Protocols for assessing the distribution of pathogens in individual Hymenopteran pollinators

Joachim R de Miranda (joachim.de.miranda@slu.se)
Swedish University of Agricultural Sciences https://orcid.org/0000-0002-0335-0386
Ivan Meeus
University of Gent
Orlando Yañez
University of Bern
Niels Piot
University of Gent
Laura Jara
Universidad de Murcia
Piero Onorati
Swedish University of Agricultural Sciences
Pilar De la Rúa
Universidad de Murcia
Anders Nielsen
5. Norwegian Institute of Bioeconomy Research
Ivana Tlak-Gajger
University of Zagreb
Erkay Özgör
Cyprus International University
Anne Dalmon
Institut National de Recherche pour l'Agriculture, l'alimentation et l'Environnement
Anna Gajda
Warsaw University of Life Sciences
Peter Neumann
University of Bern
Guy Smagghe
University of Gent
Robert J Paxton
Martin-Luther-Universität Halle-Wittenberg

Method Article

Keywords: bee, pathogen, transmission, honeybee, bumblebee, solitary bee, disease, protocol, barcode

Posted Date: April 30th, 2021

DOI: https://doi.org/10.21203/rs.3.pex-1453/v1

Abstract

This protocol was developed for the COST-Action “Super-B”, whose purpose was to coordinate research, outreach and policy towards sustainable pollination[1-3]. The protocol addresses the detection of parasites and pathogens across bee species, as one of several possible drivers of bee decline[4,5]. It consists of four major components:

1. A sample collecting protocol, based around a dominant bee species driving pathogen distribution among other bee species

2. A sample processing protocol, based on analyzing nucleic acids from individual bees

3. A pathogen detection protocol, based on RT-qPCR with broad-range primers for several common pathogens

4. A barcoding protocol, for accurate bee species identification

The protocols have largely been adapted from existing knowledge and protocols but also include two key innovations: the use of passive reference nucleic acids and synthetic positive controls, that significantly improve the quality and robustness of the raw data, and thus the reliability of the analyses and conclusions.

Introduction

The main objective of COST-Action FA1307: Sustainable pollination in Europe - joint research on bees and other pollinators (Super-B) was to coordinate research, outreach and policy towards sustainable pollination services in Europe[1-3]. One of the Super-B working groups, WG4, addressed the possible drivers for the decline in bee numbers and diversity in Europe, one of which was identified as the transmission of parasites and pathogens between bee species[4,5]. The increasing homogenization of the agricultural landscapes across Europe is reflected in the loss of pollinators in these landscapes[6-9], with most of the pollination services provided by a few dominant bee species[6,8]. The key question is whether, and to what extent, such changes in the communities of pollinators are related the distribution and transmission of parasites and pathogens between bee species[10-15]. The protocols do not specifically distinguish between replicating and non-replicating microbial agents, but the material collected allows for such a determination.

Development of the protocol

The sample processing and pathogen detection protocols were developed, adapted, improved and optimized from the BeeBook[16], a collection of protocols and manuals for honeybee research, particularly the chapters on molecular biology[17] and virus research[18]. The use of exogenous reference nucleic acids was first pioneered to correct sample-specific quantification errors at the cDNA stage[19-21], and adapted

here to encompass the whole sample. The effects of different homogenization methods, primary extraction buffers and incubation times on bee nucleic acid integrity and suitability for RT-qPCR-based quantification was explored in several different studies[21-24] and was used here to inform the choices and optimization of sample storage, management, nucleic acid extraction and transport[24]. An additional optional protocol for DNA barcode-based identification of bee species was included[25-27] to help correctly identify cryptic bee species and avoid biases in the data due to differences in field identification skills between collectors.

Applications of the method

Although these protocols are designed around bees and their shared pathosphere[28-31], they can easily be adapted to other groups of organism with common pathogens and distributions. The protocols are based on the detection and quantification of pathogen nucleic acid in a sample, rather than the disease consequences of the presence and loads of these pathogens for individual bee species[18]. However, it is not difficult to include direct or indirect measures of bee health, both through biological analyses of the original samples[32] and more thorough exploration of the extracted nucleic acids, particularly the RNA, since RNA allows analysis of both composition and function[33]. As such, the protocols are flexible and adaptable to different purposes, research questions and technologies.

Comparison with other methods

The sample processing and pathogen quantification protocols are mostly variations on common protocols for nucleic acid extraction and analysis. An alternative to targeted screening of individual pathogens by RT-qPCR is target-free screening, using mass parallel sequencing techniques[28,34,35]. With target-free screening, the entire microbial complexity of either the RNA or the DNA phase is recovered, by mass parallel sequencing of either universal DNA barcode regions[36,37], whole DNA metagenomes[28,38] and/or the non-ribosomal fraction of the RNA[34,39,40]. The identity of each individual sequenced nucleic acid is determined through comparison with the large public databases, after which the composition of the sample is reconstructed from the individual identified reads[28,34,35]. Although this is initially much more costly than targeted screening, the advantage is that all microorganisms are investigated simultaneously with the same assay methodology. This avoids both assay-specific biases, saves on precious nucleic acid and becomes quickly more cost-effective with every new pathogen or assay target included in the screening. Instead of saving biological or nucleic acid samples for future re-screening, this approach generates the sequencing data, which can be re-screened at any time in the future as the databases are updated with new information. The approach furthermore also allows genetic analysis of the sequenced targets, which opens up a whole new level of refinement for epidemiological analyses, host-pathogen interaction studies and pathogen population genetics and evolution. This applies only to the third and final component of the entire protocol (detection) and is entirely compatible with the first two components (collection and processing).

Experimental design

The experimental design resides largely in the sample collection strategy. The design consists of collecting, as they are encountered, thirty bees of the dominant bee species (the putative driver of transmission); thirty bees of all other species/targets of transmission, and fifteen additional specimens of the most common of the target bee species.

Expertise required

All parts of the protocol can be easily implemented by trained students and technicians, with knowledge of basic molecular techniques. Due to the detailed and easy implementation of the protocol, it will be the preferred standard technique for future studies of this nature.

Limitations

The main limitations of the protocol, and source of methodological variation in the data, concern the early stages of the sample processing and nucleic acid extraction, particularly the choices made for sample homogenization and early sample incubations. These affect not only the nucleic acid's extraction efficiency (which can be corrected using the exogenously added reference nucleic acids[19-21]) or its integrity (which can be corrected using internal reference standards[17,18,41,42]), but also its composition[24], which is much more difficult to correct. The best approach for guaranteeing comparability and minimizing bias between individual samples is high standardization in these early steps[24]. Since these limitations involve the very first steps in sample processing, their consequences persist throughout all subsequent stages, irrespective of which alternative approaches are implemented later in the protocol. It is therefore crucial to ensure these early steps are as tight and controlled as possible.

Reagents

Below (Table 1) is a list of *essential* reagents and equipment, *i.e.* those that are critical for the performance of the protocol. Non-essential reagents and equipment, such as collecting vessels, transport options or dissection tools, are given in the procedures with a note defining those features that are important.

Reagents

Table 1. Table of essential reagents, the suppliers, their websites and the catalog numbers of the reagents.

Reagent Set-up

1. Sterile TBS buffer: 50 mM TRIS.HCl pH7.4, 150 mM NaCl

2. TBS/RNA250/pJET buffer: On the day of use, add to TBS buffer (see above) RNA250 to a final concentration of 10 ng/mL and pJET1.2 to a final concentration of 1 ng/mL, based on the concentrations in the product data sheets.

3. Lysozyme Lysis buffer: 20 mM TRIS.HCl pH8.0, 2 mM EDTA, 1.2% Triton X100, 20 mg/mL Lysozyme

4. Positive qPCR controls: Prepare an 8-step, 10-fold dilution series of each positive control, diluting in TE8.0 buffer (10 mM TRIS.HCl pH.8.0, 1 mM EDTA)

Equipment

Equipoment

Table 2. Table of essential equipment, the suppliers, their websites and the catalog numbers of the equipment.

Equipment Set-up

1. Bead-mill (TissueLyser II) homogenization: Add three 3 mm steel beads and one 5 mm steel bead to the sample and buffer in an appropriate tube for bead-milling. Shake for 2 minutes at 30 Hz followed by 2 minutes at 20 Hz.

2. ThermoCyclers: Set up thermocycling profiles using the appropriate software for each thermocycler.

Procedure

The full procedure is divided into four distinct protocols, concerning (A) the collection of the bee samples; (B) the processing and management of the samples; (C) the pathogen and parasite assays, and (D) the genetic identification of the bee species. Many different types of data are collected throughout the procedure, some for use in data analysis and others as intermediate stages in the conversion and normalization of the raw data into processed data that is suitable for use in statistical analyses. A metafile describing in detail each of these primary and derived parameters is given as a Supplementary Table at the end of the procedure.

A. SAMPLE COLLECTION PROTOCOL

The field sampling protocol below is based on honey bees (*Apis mellifera*) as a driver of pathogen distribution in wild bees, but this can be substituted for any other biological driver, *e.g.* a particular species of bumblebee. The speed at which the bees are collected, recorded by time, can be used as a proxy for the absolute density of bees in the area. The order in which different bee species are collected

(up to their maximum) can be used as a proxy for the relative density of the different bee species. Bee foraging behavior is affected by weather conditions[43] and species of flower[44-46], and differs furthermore between bee species[44-49]. Since bee pathogens can be transmitted through floral visitation networks[11-15], and are also directly affected by temperature[50], weather conditions and the floral character of the landscape are important metadata to collect.

1. Sample collection:

a) Collect 30 individual honey bees (*Apis mellifera*) and 30 individual wild bees, in the order in which they are encountered

b) Collect 15 additional specimens of the commonest wild bee species

c) Collect all samples on the same day during high flight activity of bees (above 15°C and less than 80% cloud cover) in spring/summer (April-August in the Northern Hemisphere).

d) Collect all bees within the same 100 m x 100 m flower-rich area from flowers

e) If a particular bee species is at a high density on one patch of flowers, collect a maximum of five individuals from that patch then move to collect from other patches within your single 100 m x 100 m area

f) Place each insect individually into a tube, labelled in the order in which the bees were collected. Place the tube on ice.

g) nStore tubes the same evening in a -80 °C freezer for downstream molecular analysis

2. Data collection:

For each bee, record the following:

a) Time of collection

b) Species of bee (field ID, to be confirmed afterwards by microscopy or barcoding)

c) Gender of the bee (can be determined afterwards)

d) The plant species whose flower the bee was on (can be determined afterwards)

e) Temperature in the open; temperature in the shade

f) Cloud cover

g) Longitude of the location; latitude of the location

h) Surrounding landscape type (can be determined afterwards)

i) General floral density and diversity within the 100 m x 100 m area

B. SAMPLE PROCESSING PROTOCOL

The sample processing is designed to allow the following two aims:

To determine the prevalence of known (honey)bee pathogens across bee species

To identify and characterize yet unknown, novel pathogens in wild bees

The strategy for both aspects is to prepare a primary homogenate in a neutral buffer, extract nucleic acids from a small amount of extract for the “current-pathogen” analyses (aim 1) and retain the rest for “new-pathogen” prospecting analyses (aim 2), which may involve additional steps prior to nucleic acid extraction.

There is strong emphasis in the sample collection protocol on minimizing cross-contamination between insects, in order to avoid misclassifying the pathogen status of individual bees. Aside from sampling artefact, the larger purpose of this is to attempt to distinguish between those microbial agents that are infectious to the bee tissues, and are therefore a potential health threat; those that are part of the bee microbiome, both internal and external, and those that are passively associated with the bee but not infectious to bee tissues. We have therefore also included several simple processing and assay strategies that maximize our ability to distinguish between infectious and non-infectious agents. One means to do this is to separate the body parts of the bees:

Abdomen, where nearly all internal bee pathogens replicate, in the bee tissues, and shed their replicative propagules (spores/oocysts/virus particles *etc.*) in the gut lumen for voiding into the environment, with the faeces. The abdomen also contains passively acquired, non-infectious agents, both internally in the gut and externally on the exoskeleton.

Head, where many of the viral pathogens actively replicate, and occasionally shed particles into the salivary and hypopharyngeal glands, especially in honey bees.

Thorax, containing few internal (or external) microbes.

Wings & Legs, containing pollen baskets. Pathogens can be shared between bees through flower-visitor networks, especially on pollen.

Since the abdomen contains by far the highest concentrations of pathogens, there is little loss of detection sensitivity from excluding the head, thorax and legs-wings, which can be retained separately for other studies. For example, the DNA/RNA from legs, wings and thoraxes contain very little contaminating microbial nucleic acids and are therefore optimally suited for possible host bee genetic analyses. In order to allow for the widest possible types of additional, future analyses on the material, it is best to prepare the primary homogenate in a neutral, aqueous buffer and only extract RNA and DNA from an aliquot of primary homogenate, with the remainder stored at -80 $^{\circ}$C. There is no loss in detection sensitivity of the pathogens (or host mRNAs) from such a neutral primary extract as long as the extract is either frozen or added to the nucleic acid extraction buffers within 5 minutes. There are also several options at the assay level to distinguish active and passive infections, *vide infra*. Such measures are not entirely fool proof, but greatly improve the chances of finding truly infecting pathogens.

The protocols described below are loosely based on the COLOSS BeeBook chapters “Standard methods for virus research in *Apis mellifera*”[18] “Standard methods for research on *Apis mellifera* gut symbionts”[37] and “Standard methods for molecular research in *Apis mellifera*”[17].

3. Bee size estimation:

a) Remove from the freezer the number of bees needed for a single extraction run

b) Record the weight of each bee on a fine scale (in mg)

c) Using a fine caliper, record the inter-tegular distance (ITD) of each bee, *i.e.* the distance across the thorax between the insertion points of the wings (to the nearest 0.1 mm)

4. Dissection:

a) Perform the dissections on a frozen (carbo-ice) dissection plate, if possible

b) Separate **H**ead, **T**horax, **W**ings/legs from the **A**bdomen using a sterile scalpel/razor blade and place each in a separate marked container

c) The scalpel can be sterilised between bees by dipping in ethanol and burning off the ethanol

with a flame

d) Store the H, T, W and A sub-samples at -80 °C until further use

5. Buffer preparation:

a) Sterile TBS buffer: 50 mM TRIS.HCl pH7.4, 150 mM NaCl

b) Lysozyme Lysis buffer: 20 mM TRIS.HCl pH8.0, 2 mM EDTA, 1.2% Triton X100, 20 mg/mL Lysozyme

c) TBS/RNA250/pJET buffer: On the day of use, add to TBS buffer (see above) RNA250 to a final concentration of 10 ng/mL and pJET1.2 to a final concentration of 1 ng/mL, based on the concentrations given in the product sheets, as passive RNA and DNA reference standards, for posterior normalization of the data for methodological differences between individual samples in homogenization and nucleic acid extraction efficiency

d) The amount of TBS/RNA250/pJET buffer needed for an individual bee abdomen is:

Bumble bee sized abdomen	800 µL TBS/RNA250/pJET
Honey bee sized abdomen	500 µL TBS/RNA250/pJET
Small sweat bee sized abdomen	200 µL TBS/RNA250/pJET

e) Orchard and mason bees would be similar in size to honey bees; others more like sweat bees, so use your judgement and record which actual volume you used for each bee

6. Homogenization:

a) ***CRITICAL:*** work rapidly and at 4 °C until the nucleic acids are extracted and frozen

b) Homogenize each abdomen in the TBS-RNA250/pJET buffer:

If a bead-mill (*e.g.* the Qiagen TissueLyser II or equivalent) is available, add three 3 mm steel beads and one 5 mm steel bead to the sample and buffer in an appropriate tube for bead-

milling, and shake for 2 minutes at 30 Hz followed by 2 minutes at 20 Hz

If no bead-mill is available, use disposable micro pestles to grind the abdomen by hand in the TBS-RNA250 buffer

c) Centrifuge briefly to pellet the exoskeleton and transfer the supernatant homogenate to a clean 2 mL screwcap storage tube and store immediately at -80˚C

d) Repeat until enough bees have been processed for a single extraction run

7. RNA & DNA extraction:

a) The RNA extraction follows the Qiagen Plant RNeasy protocol

b) The DNA extraction follows the Qiagen Blood and Tissue DNA kit protocol for Gram-positive bacteria, which includes *Paenibacillus* (AFB), *Melissococcus* (EFB), the lactic acid bacteria and many other bee gut bacteria

c) The main feature of the DNA protocol is a long lysozyme/proteinase-K incubation prior to DNA purification, to digest the bacterial cell walls and the spores, oocysts and propagules of bacteria, microsporidia, and various other bee eukaryotic pathogens

d) RNA and DNA extractions can be performed in parallel, in batches of 12

e) The DNA samples will be incubating for about 1 hour while you purify the RNA samples, after which you can purify the DNA samples

f) For the **RNA** extraction, prepare 12 microcentrifuge tubes with 350 µL RLT buffer (from the Qiagen Plant RNeasy purification kit) to which 1% beta-mercaptoethanol has been added on the day of use.

g) For the **DNA** extraction, prepare 12 microcentrifuge tubes with 180 µL Lysozyme Lysis buffer (see above; do NOT use the buffer provided in the kit!)

h) Thaw enough primary homogenates for a single run of RNA and DNA extractions (usually 12) and

store on ice (4 ^{o}C)

i) ***CRITICAL:*** Work as quickly as possible while the homogenates are thawed, and return immediately to frozen storage (< -20 ^{o}C) when no longer needed.

j) Briefly mix each homogenate and add 100 µL to a tube for RNA extraction (RLT buffer) and 100 µL to a tube for DNA extraction (Lysozyme Lysis buffer)

k) ***CRITICAL:*** Freeze the remaining homogenate immediately after mixing the two 100 µL aliquots to the RNA and DNA extraction buffers

l) Proceed to the next homogenate, until all homogenates in the extraction run have aliquoted 100 µL each to an RNA and a DNA extraction tube

8. DNA purification:

a) Incubate the DNA extractions for 30 minutes at 37˚C

b) Add 25 µL Proteinase-K solution (supplied in the Blood and Tissue kit) and mix

c) Add 200 µL buffer AL (excluding ethanol; supplied in the Blood and Tissue kit) and mix

d) Incubate 30 minutes 56˚C

e) Add 200 µL ethanol and mix thoroughly

f) Transfer all onto the DNeasy Mini spin column nested in a 2 ml collection tube

g) Follow the DNeasy protocol, either manually or with a robot

h) Elute the DNA in either 100 µL (bumblebee), 70 µL (honeybee) or 50 µL (sweat bee) AE buffer (supplied in the Blood and Tissue kit)

i) Determine the approximate yield and purity of the DNA, using a NanoDrop or similar instrument

j) Adjust the DNA concentrations of all samples to 20 ng/µL by adding AE buffer (supplied in the Blood and Tissue kit)

k) Store the DNA at -80˚C until further use

9. RNA purification:

a) Meanwhile, for the RNA extractions, follow the Plant RNeasy protocol, either manually or with a robot, while the DNA extractions are incubating

b) Elute the RNA in either 50 or 30 µL ultra-pure sterile water, depending on the size of the bee

c) Determine the approximate yield and purity of the RNA using a NanoDrop or similar instrument

d) Adjust the RNA concentrations of all samples to 100 ng/µL by adding sterile water (supplied in the RNA extraction kit)

e) Store the RNA at -80 °C until further use

10. cDNA synthesis:

a) Prepare first strand cDNA with random hexamer primers from 1 µg RNA (10 µL) in a 20 µL volume using a standard first strand cDNA kit containing the M-MLV reverse transcriptase and an RNAse inhibitor, following the manufacturers recommendations

b) After the final incubation, dilute the cDNA 10-fold in sterile water and store at -80°C until further use.

C. PATHOGEN ASSAYING PROTOCOL

The presence and abundance of a range of bee microorganisms, whether detrimental or beneficial, is determined by quantitative PCR of the cDNA and DNA templates using broad-range primers (*i.e.* those encompassing several strains or species within a complex). The reason for this is in part efficiency (fewer assays to run) and in part to avoid false negative results due to assay insufficiencies (Type-II errors). Type-I errors (false positive results) are mostly due to trace contamination and are easily avoided by restricting the number of amplification cycles to 35, rather than 40[17,42,51] For accurate quantification we recommend buying synthetic external quantification standards (*e.g.* ThermoFisher), based on the product sequences (Supplementary Information 3), rather than home-made standards from either purified plasmid clones of the fragment, or purified PCR product, primarily to guarantee uniformity of standards and quantification between different labs, as well as better absolute quantification with synthetically

produced, and accurately quantified, standards. A key element of these quantitative assays is the use of the passive external reference nucleic acids, RNA250 and pJET1.2. These were added in exact known quantities at the start of the homogenization and nucleic acid extraction protocol and the amounts of these left in each cDNA and DNA template can be measured exactly through qPCR, similar to how the pathogen amounts are measured. The ratio of RNA250/pJET measured by qPCR in a standard template volume to the known amount originally added prior to extraction is therefore a simple, one-step conversion factor for all the individual, sample-specific methodological errors and losses incurred from homogenization through qPCR, for accurate absolute quantification of the amounts of each target in the original bee. It will reduce, if not eliminate, random methodological noise from the dataset and improve the chances of detecting true biological differences between the samples. Another key feature is that the qPCR assays are all designed to work with the same thermocycling profile, to enable different assays to be run in the same thermocycling run, and produce similar medium-sized PCR products, to facilitate distinguishing true product from illegitimate secondary products through either Melting Curve analysis or agarose gel electrophoresis.

11. qPCR amplification:

a) Add 2 µL diluted DNA or cDNA template to 18 µL qPCR reaction mixture (*e.g.* BioRad EvaGreen) containing 0.2 µM each of forward and reverse primers for the assay being run (Supplementary Table 1)

b) Include in each run a 7-step ten-fold dilution series of a (preferably synthetic) positive control of known concentration, as well as a template-free negative control (*e.g.* ddH_2O)

c) Amplify the target with the manufacturer's recommended thermocycling profile for a ~500 bp fragment, a fluorescent dye-based detection chemistry (*e.g.* SYBR-green, EvaGreen), for primers with a melting temperature of 58 $^{\circ}$C, and for no more than 35 cycles, *e.g.* 95 $^{\circ}$C:30sec + 35x[95 $^{\circ}$C:15sec – 58 $^{\circ}$C:20sec – 72 $^{\circ}$C:30sec - read] + 72 $^{\circ}$C:60sec

d) Following the amplification stage with a Melting Curve (MC) analysis profile for confirming product identity, *e.g.* reading the fluorescence at 0.5 $^{\circ}$C intervals from 55 $^{\circ}$C to 95 $^{\circ}$C

e) Confirm the absence of amplification in the template-free negative control

f) Discard all positive results whose MC profile and peak does not match to within 1.0 $^{\circ}$C of the MC

profiles and peaks for the positive controls

12. Absolute quantification and passive reference normalization:

a) Obtain for each sample-assay combination the SQ (Starting Quantity) value, estimated by plotting the Cq values against the known amounts (preferably as '**copy number**') of the synthetic external reference standards

b) Calculate the total number of **copies** of RNA250 and pJET1.2 molecules added to each abdomen prior to extraction, from the volume of TBS-RNA250-pJET buffer added and the molecular weights of RNA250 (641991 g/mol) and pJET1.2 (1837503 g/mol)

c) For each sample, calculate the RNA250out/RNA250in ratio by dividing the sample's individual SQ value for its RNA250 amplification (RNA250out) by the calculated amount of RNA250 added to the abdomen prior to extraction (RNA250in)

d) Divide each sample's SQ values for the different RNA pathogens (cDNA amplifications) by the sample's RNA250out/RNA250in ratio, to obtain the estimated total amount of pathogen RNA in the original abdomen

e) Follow the same procedure for the pJETout/pJETin ratios and the DNA pathogen amplifications

D. BARCODE SPECIES IDENTIFICATION PROTOCOL

Some bees may be difficult to identify in the field, either because they are cryptic (*i.e.* morphological features that overlap with other bee species), dirty or damaged. Occasionally a species may lack a morphological identification key. In these cases, it is usually possible to identify specimen through DNA barcode analysis, *i.e.* by sequencing one of several well-established barcoding genes in the bee nuclear or mitochondrial genomes. Most animals, particularly insects, are barcoded using a *c.a.* 650 bp fragment of the mitochondrial Cytochrome Oxidase I (*cox1*) gene, which is sufficiently variable to be able to uniquely distinguish even closely related species[25,26]. A critical requirement is that the DNA template is sufficiently pure to only amplify the *cox1* region of the bee, and not of any contaminating DNA from eukaryotic parasites or plants. This effectively rules out using the purified DNA from bee abdomens for use in barcoding analysis, since this DNA consists of the genomes of all organisms residing in the bee gut, as

well as some of the bee. The simplest approach is to amplify the *cox1* directly from a snippet of leg, wing or thorax tissue[27]. Each bee cell contains many mitochondria, each of which contains numerous copies of the mitochondrial genome[52,53], so that even the tiniest trace amount of tissue contains thousands of copies of the mitochondrial genome, more than enough for PCR amplification. In fact, there is greater danger that too much template is added to the PCR reaction, which can inhibit the PCR and the production of sufficient PCR product for sequencing[51].

13. *Cox1* barcode amplification:

a) Aliquot 10 µL PCR master mix containing 0.2 mM dNTP and 0.4 µM each of primers LCO-1490 (GGTCAACAAATCATAAAGATATTGG) and HCO-2198 (TAAACTTCAGGGTGACCAAAAAATCA)

b) Add with sterile forceps a snippet of a bee wing or a leg tarsa, or poke a disposable 10 µL pipette tip in the bee thorax and stir in the buffer

c) Enough DNA will transfer for amplification

d) Amplify the *cox1* gene with the manufacturer's recommended thermocycling profile for a ~650 bp fragment, for primers with a melting temperature of 51 °C, and for no more than 35 cycles, *e.g.* 95 °C:30sec + 35x[95 °C:15sec – 51 °C:15sec – 72 °C:30sec] + 72 °C:60sec

e) Run 1 µL PCR product on an agarose gel to check the product size and absence of secondary products

f) Submit the PCR product for sequencing by any of a number of commercial sequencing companies

g) Most bees can be barcoded with these primers but some species are difficult to amplifying at the *cox1*barcode with these primers[54]

h) If the *cox1* barcode of a particular bee does not amplify with these primers[54], then try an alternative forward primer, LCO-1790 (GCTTTCCCACGAATAAAATAATA), which can be used[26] together with HCO-2198 (TAAACTTCAGGGTGACCAAAAAATCA) primer

i) Amplify the shorter *cox1*product using the manufacturer's recommended PCR protocol for a ~400 bp fragment, for primers with a melting temperature of 54 °C, and for no more than 35 cycles, *e.g.* 95 °C:30sec + 35x[95 °C:15sec – 54 °C:15sec – 72 °C:30sec] + 72 °C:60sec

14. *Cox1* barcode sequencing & analysis:

a) ***OPTIONAL:*** To ascertain whether the amplification has been successful or not, run 1 µL of each PCR reaction (or a representative selection of PCR products) on a 0.8% agarose gel in a suitable TRIS/EDTA

buffer system, including a nucleic acid binding dye, as recommended by a standard molecular laboratory manual[55], and visualize the bands alongside a molecular weight marker.

b) If the *cox1* barcode amplifications have been successful, submit the reactions to a commercial sequencing service together with the corresponding LCO and HCO amplification primers for sample purification and Sanger sequencing of both strands of each DNA template

c) Check the quality of the sequence produced (usually included in the sequencing results)

d) Combine the forward and reverse sequences into a single consensus sequence for each sample, using any of a number of public or proprietary sequence analyses programmes (*e.g.* SeqScanner, 4peaks, Lasergene, Geneious)

e) Submit your consensus sequence to the BOLD database (www.boldsystems.org) for species identification of your sample

f) If you have a shorter sequence, search for a possible match in the GenBank database (www.ncbi.nlm.nih.gov) using the BLASTn programme

Troubleshooting

By and large, these protocols are very robust and free from major unaccounted sources of variation. The main sources of risk are outlined in Table 3. Those steps where the most methodological variability is introduced, *i.e.* the homogenization and nucleic acid extraction steps in protocol B[24], are largely accounted for through the use of exogenous reference nucleic acids and individual conversion factors for each sample based on these reference nucleic acids. The main form of troubleshooting concerns Type-I and Type-II errors, *i.e.* false positive or false negative results. These are generally discovered during the assaying protocol, even though they may arise during earlier stages. The solution to these is generally to discard the results, or the contaminated nucleic acid samples, and repeat the entire protocol from Protocol B onwards. There are checks and controls throughout the protocol, such as determining the nucleic acid concentrations (steps 8i and 9c) and the inclusion of positive and negative controls during PCR (steps 11b, 11e, 11f) that enable the detection of a problem, and the subsequent action required. These are standard troubleshooting controls and checks that are also included in the various commercial kits used in the protocols, and so are automatically included in the use of these kits.

Time Taken

By far the most time consuming part of the overall protocol, per individual bee, is protocol B (Sample Processing) and in particular steps 3 through 6, when each bee is individually weighed, measured, dissected and homogenized. The next most time consuming section is the production of DNA, RNA and cDNA template for assaying; steps 7-10. If extraction robots are used then part of this time can be allocated to other tasks. The least time consuming part of the protocol, per individual bee/pathogen

combination, is protocol C (Pathogen Assaying). Below is a breakdown of the amount of time that should be budgeted for each part of the overall protocol, summarized in Table 4.

11. Protocol A – Sample Collection

The sample collection normally takes about 3-4 hours, but this depends very much on the circumstances and environmental conditions when the collection takes place, some of which may be beyond our control due to the nature and design of the experiment. The primary way this protocol can be made more efficient, if the experiment allows this, is to plan the collection around when and where the bees are most abundant, active and accessible for collection:

a) Middle-late summer (relatively high pollinator-flower ratio and suitable weather for most bees)

b) Late morning-early afternoon (most active time for flowers and bees)

c) Dry weather with little wind, not too warm or too cold (best conditions for bee flight)

d) Gardens, parks, meadows, hedgerows, field edges, forest clearings (local concentration of bee-attractive flowers, attracting bees from the area)

e) Collecting in teams of two (one for catching, identification and storing; one for recording data)

f) Pre-collection training in bee and plant identification (faster identification and data collection)

12. Protocol B – Sample Processing

This is the most time-consuming part of the entire procedure due to the great amount of attention given to each individual bee: weighing, measuring, dissecting, homogenizing and finally extraction of two types of nucleic acid in separate sub-protocols. These protocols are based around the numerical standards used in molecular biology, based around the 96-well (8 x 12) assay plate. The most convenient approach is to work in batches of 12 bees, both to minimize the amount of time the bees, extracts and nucleic acids are thawed (and thus in a sensitive stage for degradation) while still ensuring a reasonable processing rate for efficiency and optimum use of labour and resources. Under optimal conditions, a single technician can process a batch of 12 bees from step 7 to 14 in about 3-4 hours, which means about 24 bees per day between morning and afternoon. This includes the following time-saving efficiencies:

a) Work in batches

b) Prepare all buffers and tubes beforehand

c) Combine the weighing, measuring, dissecting and barcode sampling of each bee

d) Synchronize the DNA incubation-extraction and RNA extraction-cDNA synthesis steps

13. Protocol C – Pathogen Assaying

A single run of pathogen assays for 88 bees takes around 4 hours, of which about 1 hour is preparation and 3 hours incubation. However, if several thermocyclers are available, this process can be made much more efficient by synchronizing the preparation and incubation steps of several different pathogen assay runs. This means that usually 4 runs can be made in a single day. The total time for this step is therefore mostly determined by how many pathogens are assayed. Since the DNA, RNA and cDNA templates are sufficient for hundreds of assays, this part of the protocol can be as long or short as desired. Time-saving efficiencies can be made by:

a) Synchronizing consecutive assaying runs

b) Assaying the same set of templates (DNA or cDNA) for multiple pathogens on the same day, preventing excessive freeze-thaw cycles

14. Protocol D – Barcode species identification

The laboratory part of the barcoding protocol is similar to running a single pathogen assay, and takes about 4 hours. If the product sizes are checked on a gel prior to submitting for sequencing (step 18a), then another hour is added to the protocol. The most time-consuming stage is actually waiting for the sequencing results and analyzing the data, which is separate to these protocols. Time can be saved by:

a) Preparing the agarose gel and sample tubes while waiting for the incubation to finish

Anticipated Results

The results of the various sub-protocols are collected in a data spreadsheet, with numerous columns for the various data items associated with each sample. These include purely administrative items, such as the sample ID (and perhaps various sub-IDs) for tracing the samples, as well as dates, times, procedural information and for coordinating the sampling, processing and assaying between participating partners. A second major group is the data collected during processing, for administrative record keeping and help with data processing. A third major group is the data collected during sampling, concerning the location, date, time and other interesting features of where the bees were collected, which can be used in analyses to explain the biological data. Then the final major group is the biological data obtained from each sample, concerning the identity of the bee and the presence and amounts of various pathogens. An example is given in Supplementary Table S1, as a metafile with some typical examples of the type of data obtained and how this is registered.

References

1. Potts, S.G. *et al.* Summary for policymakers of the assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production. Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Bonn, Germany (2016).

2. Rortais, A. *et al.* Risk assessment of pesticides and other stressors in bees: Principles, data gaps and perspectives from the European Food Safety Authority. *Sci. Total Environ.* 587-588, 524-537 (2017).

3. More, S.J., Auteri, D., Rortais, A. & Pagani, S. EFSA is working to protect bees and shape the future of environmental risk assessment. *EFSA J.* 19, e190101 (2021).

4. Brown, M.J. *et al.* A horizon scan of future threats and opportunities for pollinators and pollination. *Peer J.* 4, e2249 (2016).

5. Goulson, D., Nicholls, E., Botias, C. & Rotheray, E.L. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* 347, e1255957 (2015).

6. Bommarco, R., Lundin, O., Smith, H.G. & Rundlöf, M. Drastic historic shifts in bumble-bee community composition in Sweden. *Proc. Biol. Sci.* 279, 309-315 (2012).

7. Theodorou, P. *et al.* Urban areas as hotspots for bees and pollination but not a panacea for all insects. *Nature Commun.* 11, 576 (2020).

8. Carvalheiro, L.G. *et al.* Species richness declines and biotic homogenisation have slowed down for NW-European pollinators and plants. *Ecol. Lett.* 16, 870-878 (2013).

9. Kleijn, D. *et al.* Delivery of crop pollination services is an insufficient argument for wild pollinator conservation. *Nature Commun.* 6, 7414 (2015).

10. Tehel, A., Brown, M.J.F. & Paxton, R.J. Impact of managed honey bee viruses on wild bees. *Curr. Opin. Virol.* 19, 16–22 (2016).

11. Alger, S.A., Burnham, P.A. & Brody, A.K. Flowers as viral hot spots: Honey bees (Apis mellifera) unevenly deposit viruses across plant species. *PLoS One* 14, e0221800 (2019).

12. Figueroa, L.L. *et al.* Bee pathogen transmission dynamics: deposition, persistence and acquisition on flowers. *Proc. R. Soc. B Biol. Sci.* 286, e20190603 (2019).

13. Piot, N., Smagghe, G. & Meeus, I. Network centrality as an indicator for pollinator parasite transmission via flowers. *Insects* 11, e872. (2020).

14. Figueroa, L.L. *et al.* Landscape simplification shapes pathogen prevalence in plant-pollinator networks. *Ecol. Lett.* 23, 1212–1222 (2020).

15. Graystock, P. *et al.* Dominant bee species and floral abundance drive parasite temporal dynamics in plant-pollinator communities. *Nature Ecol. Evol.* 4, 1358–1367 (2020).

16. Dietemann, V., Ellis, J.D. & Neumann, P. The COLOSS BEEBOOK Volume I, Standard methods for *Apis mellifera* research: Introduction. *J. Apic. Res.* 52, DOI: 10.3896/IBRA.1.52.4.23 (2013).

17. Evans, J.D. *et al.* Standard methodologies for molecular research in *Apis mellifera*. In The COLOSS BeeBook, Volume I: standard methods for *Apis mellifera* research (eds. Dietemann, V., Ellis, J.D. & Neumann, P.). *J. Apic. Res.* 52, DOI: 10.3896/IBRA.1.52.4.11 (2013).

18. de Miranda, J.R. et al. Standard methods for virus research in *Apis mellifera*. In In The COLOSS BeeBook, Volume II: standard methods for *Apis mellifera* research (eds. Dietemann, V., Ellis, J.D. & Neumann, P.). *J. Apic. Res.* 52, DOI: 10.3896/IBRA.1.52.4.22 (2013).

19. Smith, R.D., Brown, B., Ikonomi, P. & Schechter, A.N. Exogenous reference RNA for normalization of real-time quantitative PCR. *Biotechniques* 34, 88-91 (2003).

20. Mondet, F., de Miranda, J.R., Kretzschmar, A., Le Conte, Y. & Mercer, A.R. On the front line: Quantitative virus dynamics in honeybee (*Apis mellifera* L.) colonies along a new expansion front of the parasite Varroa destructor. *PLoS Pathog.* 10, e1004323 (2014).

21. Tentcheva, D. *et al.* Comparative analysis of deformed wing virus (DWV) RNA in Apis mellifera and Varroa destructor. *Apidologie* 37: 41-50 (2006).

22. Chen, Y.P., Evans, J.D., Hamilton, M. & Feldlaufer, M. The influence of RNA integrity on the detection of honey bee viruses: molecular assessment of different sample storage methods. *J. Apic. Res.* 46, 81–87 (2007).

23. Dainat, B., Evans, J.D., Chen, Y.P. & Neumann, P. Sampling and RNA quality for diagnosis of honey bee viruses using quantitative PCR. *J. Virol. Methods* 174, 150–152 (2011).

24. Forsgren, E., Locke, B., Semberg, E., Laugen, A. & de Miranda, J.R. Sample preservation, transport and processing strategies for honeybee RNA extraction: Influence on RNA yield, quality, target quantification and data normalization. *J. Virol. Methods* 246, 81-89 (2017).

25. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrigenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 3, 294-299 (1994).

26. Schmidt, S., Schmid-Egger, C., Morinière, J., Haszprunar, G. & Hebert, P.D.N. DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera, Apoidea partim). *Mol. Ecol. Resour.* 15, 985-1000 (2015).

27. Ärfstäd, L. Determining genetic relatedness in honey bees, *Apis mellifera*, using microsatellite analysis. MSc thesis, Uppsala University, Uppsala, Sweden (2012).

28. Regan, T. *et al.* Characterisation of the British honey bee metagenome. *Nature Commun.* 9, e4995 (2018).

29. Schwarz, R.S., Huang, Q. & Evans, J.D. Hologenome theory and the honey bee pathosphere. *Curr. Opin. Insect Sci.* 10, 1-7 (2015).

30. Beaurepaire, A. *et al.* Diversity and global distribution of viruses of the western honey bee, *Apis mellifera*. *Insects* 11, e239 (2020).

31. Yañez, O. *et al.* Bee viruses: Routes of infection in Hymenoptera. *Front. Microbiol.* 11, e943 (2020).

32. Alaux, C. *et al.* A 'Landscape physiology' approach for assessing bee health highlights the benefits of floral landscape enrichment and semi-natural habitats. *Sci. Rep.* 7, e40568 (2017).

33. Doublet, V. *et al.* Unity in defence: honeybee workers exhibit conserved molecular responses to diverse pathogens. *BMC Genomics* 18, e207 (2017). Erratum: *BMC Genomics* 18, e256 (2017).

34. Schoonvaere, K., Smagghe, G., Francis, F., de Graaf, D.C. Study of the metatranscriptome of eight social and solitary wild bee species reveals novel viruses and bee parasites. *Front Microbiol.* 9, e177 (2018).

35. Bartolomé, C. *et al.* Wide diversity of parasites in *Bombus terrestris* (Linnaeus, 1758) revealed by a high-throughput sequencing approach. *Environ. Microbiol.* 23, 478–483 (2021).

36. Keller, A. *et al.* Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. *Plant Biol.* 17, 558-66 (2015).

37. Engel, P. *et al.* Standard methods for research on *Apis mellifera* gut symbionts. In The COLOSS BeeBook, Volume I: standard methods for *Apis mellifera* research (eds. Dietemann, V., Ellis, J.D. & Neumann, P.). *J. Apic. Res.* 52, DOI: 10.3896/IBRA.1.52.4.07 (2013).

38. Huson, D.H. *et al.* MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* 12, e1004957 (2016).

39. Shi, M. *et al.* Redefining the invertebrate RNA virosphere. *Nature* 540, 539–543 (2016).

40. Roberts, J.M.K., Anderson, D.L. & Durr, P.A. Metagenomic analysis of Varroa-free Australian honey bees (*Apis mellifera*) shows a diverse Picornavirales virome. *J. Gen. Virol.* 99, 818–826 (2018).

41. Meeus, I., de Graaf, D.C., Jans, K. & Smagghe, G. Multiplex PCR detection of slowly-evolving trypanosomatids and neogregarines in bumblebees using broad-range primers. *J. Appl. Microbiol.* 109, 107-115 (2010).

42. Bustin, S.A. *et al.* The MIQE Guidelines: minimum information for publication of quantitative Real-Time PCR experiments. *Clinic. Chem.* 55, 611-622 (2009).

43. Gérard, M., Vanderplanck, M., Wood, T. & Michez, D. Global warming and plant-pollinator mismatches. *Emerg. Top Life Sci.* 4, 77-86 (2020).

44. Ogilvie, J.E. & Forrest, J.R. Interactions between bee foraging and floral resource phenology shape bee populations and communities. *Curr. Opin. Insect Sci.* 21, 75-82 (2017).

45. Scheper, J. *et al.* Museum specimens reveal loss of pollen host plants as key factor driving wild bee decline in The Netherlands. *Proc. Natl. Acad. Sci. USA* 111, 17552-17557 (2014).

46. Urbanowicz, C., Muñiz, P.A. & McArt, S.H. Honey bees and wild pollinators differ in their preference for and use of introduced floral resources. *Ecol. Evol.* 10, 6741-6751 (2020).

47. Le Provost, G. *et al.* Grassland-to-crop conversion in agricultural landscapes has lasting impact on the trait diversity of bees. *Landsc. Ecol.* 36, 281-295 (2021).

48. Magrach, A., González-Varo, J.P., Boiffier, M., Vilà, M. & Bartomeus, I. Honeybee spillover reshuffles pollinator diets and affects plant reproductive success. *Nature Ecol. Evol.* 1, 1299-1307 (2017).

49. Gómez-Martínez, C. *et al.* Forest fragmentation modifies the composition of bumblebee communities and modulates their trophic and competitive interactions for pollination. *Sci. Rep.* 10, e10872 (2020).

50. Dalmon, A., Peruzzi, M., Le Conte, Y., Alaux, C. & Pioz, M. Temperature-driven changes in viral loads in the honey bee Apis mellifera. *J. Invertebr. Pathol.* 160, 87-94 (2019).

51. SantaLucia, J. Physical principles and visual-OMP software for optimal PCR design. *Methods Mol. Biol.* 402, 3-34 (2007).

52. Syromyatnikov, M.Y., Gureev, A.P., Vitkalova, I.Y., Starkov, A.A. & Popov, V.N. Unique features of flight muscles mitochondria of honey bees (*Apis mellifera* L.). *Arch. Insect Biochem. Physiol.* 102, e21595 (2019).

53. Mesquita, R.D. *et al.* Cytochrome c Oxidase at full thrust: Regulation and biological consequences to flying insects. *Cells* 10, e470 (2021).

54. Françiso, E. & Arias, M.C. Cytochrome c oxidase I primers for corbiculate bees: DNA barcode and mini-barcode. *Mol. Ecol. Resour.* 13, 844–850 (2013).

55. Sambrook, J. Detection of DNA in agarose gels. In Molecular Cloning: A Laboratory Manual Edn. 3 Vol. 1 (eds. Sambrook., J. & Russell, D.W.), chapter 5.14. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA (2001).

Acknowledgements

Author Contributions Statement

All authors participated in the conception of the project. IM and RJP designed the sample collection strategy. JRM, IM, GS, PN and OY designed the initial sample processing methods. JRM, IM, OY, GS and PN designed the initial pathogen assay methods. RJP and JRM designed the barcoding identification method. All authors contributed to the optimization and validation of the protocols. All authors contributed to the sample collection and processing. JRM, IM, OY, PN, PO, RP, AD performed the pathogen assays. JRM curated the data. JRM wrote the first draft of the manuscript. All authors contributed to editing the manuscript. All authors read and approved the final manuscript.

Competing Interests Statement

The authors declare no competing interest.

Acknowledgements

The authors acknowledge the support of the EU COST-Action FA1307 Sustainable pollination in Europe - joint research on bees and other pollinators (Super-B) for providing the platform to initiate this project, as well as internal funders from each country and institute for financing the sample collection and processing costs. We also would like to thank the technicians and students at the collecting-processing labs and the assaying labs for their invaluable help and competence.

Figures

Reagent	Company	www	Catalog number
RNA250	ThermoFisher	www.thermofisher.com	AM7155
pJET2.1	ThermoFisher	www.thermofisher.com	K1231
Lysozyme	ThermoFisher	www.thermofisher.com	89833
RNeasy Plant Mini Kit	Qiagen	www.qiagen.com	74904
DNeasy Blood & Tissue Kit (50)	Qiagen	www.qiagen.com	69504
First Strand cDNA Synthesis Kit	ThermoFisher	www.thermofisher.com	10774691
Applied Biosystems™ Master Mix PCR Power SYBR™ Green	ThermoFisher	www.thermofisher.com	10310115
Synthetic positive qPCR controls	ThermoFisher	www.thermofisher.com	Check supplier

Figure 1

List of essential reagents for the protocols, together with their suppliers, the company websites and the catalog numbers.

Equipment	Company	www	Catalog number
TissueLyser II	Qiagen	www.qiagen.com	85300
Eppendorf Mastercycler® Nexus SX 1	Eppendorf	www.eppendorf.com	6331000017
StepOne-Plus Real-Time PCR System ®	ThermoFisher	www.thermofisher.com	4376600
CFX Connect Real-Time PCR Detection System	BioRad	www.bio-rad.com	1855201
Nanodrop® spectrophotometer	ThermoFisher	www.thermofisher.com	ND-2000

Figure 2

List of essential equipment for the protocols, together with their suppliers, the company websites and the catalog numbers.

STEP	PROBLEM	POSSIBLE CAUSE	SOLUTION
7-10	No nucleic acid	Faulty purification kit	Repeat with fresh kit
7-10	No nucleic acid	Faulty NanoDrop instrument	Check instrument
11	No exogenous reference standard detected by PCR	Exogenous standards omitted from TBS buffer	Add exogenous reference standards to remaining extract and repeat protocol B & C
11	No exogenous reference standard detected by PCR	Faulty PCR process	Check thermocycler, cycling profile, positive PCR controls, buffers, kits. Repeat with new kits.

Figure 3

Table describing the major risks associate with the protocols, the steps involved, and how to troubleshoot these.

PROTOCOL	STEPS	TOTAL TIME	TIME/BEE/ASSAY
A. Sample Collection	1-2	~4 hours	~3 minutes
B. Sample Processing	3-4	~6 hours	~5 minutes
	7-10	~18 hours	~15 minutes
C. Pathogen Assaying	11-12	~4 hours*	~3 minutes
D. Barcode Analysis	13-14	~5 hours**	~4 minutes
* for a single pathogen assay			**including agarose gel

Figure 4

Table describing the approximate time required for different subprotocols and their steps, both in total and broken down to per individual sample and/or assay.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- deMiranda2020SuperBprotocolsNatureProtocolsTABLES1.pdf
- deMiranda2020SuperBprotocolsNatureProtocolsTABLES2.pdf
- deMiranda2020SuperBprotocolsNatureProtocolsExtendeddata.pdf