and imported into R-Studio v1.2.5033 (Team, 2015) using the package Phyloseq v1.27.6 (McMurdie & Holmes, 2013).

Data analysis

The relative abundances (RA) of high-quality reads were used as a proxy of pollen quantity estimates at both family and species levels, after discarding taxa below 1% abundance. A species accumulation curve was used to assess the sequence depth of each sample using the package *ranacapa* (Kandlikar et al., 2018) in R-Studio v1.2.5033 (Team, 2015). As a measure of pollen diversity, the Shannon–Wiener index (H') was calculated for each sample replicate from both family and species classification data. H' values, which combine a component on the number of taxa (richness) with another on their relative abundances (evenness), were computed using the BiodiversityR package v 2.13–1 (Kindt & Coe, 2005) in R-Studio v1.2.5033 (Team, 2015). Statistical comparisons among the four preservation methods were also performed in R-Studio by using Kruskal–Wallis and χ^2 tests. When statistical differences were found, multiple comparisons were performed using Dunn's test with Bonferroni P value adjustment, using the FSA package v0.8.32 (Ogle et al., 2021) in R-Studio v1.2.5033 (Team, 2015). All descriptive statistics are reported as the median and interquartile range (IQR). The level of statistical significance for all tests was set at $\alpha = 0.05$.

Results and discussion

Determining the amount of silica gel for efficient pollen desiccation

The baseline values for water content and activity measured soon after sampling (day 0) were 21.5% and 0.69, respectively (Table 3). These values decreased

Table 3 Temporal variation of water content (WC, %) and activity (WA) measured on 5 g of pollen desiccated with different amounts silica gel (SG) and at room temperature (RT)

| Method | 13 July (day 0) | | 15 July (day 1.5) | | 17 July (day 3.5) | | 20 July (day 6.5) | |
|---------|-----------------|------|-------------------|------|-------------------|------|-------------------|------|
| | WC | WA | WC | WA | WC | WA | WC | WA |
| RT | 21.5 | 0.69 | 16.4 | 0.53 | 13.9 | 0.48 | 14.8 | 0.53 |
| 1 g SG | | | 21.0 | 0.68 | 21.0 | 0.69 | 18.0 | 0.57 |
| 5 g SG | | | 16.2 | 0.54 | 13.2 | 0.44 | 11.1 | 0.37 |
| 10 g SG | | | 14.5 | 0.41 | 11.3 | 0.28 | 9.4 | 0.24 |

substantially after a week of pollen desiccation with silica gel (SG) or at room temperature (RT). The lowest values of water content were observed when pollen was dried with 10 g of SG, reaching 9.4% at the last measurement on July 20th (day 6.5). This was lower than the water content obtained for pollen dried at RT (14.8%) and almost half of the value obtained with 1 g of SG (18.0%) for the same period.

Water activity <0.6 is required to prevent microbial growth in pollen (Beuchat, 1983). This threshold was already exceeded at day 1.5 for RT and with 5 and 10 g of SG. However, while water activity levelled out around 0.5 for RT, with some variation probably due to changes in environmental humidity, it continued steadily decreasing with time for pollen dried with 5 and 10 g of SG. In contrast, it was only at day 6.5 when 1 g of SG was able to lower water activity below 0.6. While 1 g of SG was clearly insufficient for drying 5 g of pollen, 10 g of SG enabled the fastest and most efficient desiccation, preventing moulds from developing. Fungal contamination can be a problem for pollen samples requiring botanical identification via ITS2 metabarcoding. The DNA of fungi will be extracted alongside that of pollen and might eventually be co-amplified in the PCR. Due to high taxonomic resolution, ITS2 has been one of the markers of choice in pollen metabarcoding (e.g. Bell et al., 2017; Cornman et al., 2015; Danner et al., 2017; Jones et al., 2021; Keller et al., 2015; Milla et al., 2021; Nürnberger et al., 2019; Richardson et al., 2015b; Sickel et al., 2015). The problem is that PCR using ITS2 barcodes may suffer from non-specificity, resulting in co-amplification of fungal contaminants (Cheng et al., 2016). While these contaminants will not prevent ITS2 plant fragments from being sequenced in

high-throughput sequencing platforms, they will consume sequence reads and thereby limit the number of pollen samples that can be pooled in a sequencing run (Bell et al., 2016; Cornman et al., 2015).

Yield and quality of the DNA isolated from the 17 samples, stored for 10 days at RT and with varying amounts of SG, are shown in Fig. 2 (see Online Resource 1 for details). The highest DNA concentration was observed for pollen desiccated with 10 g of SG (41.69 (6.85) ng/μL) and the lowest for RT (29.48 (4.23) ng/μL), although the differences across the four methods were non-significant (*P* value=0.080, Kruskal–Wallis test). Likewise, there were no statistical differences in DNA quality (*P*=0.408), with all methods showing 260/280 absorbance ratios slightly below the optimal 1.8 value (median=1.7; Online Resource 1). These results suggest that all pollen desiccation methods are able to properly preserve DNA for downstream metabarcoding analysis. Using more samples might lead to a significant difference between desiccation with 10 g of SG and RT in terms of DNA concentration.

Although 10 g of SG was sufficient for desiccating 5 g of bee-collected fresh pollen, in the ensuing experiment pollen samples were preserved with two pre-prepared sachets of 6 g each available on the market. According to the SG manufacturer, one 6-g sachet of well-dried SG can absorb 1 g of water, corresponding to 100% water in 5 g of pollen with 20% water content. Since water content in bee-collected fresh pollen typically varies between 15 and 30% (Canale et al., 2016; Herbert & Shimanuki, 1978), one sachet would suffice to desiccate 5 g of pollen and two would allow over-absorbance, hence assuring

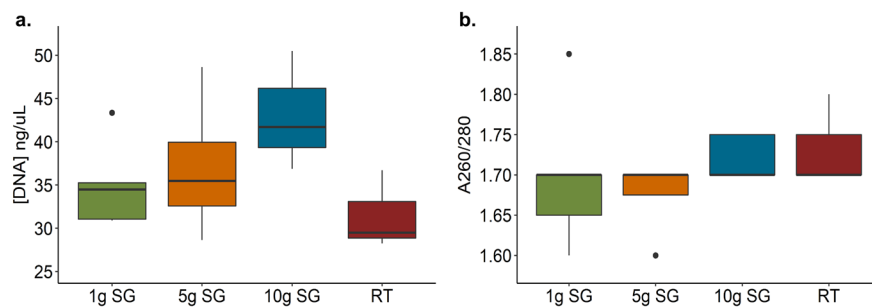


Fig. 2 Boxplots for **a** DNA concentration (ng/μL) and **b** quality (260/280 absorbance ratio) measured for pollen desiccated at room temperature (RT; *N*=3) and using 1 g (*N*=4), 5 g

(*N*=5) and 10 g (*N*=5) of silica gel (SG). The optimal DNA quality is obtained when the 260/280 absorbance ratio reaches 1.8

proper DNA preservation in a wide range of environmental conditions.

Comparing concentration and quality of DNA across preservation methods

DNA quality and concentration measured for extracts prepared from the 87 pollen samples stored in ethanol (EtOH), with 12 g of silica gel (SG), at -20°C (FRZ), and at room temperature (RT) are shown for Austria and Denmark in Fig. 3 (see details in Online Resource 2). Statistical analysis detected significant differences in DNA quality among preservation methods for the Austrian samples ($P=0.008$, Kruskal–Wallis test) for comparisons of RT [1.83 (0.10)] with FRZ [1.80 (0.05); $P=0.011$, Dunn's test] and with SG [1.80 (0.05); $P=0.027$, Dunn's test]. For Denmark, DNA quality was similar across the four methods ($P=0.045$, Kruskal–Wallis test; however, Dunn's test did not find any significant difference between pairs of methods), with most samples exhibiting A260/280 absorbance ratios close to the optimal 1.8 threshold (median = 1.80 calculated over all samples; Online Resource 2). In contrast, DNA concentration was more heterogeneous, with RT samples showing about three times as much DNA as EtOH samples in both Austria [RT 133.51 (70.14) ng/ μL and EtOH 30.19 (7.97) ng/ μL , $P=9.734\times 10^{-6}$, Dunn's test] and Denmark [RT 116.65 (60.41) ng/

μL and EtOH 35.37 (3.96) ng/ μL ; $P=5.460\times 10^{-5}$, Dunn's test]. Statistical differences were also found for the comparisons between EtOH and FRZ in both countries ($P>2.012\times 10^{-5}$) and EtOH and SG in Austria ($P=0.042$).

The downward bias observed for DNA extracted from pollen preserved in EtOH could be due to a technical artefact. In the homogenisation step, a lower amount of pollen was often collected at the bottom of the 1.5-mL tube upon centrifugation, despite the 33% excess of starting material (3 g instead of 2 g of sample) to compensate for the ethanol mass. It is also possible that samples exposed at RT were contaminated by organisms that feed on pollen. The observation of a *Lepidoptera* larva developing in a RT sample is a living proof of such a circumstance. While insect larvae can be detected during sample preparation for DNA extraction, eggs would go unnoticed and therefore extracted alongside with pollen.

Pollen contamination may result in higher DNA yields, depending on the amount of tissue and genome size of contaminants (Bell et al., 2018; Pornon et al., 2016). This is especially the case when employing kits for isolating DNA from food matrices, as here, which are tailored for efficient DNA recovering from a wide range of organisms. It should be noted, however, that non-plant contaminants will not hamper downstream metabarcoding analysis as they will never be co-amplified with pollen when using

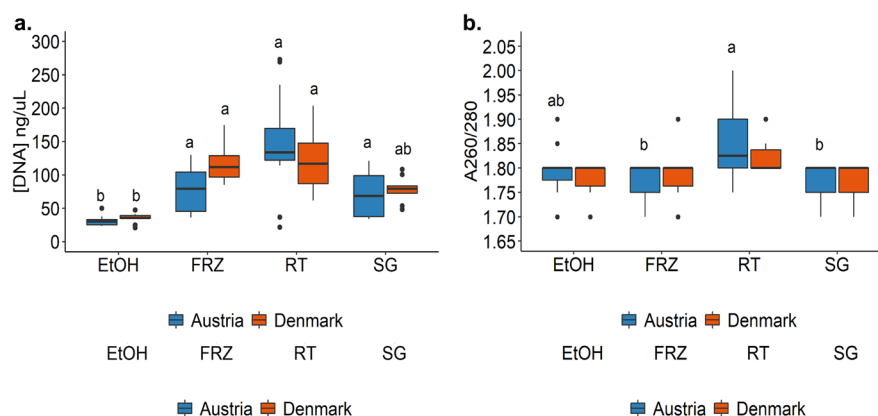


Fig. 3 Boxplots for **a** DNA concentration (ng/ μL) and **b** quality (260/280 absorbance ratio) measured for pollen stored in ethanol (EtOH; Austria $N=11$, Denmark $N=10$), frozen at -20°C (FRZ; Austria $N=11$, Denmark $N=10$), at room temperature (RT; Austria $N=12$, Denmark $N=10$) and with

12 g of silica gel (SG; Austria $N=13$, Denmark $N=10$). The optimal DNA quality is obtained when the 260/280 absorbance ratio is 1.8. Different letters above boxplots indicate significant differences ($P < 0.05$) in DNA quality and concentration among preservation methods

plant-specific ITS2 primers (Bell et al., 2016; Cheng et al., 2016). The problem for metabarcoding studies on bee-collected pollen arises when contamination originates from airborne pollen grains (Kraaijeveld et al., 2015), as it may comprise an important proportion of the sequencing reads. Therefore, while desiccation at RT offers an attractive alternative for preserving pollen in citizen science projects, the risk of contamination (plant or non-plant) makes this method less suitable for recommendation for use in metabarcoding studies.

Comparing pollen preservation methods by ITS2 metabarcoding

The sequencing run in the Illumina MiSeq platform produced a total of 2,549,419 high-quality reads (calculated over all samples and countries, $N=87$) and all of them were classified as Viridiplantae. Of these, 2,437,347 reads were classified at the species level and 112,072 did not pass beyond the genus level. The median number of reads per sample was 29,507.0 (14,038.0) at the family level (Online Resource 3) and 28,360.0 (12,970.0) at the species level (Online Resource 4). This sequencing depth allowed sequencing each sample to saturation, with the lowermost 1660 number of reads obtained for a RT sample (Online Resource 4) well above the plateau reached at ~1000 reads in the species accumulation curve (Fig. 4). These results suggest that a relatively complete list of species-level assignment was recovered for all the samples with ~1000 high-quality reads, not far from the 3000 reads reported by Sickel et al. (2015) and below the ~50,000 reads reported by Cornman et al. (2015).

The distribution of the total number of reads was similar across methods for both Austria ($P=0.238$, Kruskal–Wallis test) and Denmark ($P=0.120$), as shown in Fig. 5. This sequencing result suggests that DNA integrity was not differentially affected by the mode of preserving pollen samples.

The classification algorithm identified a total of 51 plant species, belonging to 40 genera and 22 families, in the 87 pollen samples. Floral spectra sampled by honey bees differed between Austria and Denmark, with only nine families and seven species overlapping (Online Resources 5 and 6). Nonetheless, all detected taxa have been listed elsewhere as foraging sources

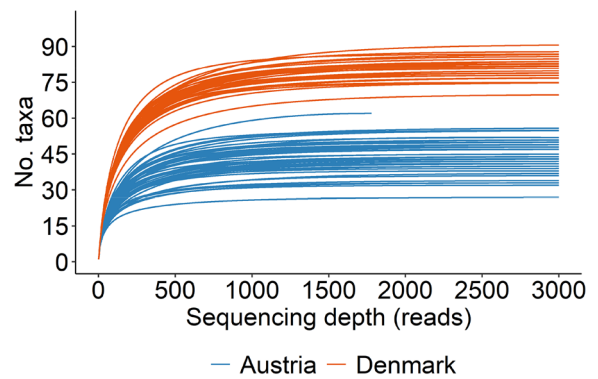


Fig. 4 Species accumulation curve obtained for the 87 pollen samples. The x-axis was trimmed at 3000 reads as the saturation of nearly all samples was below this threshold

for honey bees (Brodschneider et al., 2019; Danner et al., 2017; Lau et al., 2019), suggesting that pollen preserved using a range of methods can be reliably identified via metabarcoding. Plant diversity was also distinct between the two countries with the Shannon index (H') showing higher values, and therefore higher diversity, in Denmark, with median of 2.14 (0.10) for families and species, than in Austria, with median of 1.80 (0.10) for families and 1.89 (0.12) for species (calculated over all samples).

Relative abundances (RA) are depicted for the 22 families and 51 plant species in Fig. 6, facilitating comparisons on the performance of the four preservation methods at both taxonomic levels (see Online Resources 7 and 8 for details). A total of 15 families were detected in Austria and 16 in Denmark, and these numbers lie within the range that has been reported for single sampling events in central Europe (Brodschneider et al., 2019; Danner et al., 2017). Most Austrian (9) and Danish (10) families were rare, exhibiting RA values <5% (Online Resource 7). Just three families accounted for the majority of the pollen collected in Austria [median of 62.2% (13.7)] and Denmark [median of 57.0% (7.3), calculated over all samples; Online Resource 7]. Of these top three families, *Ranunculaceae* was the least abundant and the only one shared between the two countries, with similar proportions [15.1% (3.8) in Austria; 14.6% (3.5) in Denmark]. The other two dominant families were *Balsaminaceae* [32.3% (4.5)] and *Brassicaceae* [14.8% (5.4)] in Austria and *Asteraceae* [23.0% (2.3)] and *Papaveraceae* [19.4% (1.5)] in Denmark. While

Asteraceae, *Brassicaceae* and *Papaveraceae* are often listed as ubiquitous pollen sources for honey bees, *Balsaminaceae* is less common (Brodschneider et al., 2019; Coffey & Breen, 1997; Danner et al., 2017; Dimou & Thrasylvoulou, 2007; Potter et al., 2019; Richardson et al., 2019). However, the relative importance of forage varies with land cover and season, and all these herb families include late summer bloomers (Bilisik et al., 2008; Coffey & Breen, 1997; Dimou & Thrasylvoulou, 2007; Donkersley et al., 2014; Lau et al., 2019).

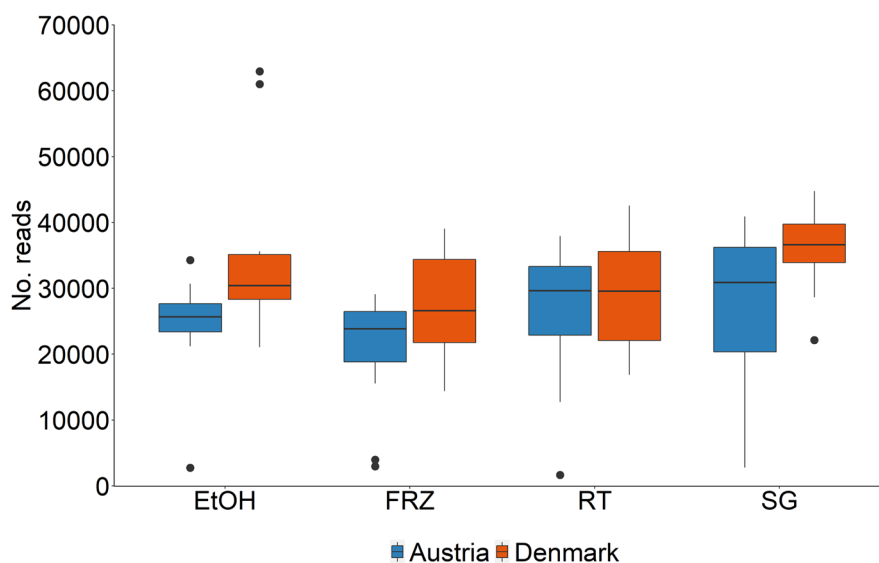
Classification at a lower taxonomic level identified 25 plant species (24 genera) in Austria and 33 (24 genera) in Denmark (Online Resource 8). This is a notable number of taxa, given the temporally limited sampling undertaken herein, with pollen traps activated < 12 h at the single sampling event in August, in Denmark, and at any of the three sampling events in September, in Austria. Yet, this level of richness is supported by reports that honey bees tend to collect greater diversity of pollen later in the season, to compensate for lower availability of mass-blooming sources (Danner et al., 2017; Rasmussen et al., 2021). Despite the high number of visited plants, most pollen was collected from only three species in Austria and four in Denmark, which accounted for 60.2% (14.0) and 54.6% (9.1) of the total abundance in the former and latter country, respectively (median calculated over all samples; Online Resource 8). While our results are consistent with studies that have observed

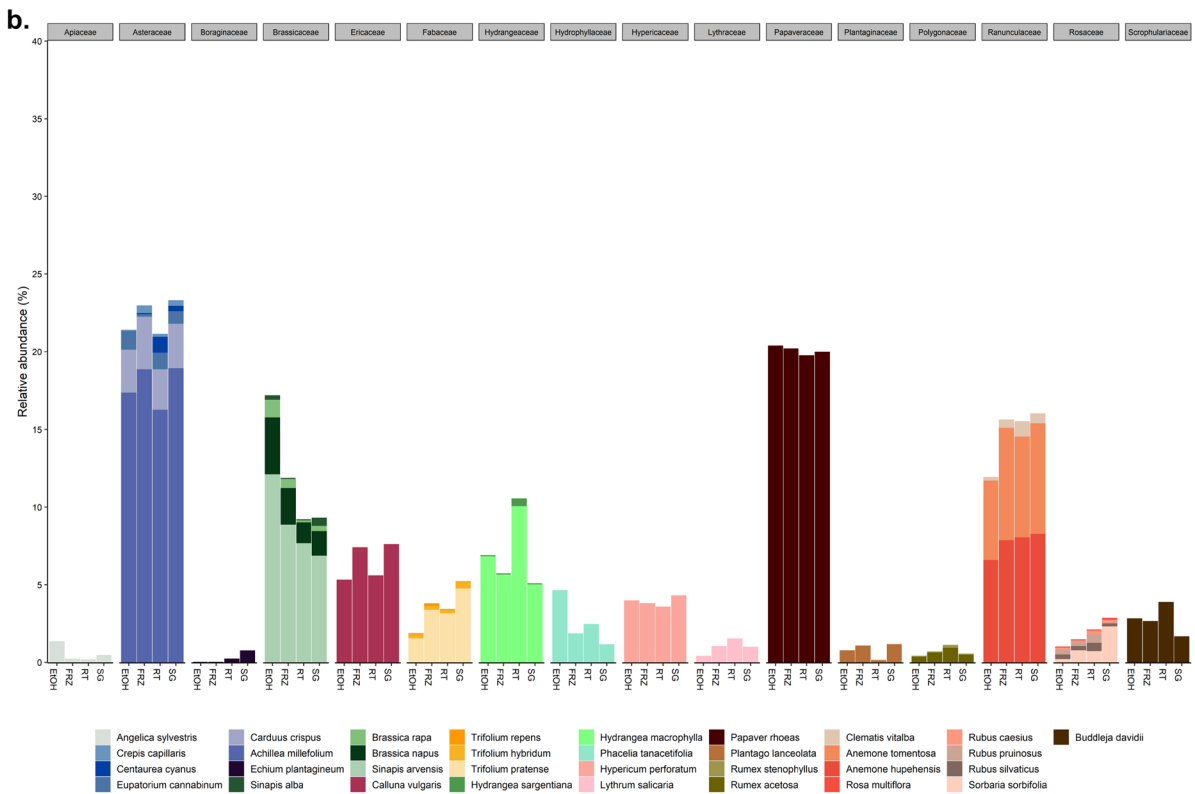
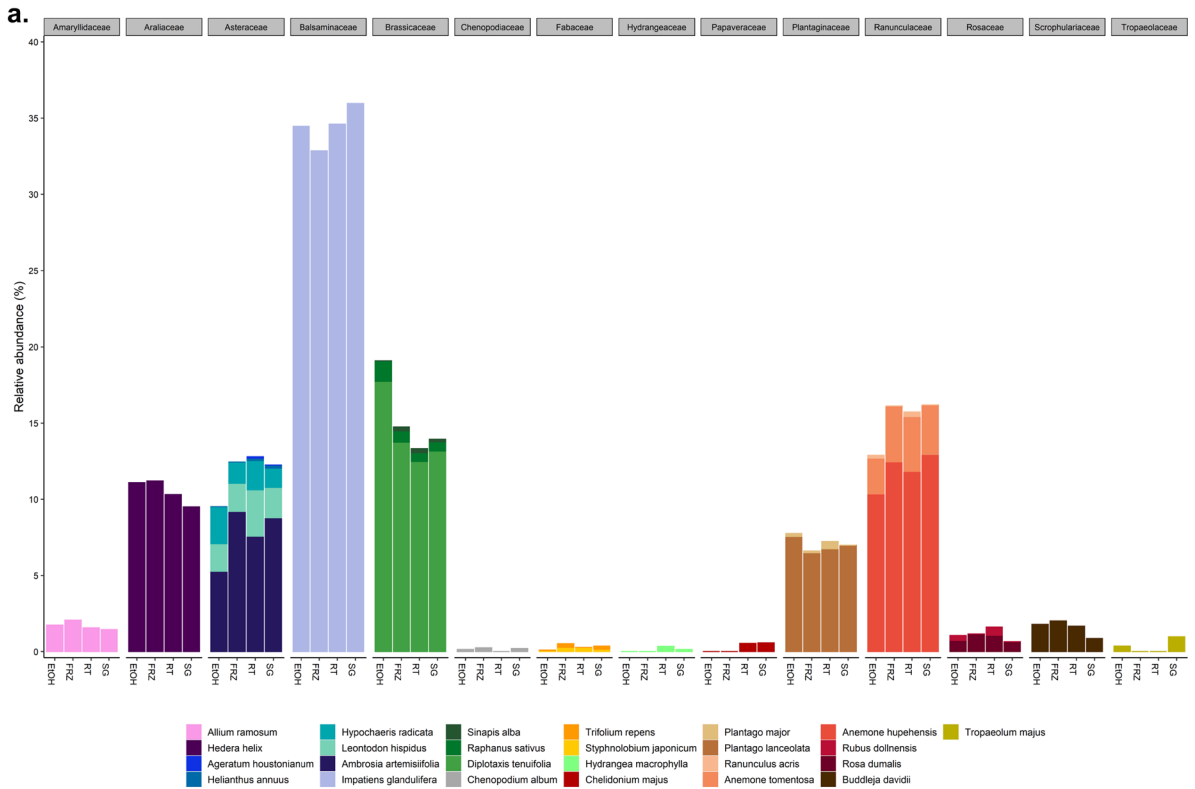
Fig. 6 Relative abundances (%) estimated from 87 pollen samples collected in **a** Austria and **b** Denmark and preserved in ethanol (EtOH; Austria $N=11$, Denmark $N=10$), frozen at $-20\text{ }^{\circ}\text{C}$ (FRZ; Austria $N=11$, Denmark $N=10$), at room temperature (RT; Austria $N=12$, Denmark $N=10$) and with 12 g of silica gel (SG; Austria $N=13$, Denmark $N=10$). Relative abundances, shown here for species and families, were inferred from sequence reads obtained by ITS2 metabarcoding. In Austria, *Chenopodiaceae* was only detected in 05/09/20, *Hydrangeaceae* in 10/09/20, and *Papaveraceae* and *Tropaeolaceae* in 15/09/20 samples

foraging preferences for a few plant sources (Bilisik et al., 2008; Brodschneider et al., 2019; Danner et al., 2017), they also further support the claim that honey bees forage for a diverse pollen diet to ensure colony health (Alaux et al., 2010; Di Pasquale et al., 2013, 2016; Goulson et al., 2015; Kaluza et al., 2018; Omar et al., 2017).

Among the 22 families, 19 (11 in Austria and 15 in Denmark; Figs. 6 and 7) were detected across all methods whereas three very rare families [median of 0% (0), Online Resources 5 and 7] were detected only in pollen preserved in (1) EtOH and SG (*Tropaeolaceae*), (2) SG and RT (*Boraginaceae*), and (3) EtOH, SG and FRZ (*Chenopodiaceae*). The same pattern was observed at the species level; of the 51 species, 17 occurring at very low abundances [median of 0% (0), Online Resources 6 and 8] could only be identified in samples preserved by one (four species), two (seven species) or three (five species)

Fig. 5 Boxplots for the total number of high-quality sequence reads obtained for pollen samples stored in ethanol (EtOH; Austria $N=11$, Denmark $N=10$), frozen at $-20\text{ }^{\circ}\text{C}$ (FRZ; Austria $N=11$, Denmark $N=10$), at room temperature (RT; Austria $N=12$, Denmark $N=10$) and with 12 g of silica gel (SG; Austria $N=13$, Denmark $N=10$)





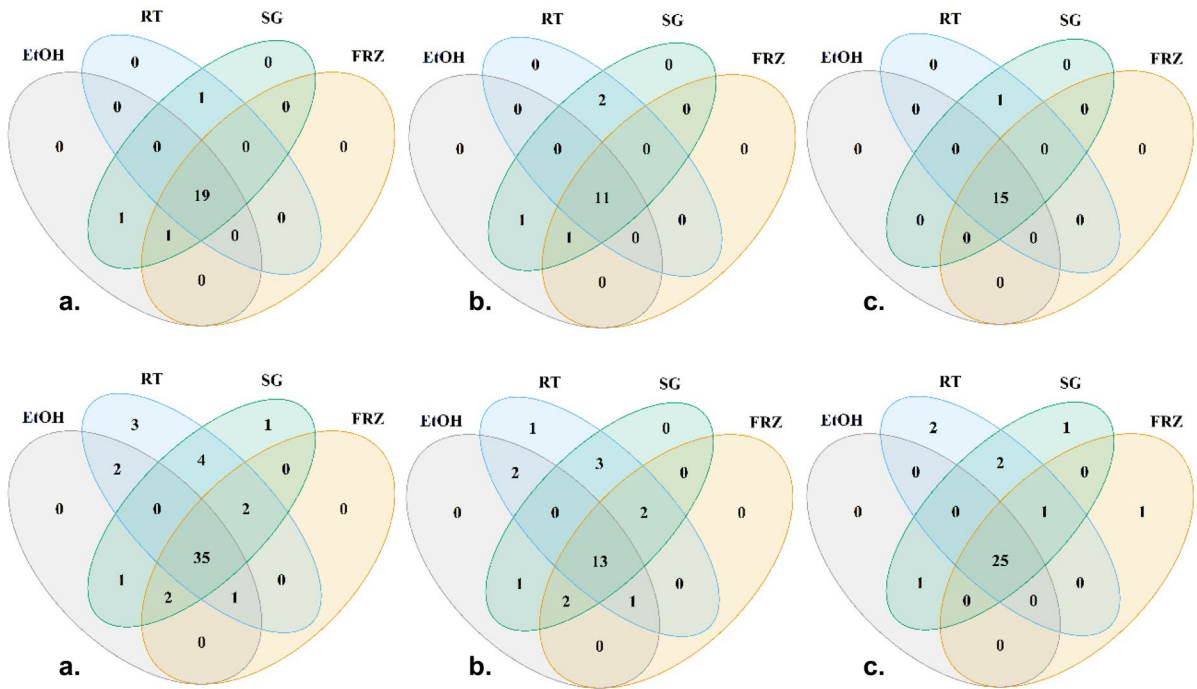


Fig. 7 Venn diagram representing the number of families (top) and species (bottom) shared between preservation methods for **a** total samples, **b** Austrian samples and **c** Danish samples

methods (Fig. 7a; Online Resource 8). Notably, the highest richness was detected in SG (46 species) and RT (47 species) samples whereas the lowest was detected in FRZ (40 species) and EtOH (41 species) samples, although the differences were not statistically significant ($P=0.474$ for Austria and $P=0.744$ for Denmark, χ^2 test). Eight low-abundance plant species, exclusively detected in RT and/or SG replicates, were major contributors to richness differences among preservation methods (Online Resource 6). Four species were singletons in samples preserved at RT (*Ageratum houstonianum*, *Hydrangea sargentiana* and *Rumex stenophyllus*) or with SG (*Rosa multiflora*), whereas four other species were simultaneously detected in several RT and SG samples (*Centaurea cyanus*, *Chelidonium majus*, *Echium plantagineum* and *Helianthus annuus*).

At least three mutually inclusive factors could explain the greater species richness detected in samples desiccated at RT and with SG. First, it is possible that these samples were contaminated by airborne pollen and any trace of contamination, in a typically small size pollen sample, is able to generate a misleading result when using high-throughput

sequencing (Bell et al., 2016, 2018; Pornon et al., 2016). Contrary to FRZ and EtOH samples, which were placed in capped vials soon after sampling, preparation of SG samples involved greater manipulation and RT samples were more exposed to airborne pollen. In a metabarcoding study of airborne pollen, Kraaijeveld et al. (2015) identified over 85% of honey bee plant taxa in the samples analysed, suggesting that contamination could have occurred during SG preparation and, to a greater extent, during ambient exposure of RT samples.

Second, it is possible that the bioinformatics pipeline assigned a wrong taxonomy to the sequence reads generated for those replicates, resulting in false-positive species. Such bias may occur when sequences are misidentified in the reference database (Bell et al., 2018). We tested this hypothesis for the sequence reads generated by the MiSeq platform for the eight species detected exclusively in SG and/or RT preserved samples, using the Basic Local Alignment Search Tool (BLAST) available from the NCBI platform. While BLAST confirmed the identity of seven plant species, strikingly, the sequence reads classified by the pipeline as *Rumex stenophyllus* for

one RT replicate aligned with the airborne fungus *Alternaria infectoria*, with 99.38% identity. This finding highlights the importance of having accurately curated sequences in the reference databases used in metabarcoding analysis (Banchi et al., 2020) and suggests a contamination event for that replicate preserved at RT.

Third, and most likely, because samples desiccated at RT and with SG have lower water content than those placed fresh in the freezer, and samples preserved in EtOH are heavier, it is possible that the higher amount of starting material observed for several RT and SG preserved samples during the DNA extraction step allowed rare pollen grains to be represented in the extracts. While studies of pollen metabarcoding have employed variable amounts of starting material during DNA extraction, with values ranging from 3 to 50 mg (Danner et al., 2017; Keller et al., 2015; Richardson et al., 2015a, 2015b; Sickel et al., 2015), it is unclear to what extent this factor influences the number of plant taxa identified in the extracts. Our results suggest that detection of rare taxa might benefit from using greater amounts of starting material, although investigation is necessary for further understanding this factor.

As illustrated in Fig. 6, deviations in relative abundances across preservation methods were greater in Denmark than in Austria. This finding might be due to more efficient blending of the pollen harvest, performed prior to its splitting into replicates or earlier sampling in Denmark. Yet, the results cannot be directly compared between the two countries because floral spectra at the different sampling dates differed substantially, with only a few families and species overlapping, and taxa-specific biases might occur (Bell et al., 2018). It is, however, noteworthy that two shared families, *Brassicaceae* and *Ranunculaceae*, were consistently over- and under-represented, respectively, in EtOH samples from both countries. Whether the relative abundances estimated from EtOH samples for these two families are more accurate than those obtained for the other methods deserves further scrutiny.

Among the top six families, the greatest differences in median relative abundances were found in *Brassicaceae* comparing between EtOH [16.9% (3.5); Online Resource 7] and SG [9.7% (2.5)] and in *Hydrangeaceae* comparing between RT [10.7% (6.8)] and SG [4.8% (5.4)], with deviations between

the two most dissimilar methods nearly doubling. On the other hand, for the top three families and species, deviations were lower than 3% in both countries. Of these, *Papaveraceae/Papaver rhoeas* displayed the greatest homogeneity across methods, with only 0.5% difference between EtOH [19.5% (1.0)] and SG [19.0% (1.3)] samples. Regardless of the magnitude of the deviations, no statistical differences could be detected among preservation methods for any family or species either for Austria ($P > 0.161$ for families and species, Kruskal–Wallis tests) or Denmark ($P > 0.206$ for families and $P > 0.100$ for species).

Cumulative abundances, depicted in Fig. 8 for families and species, reveal close resemblances in floral spectra among preservation methods in both countries. This finding is further supported by the parameter H' , in which diversity is expressed by combining richness and evenness (Fig. 9). When computed from family data, the median H' ranged from 1.76 (0.11), for SG, to 1.82 (0.14), for RT, in Austria and from 2.10 (0.08), for EtOH, to 2.17 (0.13), for RT, in Denmark (Online Resource 9). A similar pattern was found at the species level, with median H' values ranging from 1.89 (0.10), for SG, to 1.91 (0.10), for RT, in Austria and from 2.10 (0.08), for EtOH, to 2.17 (0.13), for RT, in Denmark (Online Resource 10). However, H' differences among preservation methods were found to be non-significant for both Austria ($P = 0.471$ for families and species, Kruskal–Wallis test) and Denmark ($P = 0.470$ for families and species), suggesting that bee-collected fresh pollen can be adequately preserved for ITS2 metabarcoding applications by any of the methods tested herein.

Ultra-low temperature freezing (e.g. $-80\text{ }^{\circ}\text{C}$) is the gold standard for high-quality tissue archiving of a wide range of samples for use in molecular genetic analyses (Nagy, 2010; Prendini et al., 2002). While long-term tissue preservation at higher temperatures can be compromised by enzymatic activity and DNA degradation, ultra-low temperature freezers are usually not available at citizen scientists' premises and storing at $-20\text{ }^{\circ}\text{C}$ in a household freezer becomes the only alternative. However, here this method was revealed to be effective for pollen, at least for medium-term preservation, and most studies on pollen metabarcoding have typically used material stored at $-20\text{ }^{\circ}\text{C}$ (Bell et al., 2017; Danner et al., 2017; Smart et al., 2016). The problem arises when

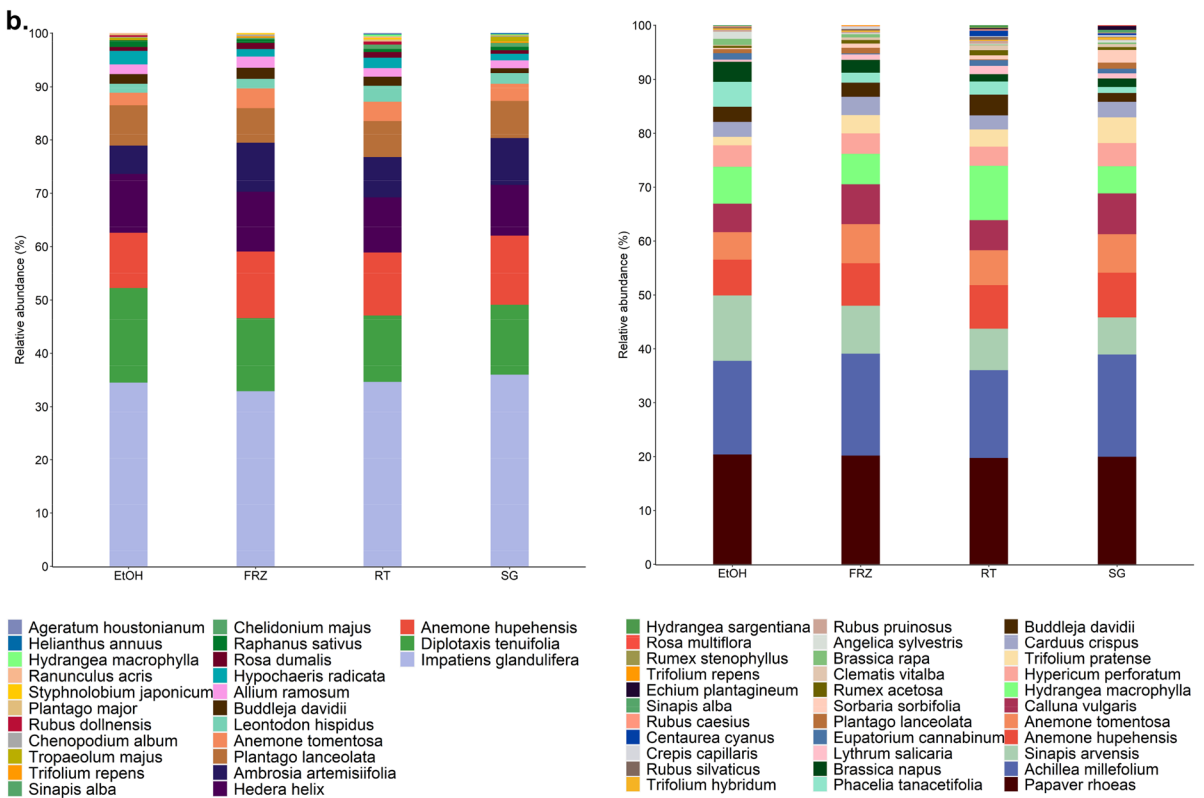
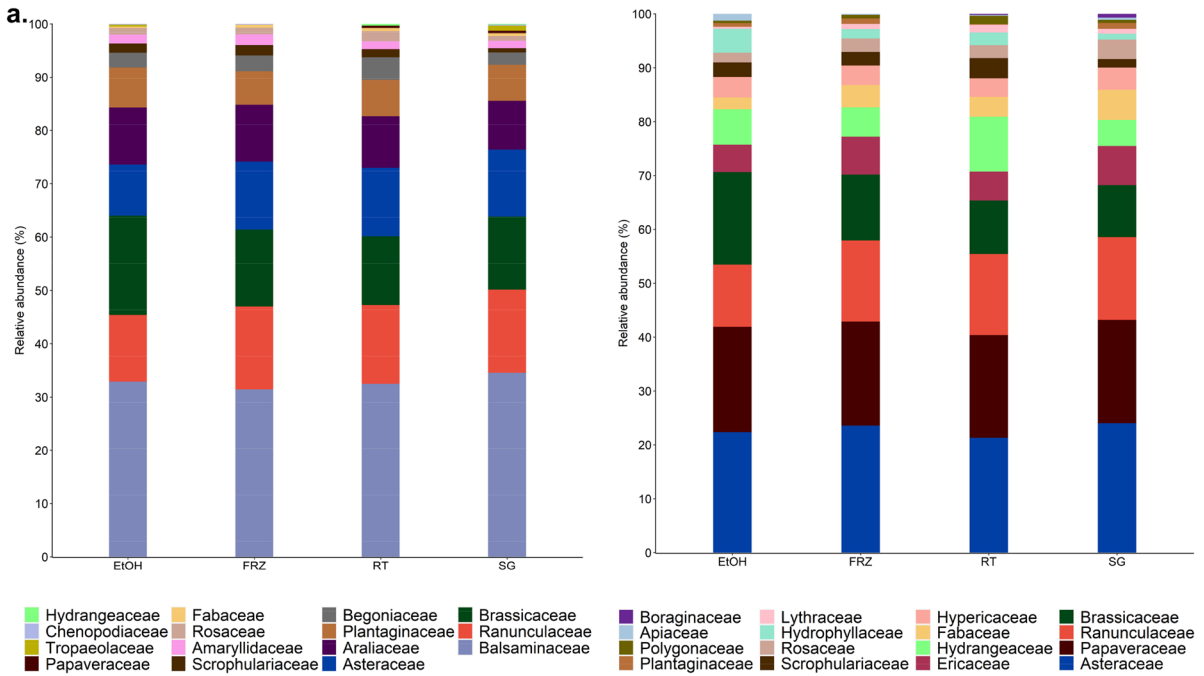
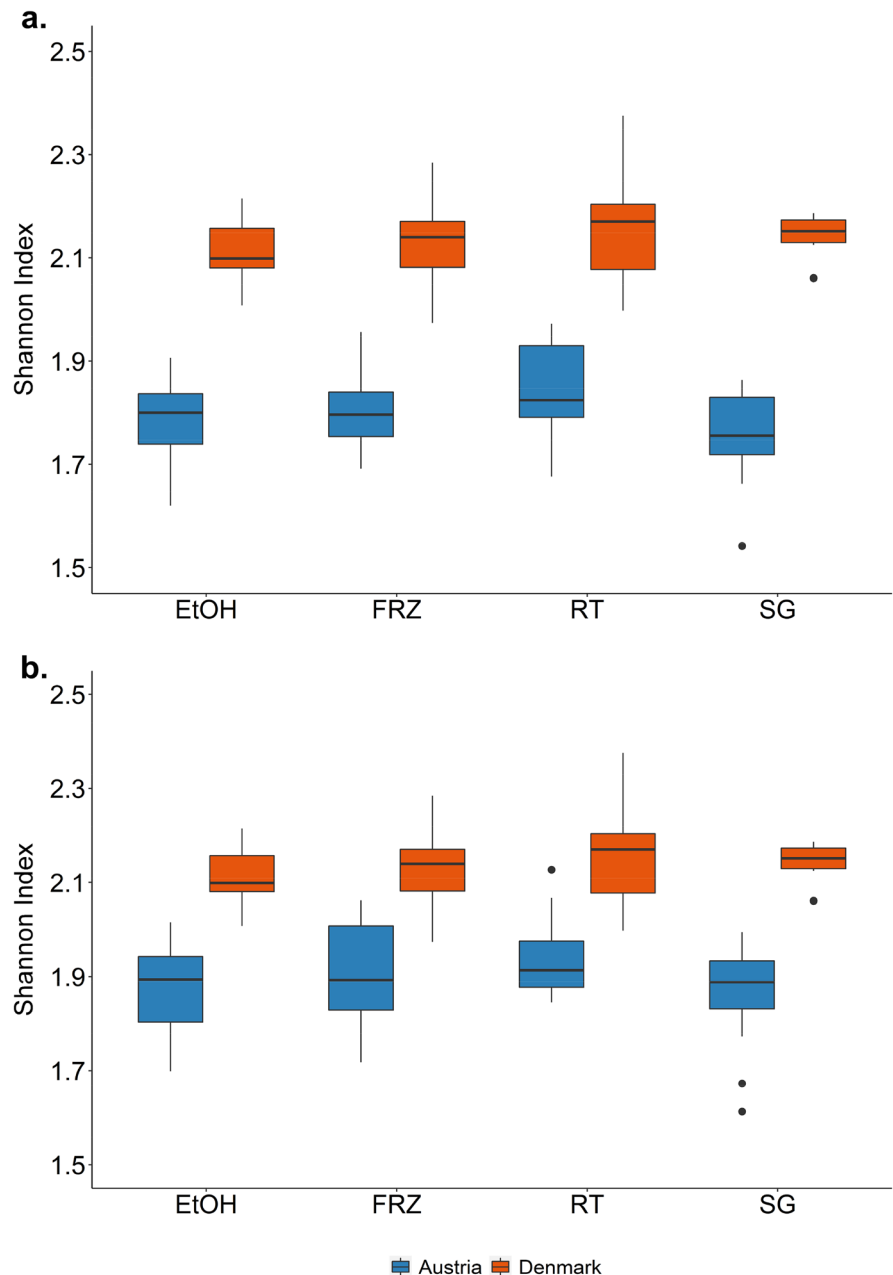


Fig. 8 Cumulative abundances across the four preservation methods. Relative abundances (%) were calculated for **a** families and **b** species from classifying sequence reads obtained by ITS2 metabarcoding. Relative abundances were calculated for pollen samples collected in Austria (left) and Denmark (right) and preserved in ethanol (EtOH; Austria $N=11$, Denmark $N=10$), frozen at $-20\text{ }^{\circ}\text{C}$ (FRZ; Austria $N=11$, Denmark $N=10$), at room temperature (RT; Austria $N=12$, Denmark $N=10$) and with 12 g of silica gel (SG; Austria $N=13$, Denmark $N=10$) until metabarcoding analysis

sample size becomes large, as large numbers of samples will take too much storage space and will involve electricity expenses for the citizen scientist. Besides, to avoid a cycle of freezing–thawing–freezing, which can be harmful to sample integrity, transportation to the analytical laboratory requires a cold chain, further increasing sample handling costs.

Fig. 9 Boxplot for Shannon index (H') estimated for **a** families and **b** species from the 87 pollen samples collected in Austria and Denmark and preserved in ethanol (EtOH; Austria $N=11$, Denmark $N=10$), frozen at $-20\text{ }^{\circ}\text{C}$ (FRZ; Austria $N=11$, Denmark $N=10$), at room temperature (RT; Austria $N=12$, Denmark $N=10$) and with 12 g of silica gel (SG; Austria $N=13$, Denmark $N=10$) until metabarcoding analysis



Ethanol is one of the most popular methods for storing animal tissue, allowing long-term preservation especially when samples are placed at $-20\text{ }^{\circ}\text{C}$ (Nagy, 2010). Although EtOH has not been commonly used for archiving plant tissues such as leaves, shoots and seeds (reviewed by Bressan et al., 2014; Murray & Pitas, 1996; Nagy, 2010; Prendini et al., 2002), here it is revealed to be adequate for preserving pollen for downstream metabarcoding analysis. However, this method might not be affordable in citizen science projects because of costs of obtaining ethanol or difficulty for a layman to obtain it, and, more importantly, because (national or international) transportation of flammable or hazardous fluids is strictly regulated, further increasing shipping costs and burden of paper work requirements.

Of the four methods tested here, desiccation at RT has no costs and does not require any material or equipment. However, while this study indicates that pollen stored at RT provides quality data comparable to the other preservation methods, there is a risk of sample contamination that cannot be overlooked. The finding of a *Lepidoptera* larva feeding on a RT sample is a living example of such risk. Moreover, environmental temperature and humidity may vary across seasons and regions, influencing the efficiency and velocity of the desiccation process, which may compromise sample stability.

Finally, SG offers the best solution for pollen desiccation and storage, emerging as the most promising method for pollen preservation in citizen science studies. When dispensed in adequate silica-to-sample ratios (here 12 g:5 g), it provides effective drying required for archival storage (Alsos et al., 2020; Chase & Hills, 1991), although silica beads should be monitored regularly for dryness when samples are kept at room temperature. The use of commercially available 6-g sachets makes use of SG straightforward for citizen scientists. Besides, SG is inexpensive and shipping to the analytical laboratory is less costly than with the freezing and ethanol methods because dry material is lighter and no special regulations or conditions are applicable.

Conclusion

While pollen metabarcoding studies have typically worked with samples preserved at $-20\text{ }^{\circ}\text{C}$, this is the first study to examine how different preservation methods affect molecular identification of mixed pollen samples by high-throughput sequencing. Overall, the results obtained in this experiment suggest that the methods involving desiccation, which are cheaper than use of ethanol and freezing, can be used by the citizen scientist for medium-term pollen storage for downstream applications involving DNA metabarcoding. Given that relative humidity at room temperature may vary temporally and geographically, and one sample dried at room temperature in this study was lost due to moth infestation, we recommend using silica gel for preserving bee-collected fresh mixed pollen samples. The method is also straightforward for laymen to use in practice, and therefore it is a robust option for widespread use in citizen science studies involving collection of pollen.

Acknowledgements We are deeply indebted to Susana Lopes and Maria Magalhães, from CIBIO—Research Centre in Biodiversity and Genetic Resources—InBIO Associate Laboratory, for their time devoted to library preparation and sequencing in the MiSeq. AQ acknowledges the PhD scholarship (DFA/BD/5155/2020) funded by FCT.

Author contribution MAP, JvdS, RB, OK and FV conceived the ideas and designed the methodology. AQ performed the laboratorial metabarcoding work as well as all the analyses with assistance of MAP and AG. FV and OK performed the preliminary silica testing. Pollen sampling and sample preservation were carried out by RB, KG, OK and FV. AK and JR assembled the reference ITS2 database and assisted with bioinformatics. JvdS acquired funding. MAP and AQ wrote the original draft. All the authors contributed with data interpretation. All the authors critically revised and edited the manuscript for important intellectual content.

Funding This work was funded by the Health and Food Safety Directorate General, European Commission through the project INSIGNIA—Environmental monitoring of pesticide use through honeybees SANTE/E4/SI2.788418-SI2.788452-INSIGNIA-PP-1-1-2018. Fundação para a Ciência e a Tecnologia (FCT) provided financial support by national funds (FCT/MCTES) to CIMO (UIDB/00690/2020).

Data availability All source data analysed is available in the supplemental material.

Declarations

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>

Alsos, I. G., Lavergne, S., Merkel, M. K., Boleda, M., Lammers, Y., Alberti, A., Pouchon, C., Denoeud, F., Pitelkova, I., Puşcaş, M., Roquet, C., Hurdu, B.-I., Thuiller, W., Zimmermann, N. E., Hollingsworth, P. M., & Coissac, E. (2020). The treasure vault can be opened: large-scale genome skimming works well using herbarium and silica gel dried material. *Plants*, 9(4). <https://doi.org/10.3390/plants9040432>

Avni, D., Hendriksma, H. P., Dag, A., Uni, Z., & Shafir, S. (2014). Nutritional aspects of honey bee-collected pollen and constraints on colony development in the eastern Mediterranean. *Journal of Insect Physiology*, 69, 65–73. <https://doi.org/10.1016/j.jinsphys.2014.07.001>

Banchi, E., Ametrano, C. G., Greco, S., Stanković, D., Muggia, L., & Pallavicini, A. (2020). PLANiTS: a curated sequence reference dataset for plant ITS DNA metabarcoding. *Database*, 2020. <https://doi.org/10.1093/database/baz155>

Beekman, M., & Ratnieks, F. L. W. (2000). Long-range foraging by the honey-bee, *Apis mellifera* L. *Functional Ecology*, 14, 490–496. <https://doi.org/10.1046/j.1365-2435.2000.00443.x>

Bell, K. L., Burgess, K. S., Botsch, J. C., Dobbs, E. K., Read, T. D., & Brosi, B. J. (2018). Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. *Molecular Ecology*, 28(2), 431–455. <https://doi.org/10.1111/mec.14840>

Bell, K. L., De Vere, N., Keller, A., Richardson, R. T., Gous, A., Burgess, K. S., & Brosi, B. J. (2016). Pollen DNA barcoding: Current applications and future prospects. *Genome*, 59, 629–640. <https://doi.org/10.1139/gen-2015-0200>

Bell, K. L., Fowler, J., Burgess, K. S., Dobbs, E. K., Gruenewald, D., Lawley, B., Morozumi, C., & Brosi, B. J. (2017). Applying pollen DNA metabarcoding to the study of plant–pollinator interactions. *Applications in Plant Sciences*, 5, 1600124. <https://doi.org/10.3732/apps.1600124>

Beuchat, L. R. (1983). Influence of water activity on growth, metabolic activities and survival of yeasts and molds. *Journal of Food Protection*, 46(2), 135–141. <https://doi.org/10.4315/0362-028X-46.2.135>

Bilisik, A., Cakmak, I., Bicakci, A., & Malyer, H. (2008). Seasonal variation of collected pollen loads of honeybees (*Apis mellifera* L. anatoliaca). *Grana*, 47(1), 70–77. <https://doi.org/10.1080/00173130801923976>

Bressan, E. A., Rossi, M. L., Gerald, L. T. S., & Figueira, A. (2014). Extraction of high-quality DNA from ethanol-preserved

tropical plant tissues. *BMC Research Notes*, 7(1), 268. <https://doi.org/10.1186/1756-0500-7-268>

Brodtschneider, R., & Crailsheim, K. (2010). Nutrition and health in honey bees. *Apidologie*, 41(3), 278–294. <https://doi.org/10.1051/apido/2010012>

Brodtschneider, R., Gratzter, K., Carreck, N. L., Vejsnaes, F., & van der Steen, J. (2021). INSIGNIA: beekeepers as citizen scientists investigate the environment of their honey bees. *Proceedings of Science*. <https://doi.org/10.22323/1.393.0019>

Brodtschneider, R., Gratzter, K., Kalcher-Sommersguter, E., Heigl, H., Auer, W., Moosbeckhofer, R., & Crailsheim, K. (2019). A citizen science supported study on seasonal diversity and monoflorality of pollen collected by honey bees in Austria. *Scientific Reports*, 9, 16633. <https://doi.org/10.1038/s41598-019-53016-5>

Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., Reris, R. A., Sheth, N. U., Huang, B., Girerd, P., Strauss, J. F., Jefferson, K. K., & Buck, G. A. (2015). The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiology*, 15(1), 66. <https://doi.org/10.1186/s12866-015-0351-6>

Canale, A., Benelli, G., Castagna, A., Sgherri, C., Poli, P., Serra, A., Mele, M., Ranieri, A., Signorini, F., Bientinesi, M., & Nicolella, C. (2016). Microwave-assisted drying for the conservation of honeybee pollen. *Materials*, 9, 363. <https://doi.org/10.3390/ma9050363>

Chase, M. W., & Hills, H. H. (1991). Silica gel: An ideal material for field preservation of leaf samples for DNA studies. *Taxon*, 40(2), 215–220. <https://doi.org/10.2307/1222975>

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y., & Leon, C. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE*, 5(1), e8613. <https://doi.org/10.1371/journal.pone.0008613>

Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., & Zhou, S. (2016). Barcoding the kingdom Plantae: New PCR primers for ITS regions of plants with improved universality and specificity. *Molecular Ecology Resources*, 16(1), 138–149. <https://doi.org/10.1111/1755-0998.12438>

Coffey, M. F., & Breen, J. (1997). Seasonal variation in pollen and nectar sources of honey bees in Ireland. *Journal of Apicultural Research*, 36(2), 63–76. <https://doi.org/10.1080/00218839.1997.11100932>

Cornman, R. S., Otto, C. R. V., Iwanowicz, D., & Pettis, J. S. (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., Keller, A., Härtel, S., & Steffan-Dewenter, I. (2017). Honey bee foraging ecology: Season but not landscape diversity shapes the amount and diversity of collected pollen. *PLoS ONE*, 12(8), e0183716. <https://doi.org/10.1371/journal.pone.0183716>

Danner, N., Molitor, A. M., 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>
- Delavaux, C. S., Bever, J. D., Karppinen, E. M., & Bainard, L. D. (2020). Keeping it cool: Soil sample cold pack storage and DNA shipment up to 1 month does not impact metabarcoding results. *Ecology and Evolution*, 10(11), 4652–4664. <https://doi.org/10.1002/ece3.6219>
- Di Pasquale, G., Alaux, C., Le Conte, Y., Odoux, J.-F., Pioz, M., Vaissière, B. E., Belzunces, L. P., & 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>
- Di Pasquale, G., Salignon, M., Le Conte, Y., Belzunces, L. P., Decourtye, A., Kretzschmar, A., Suchail, S., Brunet, J.-L., & 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>
- Dimou, M., & Thrasylvoulou, A. (2007). Seasonal variation in vegetation and pollen collected by honeybees in Thessaloniki. *Greece. Grana*, 46(4), 292–299. <https://doi.org/10.1080/00173130701760718>
- Donkersley, P., Rhodes, G., Pickup, R. W., Jones, K. C., & Wilson, K. (2014). Honeybee nutrition is linked to landscape composition. *Ecology and Evolution*, 4(21), 4195–4206. <https://doi.org/10.1002/ece3.1293>
- Doyle, J. J., & Dickson, E. E. (1987). Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon*, 36(4), 715–722. <https://doi.org/10.2307/1221122>
- Drummond, F. A., Ballman, E. S., Eitzer, B. D., Du Clos, B., & Dill, J. (2018). Exposure of honey bee (*Apis mellifera* L.) colonies to pesticides in pollen, a statewide assessment in Maine. *Environmental Entomology*, 47(2), 378–387. <https://doi.org/10.1093/ee/nvy023>
- Dunker, S., Motivans, E., Rakosy, D., Boho, D., Mäder, P., Hornick, T., & Knight, T. M. (2021). Pollen analysis using multispectral imaging flow cytometry and deep learning. *New Phytologist*, 229(1), 593–606. <https://doi.org/10.1111/nph.16882>
- Feinstein, L. M., Sul, W. J., & Blackwood, C. B. (2009). Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Applied and Environmental Microbiology*, 75(16), 5428–5433. <https://doi.org/10.1128/AEM.00120-09>
- Garbuzov, M., Couvillon, M. J., Schürch, R., & Ratnieks, F. L. W. (2015). Honey bee dance decoding and pollen-load analysis show limited foraging on spring-flowering oil-seed rape, a potential source of neonicotinoid contamination. *Agriculture, Ecosystems & Environment*, 203, 62–68. <https://doi.org/10.1016/j.agee.2014.12.009>
- Goulson, D., Nicholls, E., Botías, C., & Rotheray, E. L. (2015). Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science*, 347(6229), 1255957. <https://doi.org/10.1126/science.1255957>
- Gratzer, K., & Brodschneider, R. (2021). How and why beekeepers participate in the INSIGNIA citizen science honey bee environmental monitoring project. *Environmental Science and Pollution Research*. <https://doi.org/10.1007/s11356-021-13379-7>
- Hawkins, J., de Vere, N., Griffith, A., Ford, C. R., Allainguillaume, J., Hegarty, M. J., Baillie, L., & Adams-Groom, B. (2015). Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. *PLoS ONE*, 10(8), e0134735. <https://doi.org/10.1371/journal.pone.0134735>
- Herbert, E. W., & Shimanuki, H. (1978). Chemical composition and nutritive value of bee-collected and bee-stored pollen. *Apidologie*, 9, 33–40. <https://doi.org/10.1051/apido:19780103>
- Jones, L., Brennan, G. L., Lowe, A., Creer, S., Ford, C. R., & De Vere, N. (2021). Shifts in honeybee foraging reveal historical changes in floral resources. *Communications Biology*, 4(1), 1–10. <https://doi.org/10.1038/s42003-020-01562-4>
- Kaluza, B. F., Wallace, H. M., Heard, T. A., Minden, V., Klein, A., & Leonhardt, S. D. (2018). Social bees are fitter in more biodiverse environments. *Scientific Reports*, 8(1), 12353. <https://doi.org/10.1038/s41598-018-30126-0>
- Kandlikar, G. S., Gold, Z. J., Cowen, M. C., Meyer, R. S., Freise, A. C., Kraft, N. J., Moberg-Parker, J., Sprague, J., Kushner, D. J., & Curd, E. E. (2018). Ranacapa: an R package and shiny web app to explore environmental DNA data with exploratory statistics and interactive visualizations. *F1000Research*, 7, 1734. <https://doi.org/10.12688/f1000research.16680.1>
- Keller, A., Danner, N., Grimmer, G., Ankenbrand, M., 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, 558–566. <https://doi.org/10.1111/plb.12251>
- Kindt, R., & Coe, R. (2005). *Tree diversity analysis: a manual and software for common statistical methods for ecological and biodiversity studies*. Nairobi: World Agroforestry Centre (ICRAF).
- Koffler, S., Barbiéri, C., Ghilardi-Lopes, N. P., Leocadio, J. N., Albertini, B., Francoy, T. M., & Saraiva, A. M. (2021). A buzz for sustainability and conservation: The growing potential of citizen science studies on bees. *Sustainability*, 13(2), 959. <https://doi.org/10.3390/su13020959>
- Kraaijeveld, K., De Weger, L. A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P. S., & Den Dunnen, J. T. (2015). Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Molecular Ecology Resources*, 15, 8–16. <https://doi.org/10.1111/1755-0998.12288>
- Lau, P., Bryant, V., Ellis, J. D., Huang, Z. Y., Sullivan, J., Schmehl, D. R., Cabrera, A. R., & Rangel, J. (2019). Seasonal variation of pollen collected by honey bees (*Apis mellifera*) in developed areas across four regions in the United States. *PLoS ONE*, 14(6), e0217294. <https://doi.org/10.1371/journal.pone.0217294>
- Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J., & Burridge, C. P. (2020). A practical guide to DNA metabarcoding for entomological ecologists. *Ecological Entomology*, 45(3), 373–385. <https://doi.org/10.1111/een.12831>
- Macgregor, C. J., Kitson, J. J. N., Fox, R., Hahn, C., Lunt, D. H., Pocock, M. J. O., & Evans, D. M. (2019). Construction, validation, and application of nocturnal pollen transport networks in an agro-ecosystem: A comparison using light microscopy and DNA metabarcoding. *Ecological Entomology*, 44(1), 17–29. <https://doi.org/10.1111/een.12674>

- McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, 8, e61217. <https://doi.org/10.1371/journal.pone.0061217>
- Milla, L., Sniderman, K., Lines, R., Mousavi-Derazmahalleh, M., & Encinas-Viso, F. (2021). Pollen DNA metabarcoding identifies regional provenance and high plant diversity in Australian honey. *Ecology and Evolution*, 11, 8683–8698. <https://doi.org/10.1002/ece3.7679>
- Miller-Rushing, A., Primack, R., & Bonney, R. (2012). The history of public participation in ecological research. *Frontiers in Ecology and the Environment*, 10(6), 285–290. <https://doi.org/10.1890/110278>
- Moro, A., Beaurepaire, A., Dall’Olio, R., Rogenstein, S., Blacquièrre, T., Dahle, B., de Miranda, J. R., Dietemann, V., Locke, B., Licón Luna, R. M., Le Conte, Y., & Neumann, P. (2021). Using citizen science to scout honey bee colonies that naturally survive *Varroa destructor* infestations. *Insects*, 12(6). <https://doi.org/10.3390/insects12060536>
- Murray, M. G., & Pitas, J. W. (1996). Plant DNA from alcohol-preserved samples. *Plant Molecular Biology Reporter*, 14(3), 261–265. <https://doi.org/10.1007/BF02671661>
- Nagy, Z. T. (2010). A hands-on overview of tissue preservation methods for molecular genetic analyses. *Organisms Diversity & Evolution*, 10(1), 91–105. <https://doi.org/10.1007/s13127-010-0012-4>
- Nürnbergger, 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>
- Ogle, D. H., Doll, J. D., Wheeler, P., & Dinno, A. (2021). FSA: Fisheries Stock Analysis. *R package version 0.9.0.9000*, 636. <https://github.com/droglenc/FSA>
- Omar, A., Kadesjö, N., Palmgren, C., Marteinsdottir, M., Segerdahl, T., & Fransson, A. (2017). Assessment of the occupational eye lens dose for clinical staff in interventional radiology, cardiology and neuroradiology. *Journal of Radiological Protection*, 37(1), 145–159. <https://doi.org/10.1088/1361-6498/aa559c>
- Pacini, E., & Hesse, M. (2005). Pollenkitt – its composition, forms and functions. *Flora - Morphology, Distribution, Functional Ecology of Plants*, 200(5), 399–415. <https://doi.org/10.1016/j.flora.2005.02.006>
- Pawluczyk, M., Weiss, J., Links, M. G., Egeña Aranguren, M., Wilkinson, M. D., & Egea-Cortines, M. (2015). Quantitative evaluation of bias in PCR amplification and next-generation sequencing derived from metabarcoding samples. *Analytical and Bioanalytical Chemistry*, 407(7), 1841–1848. <https://doi.org/10.1007/s00216-014-8435-y>
- Piñol, J., Senar, M. A., & Symondson, W. O. C. (2019). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology*, 28(2), 407–419. <https://doi.org/10.1111/mec.14776>
- Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: Diet assessment using next generation sequencing. *Molecular Ecology*, 21(8), 1931–1950. <https://doi.org/10.1111/j.1365-294X.2011.05403.x>
- Pornon, A., Escaravage, N., Burrus, M., Holota, H., Khimoun, A., Mariette, J., Pellizzari, C., Iribar, A., Etienne, R., Taberlet, P., Vidal, M., Winterton, P., Zinger, L., & Andalo, C. (2016). Using metabarcoding to reveal and quantify plant-pollinator interactions. *Scientific Reports*, 6(1), 27282. <https://doi.org/10.1038/srep27282>
- Potter, C., de Vere, N., Jones, L. E., Ford, C. R., Hegarty, M. J., Hodder, K. H., Diaz, A., & Franklin, E. L. (2019). Pollen metabarcoding reveals broad and species-specific resource use by urban bees. *PeerJ*, 7, e5999. <https://doi.org/10.7717/peerj.5999>
- Prendini, L., Hanner, R., & DeSalle, R. (2002). Obtaining, storing and archiving specimens and tissue samples for use in molecular studies. In R. DeSalle, G. Giribet, & W. Wheeler (Eds.), *Techniques in Molecular Systematics and Evolution* (pp. 176–248). Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-8125-8_11
- Rasmussen, C., Dupont, Y. L., Madsen, H. B., Bogusch, P., Goulson, D., Herbertsson, L., Maia, K. P., Nielsen, A., Olesen, J. M., Potts, S. G., Roberts, S. P. M., Sydenham, M. A. K., & Kryger, P. (2021). Evaluating competition for forage plants between honey bees and wild bees in Denmark. *PLoS ONE*, 16(4), e0250056. <https://doi.org/10.1371/journal.pone.0250056>
- Requier, F., Odoux, J.-F., Tamic, T., Moreau, N., Henry, M., Decourtye, A., & Bretagnolle, V. (2015). Honey bee diet in intensive farmland habitats reveals an unexpectedly high flower richness and a major role of weeds. *Ecological Applications*, 25(4), 881–890. <https://doi.org/10.1890/14-1011.1>
- Richardson, R. T., Bengtsson-Palme, J., & Johnson, R. M. (2017). Evaluating and optimizing the performance of software commonly used for the taxonomic classification of DNA metabarcoding sequence data. *Molecular Ecology Resources*, 17(4), 760–769. <https://doi.org/10.1111/1755-0998.12628>
- Richardson, R. T., Eaton, T. D., Lin, C.-H., Cherry, G., Johnson, R. M., & Sponsler, D. B. (2019). Honey bee pollen foraging ecology across an urbanization gradient. *bioRxiv*, 824474. <https://doi.org/10.1101/824474>
- Richardson, R. T., Lin, C.-H., Quijia, J. O., Riusech, N. S., Goodell, K., & Johnson, R. M. (2015a). 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, R. T., Lin, C. H., Sponsler, D. B., Quijia, J. O., Goodell, K., & Johnson, R. M. (2015b). Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. *Applications in Plant Sciences*, 3, 1400066. <https://doi.org/10.3732/apps.1400066>
- 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>
- Rubin, B. E. R., Gibbons, S. M., Kennedy, S., Hampton-Marcell, J., Owens, S., & Gilbert, J. A. (2013). Investigating the impact of storage conditions on microbial community composition in soil samples. *PLoS ONE*, 8(7), e70460. <https://doi.org/10.1371/journal.pone.0070460>

- Schiebelhut, L. M., Abboud, S. S., Gómez Daglio, L. E., Swift, H. F., & Dawson, M. N. (2017). A comparison of DNA extraction methods for high-throughput DNA analyses. *Molecular Ecology Resources*, *17*(4), 721–729. <https://doi.org/10.1111/1755-0998.12620>
- Sickel, W., Ankenbrand, M. J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., Steffan-Dewenter, I., & Keller, A. (2015). Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology*, *15*(1), 1–9. <https://doi.org/10.1186/s12898-015-0051-y>
- Smart, M. D., Cornman, R. S., Iwanowicz, D. D., McDermott-Kubeczko, M., Pettis, J. S., Spivak, M. S., & Otto, C. R. V. (2017). A comparison of honey bee-collected pollen from working agricultural lands using light microscopy and ITS metabarcoding. *Environmental Entomology*, *46*(1), 38–49. <https://doi.org/10.1093/ee/nvw159>
- Smart, M. D., Pettis, J. S., Euliss, N., & Spivak, M. S. (2016). Land use in the Northern Great Plains region of the U.S. 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>
- Swenson, S. J., & Gemeinholzer, B. (2021). Testing the effect of pollen exine rupture on metabarcoding with Illumina sequencing. *PLoS ONE*, *16*(2), e0245611. <https://doi.org/10.1371/journal.pone.0245611>
- Team, R. (2015). *RStudio: integrated development for R*. In RStudio Inc. <http://www.rstudio.com>
- Tosi, S., Costa, C., Vesco, U., Quaglia, G., & Guido, G. (2018). A 3-year survey of Italian honey bee-collected pollen reveals widespread contamination by agricultural pesticides. *Science of the Total Environment*, *615*, 208–218. <https://doi.org/10.1016/j.scitotenv.2017.09.226>
- Weißbecker, C., Buscot, F., & Wubet, T. (2017). Preservation of nucleic acids by freeze-drying for next generation sequencing analyses of soil microbial communities. *Journal of Plant Ecology*, *10*(1), 81–90. <https://doi.org/10.1093/jpe/rtw042>
- White, T. J., Bruns, T., Lee, S. J. W. T., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In G. D. M. A. Innis, J. J. Sninsky, & T. J. White (Eds.), *PCR protocols: a guide to methods applications* (pp. 315–322). Academic Press

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