Development of Bioassays Based on Nucleic Acid Amplification and Next-Generation Sequencing

by
Dimple Vinayak Chavan

A dissertation submitted to the Department of Biology and Biochemistry,
College of Natural Sciences and Mathematics
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Biochemistry

Chair of Committee: Richard C. Willson

Committee Member: George E. Fox

Committee Member: Chandra Mohan

Committee Member: Frank McKeon

University of Houston
August 2021

ACKNOWLEDGMENTS

I would like to extend my thanks to all who have helped, encouraged, and supported me during my journey as a Ph.D. student at the University of Houston. First and foremost, I would like to express my deepest and sincerest gratitude to my supervisor, Dr. Richard C. Willson, for giving me the opportunity to work with him. His incredible guidance, critical thinking approaches, broad and in-depth knowledge, and professional advice have helped me build my analytical thinking and problem-solving abilities and a solid background towards my future career path.

I also thank my committee members, Dr. George E. Fox, Dr. Chandra Mohan, and Dr. Frank McKeon, for their guidance, encouragement, constructive comments, and recommendations on my research projects.

I consider it an honor to work with Dr. Aniko Sabo and Dr. Jay R T. Adolacion. I thank them for their mentorship and guidance in accomplishing the bioinformatic analysis of the honey project. This project would never have been completed without their support.

I want to thank Dr. Katerina Kourentzi, Dr. Binh Vu, and Dr. Mary Crum, who provided me valuable help and insightful advice on my projects. Thank you for being amazing colleagues and friends and always motivating me with inspiring discussions about research and career.

I want to thank the U.S. Department of Homeland Security research team, the staff, and research committee members of the Borders, Trade, and Immigration (BTI) Institute, led by the University of Houston, for their valuable insights and inputs in concluding the honey project. I would also like to thank members of the Willson lab and other University of

Houston colleagues for their aid in acquiring three hundred honey samples from various countries before and during the COVID-19 pandemic.

I want to thank all my former colleagues in Willson's group for the outstanding working climate. Particularly, I want to thank Heather Goux and Hui Chen, who guided me in my early Ph.D. years when I was new to the lab. I want to thank my fellow lab mates, Suman Nandy, Maxwell Smith, Adheesha Danthanarayana, Atul Goyal, Maede Chabi, Kristen Brosamer, Rongwei Lei, Victoria Hlavinka, and Vijay Maranholkar, who directly and indirectly helped me in my research and for the outstanding assistance whenever I needed help. Thanks for your support, friendship, and all the fun times we shared.

My success would not have been possible without the love and patience of my family. I am indebted to my parents and my in-laws for their love, concern, support, and strength all these years. I want to express my most grateful thanks to my partner in life and science, Ujwal Patil, for his emotional support, encouragement, and motivation to achieve my dream.

ABSTRACT

Adulteration and mis-labeling of honey to mask its true origin have become a global concern. Pollen microscopy, the current gold standard for identifying the geographical origins of honey, is very laborious and requires extensive training. In addition, filtered honey cannot be identified by pollen examination and can be spiked with pollen from a more favorable plant to disguise its origins. We targeted the nuclear ribosomal ITS2 region of plant DNA, which is known to support genus-level discrimination. We purified pollen-free DNA from honey, filtered or centrifuged to remove all pollens, using three different methods: (i) anti-dsDNA antibodies coupled to magnetic particles; (ii) Q Sepharose anion exchanger; and (iii) ceramic hydroxyapatite, Type I. The ITS2 region of the captured pollen-free DNA was PCR-amplified and subjected to next-generation sequencing (NGS). Q Sepharose showed the greatest capacity to capture trace pollen-free DNA and was applied to DNA isolation from two additional honey samples.

Additionally, using pollen DNA barcoding and NGS, we have developed a method to authenticate Manuka honey, a high-value product native to New Zealand. We targeted the nuclear ribosomal ITS2 region of plant DNA of twenty-one different manuka samples. Using our in-house developed bioinformatics pipeline, we have successfully developed an NGS-based quantitative technique for measuring manuka DNA out of total plant DNA. Enrichment of trace pollen-free DNA from filtered honey samples opens a new approach to identify the true origins of filtered honey samples. The methods developed may be useful in other applications of trace DNA analysis.

In another study, we developed an immuno-PCR-based diagnostic platform which couples detection antibodies to self-assembled, ultra-detectable DNA-avidin nanoparticles

stabilized with poly(ethylene glycol) to link DNA amplification to target protein concentration. Electrostatic neutralization and steric cloaking of the PCR-amplifiable DNA labels by avidin and PEG coating reduces non-specific “stickiness” and enhances assay sensitivity. We further optimized the detectability of the nanoparticles by incorporating four repeats of a unique synthetic DNA PCR target into each nanoparticle. Using human chorionic gonadotropin hormone (hCG) as a model analyte, this platform was able to quantitate the target hCG protein in femtomolar concentrations using only standard laboratory equipment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Background

1.1. Assay development

The need for robust methods for the capture of trace molecules and ultrasensitive detection of biomarkers (nucleic acids or proteins) has constantly challenged the assay development community. For many years, the enzyme-linked immunosorbent assay (ELISA) has been widely used to detect and quantify analytes like toxins, hormones, and oncoproteins in clinical laboratories (Moura et al., 2004; Nguyen et al., 1990). However, ELISA is not always a suitable method for the detection of small molecules or trace substances (Chen et al., 2014; Janssen et al., 2013). To overcome these problems, a new technique called immuno-PCR (iPCR), the combination of ELISA and PCR, was first introduced in 1992 (Sano et al., 1992). In iPCR, a DNA reporter molecule is either chemically conjugated to detection antibody or attached through linking avidin or streptavidin. The increased specificity and nucleic acid amplification efficiency make iPCR up to 10,000- fold more sensitive than ELISA (Assumpção & da Silva, 2016).

1.2. Nucleic acid extraction

Nucleic acids often occur in dilute, contaminated form, and extraction is a crucial step for obtaining DNA or RNA in its intact form for its downstream applications such as PCR, sequencing, cloning, and transfection. The general protocol for any nucleic acid purification involves four steps: i) cell lysis; ii) removal of interfering components like lipids, polysaccharides, proteins, polyphenols, tannins, humic acid, and other inhibitory components; iii) nucleic acid purification, and iv) quantification of purified nucleic acid material. Cell lysis is achieved either by mechanical methods and/or chemical methods. Chemical lysis involves

the use of detergents like Triton X-100 and sodium dodecyl sulfate (most common in commercial kits) and enzymes like Proteinase K. Mechanical methods like sonication, bead-beating using glass beads or stainless-steel beads are often used for tough samples like plant cells, fungal spores, and insects (Ali et al., 2017; Soares et al., 2015).

1.3. Nucleic acid purification

Nucleic acids are commonly separated and purified from the sample or cell lysate by using either phase-separation-based or column-based methods. Phase-separation-based methods involve using a mixture of organic solvents like phenol and chloroform, followed by centrifugation to separate the nucleic acids in the aqueous layer. The isolated nucleic acids are further purified using well-established methods such as ethanol or isopropanol precipitation from the aqueous layer and finally resuspended in a low-salt solution such as 10 mM Tris or nuclease-free water (Tan & Yiap, 2009).

The discovery of silica-based capture of nucleic acid in the 1990s (Boom et al., 1990) led to the widespread use of silica spin columns for DNA purification. The reversible nucleic acid binding on silica columns is achieved using chaotropic salts such as guanidinium thiocyanate (GuSCN) or guanidinium hydrochloride (GuHCI), alcohol, and water. The chaotropic agent dehydrates the nucleic acid backbone allowing the nucleic acid to interact with silica particles under low pH, high salt conditions. Alcohol wash helps to remove other contaminants. Upon the addition of low-salt buffers like TE or water and at slightly alkaline pH, rehydration occurs, reversing the interaction, and the nucleic acid is eluted from the column (Melzak et al., 1996).

Other adsorbents that are commonly used for purification for nucleic acids are anion exchange resins (Q and DEAE) and ceramic hydroxyapatite (CHT) (Giovannini & Freitag, 2001). The negatively-charged nucleic acids interact with positively charged quaternary amines of the adsorbent (Q Sepharose) in the presence of low salt. The bound nucleic acid is finally eluted with high salt (2-3 M NaCl). The negatively charged nucleic acids interact with positively charged Ca^{2+} ions of CHT in low phosphate and are eluted from CHT in the presence of high phosphate (400 mM).

1.4. Sanger sequencing

The journey of DNA sequencing technology development has witnessed several paradigm shifts in the past 53 years. A successful preliminary attempt in DNA sequencing was reported by Wu and Kaiser in 1968 to determine the 12 bases in the cohesive ends of bacteriophage lambda DNA (Wu & Kaiser, 1968). The development of the chain-terminator procedure by Sanger and Coulson (Sanger et al., 1977) and the chemical cleavage procedure developed by Maxam and Gilbert (Maxam & Gilbert, 1977) set significant milestones in the early years of DNA sequencing. Sanger sequencing (also known as chain terminator sequencing, or dideoxy sequencing) was accurate, robust, easy to use, and became the basis for automated first-generation sequencing machines by 1987, which could sequence 1,000 bases per day (Hood et al., 1987). The method involves using specific fluorescent-labeled, chain-terminating nucleotides (dideoxynucleotides, ddNTPs) that lack a 3′-OH group. DNA polymerase cannot form the phosphodiester bond at the position where ddNTP is incorporated, resulting in the termination of the growing DNA chain. The label on the

terminating ddNTP corresponds to the nucleotide identity at that terminal position (Figure 1.1).

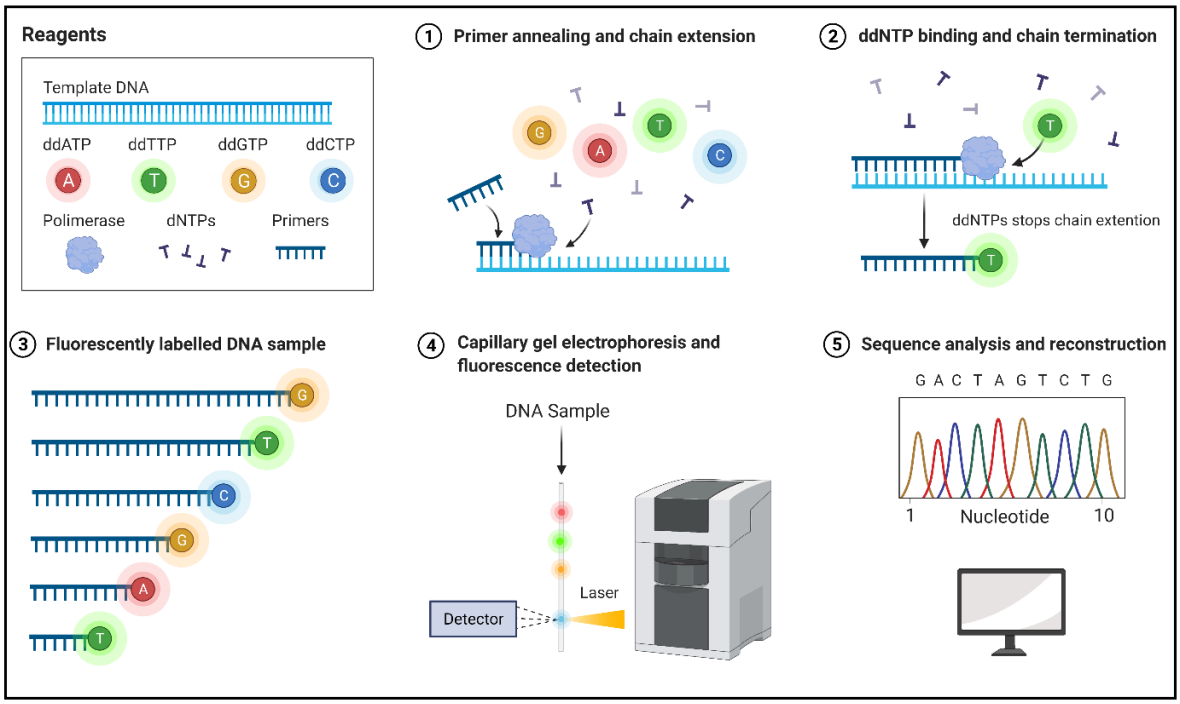


Figure 1.1. Steps involved in Sanger sequencing.

Illustration created with BioRender.com

1.5. Next-generation sequencing (NGS)

NGS (also known as massively parallel DNA sequencing, second-generation sequencing) started evolving in the late 2000s (Barba et al., 2014). The major advantage of NGS over Sanger sequencing was high sequencing throughput. The Sanger method could sequence a single DNA fragment at a time, whereas NGS allowed parallel sequencing of millions of fragments simultaneously in a single run. The two major NGS platforms (Slatko et al., 2018) currently governing the sequencing market are Illumina technology and Ion Torrent™ technology (Thermo-Fisher).

Illumina sequencing is based on a technique known as “bridge amplification”. A DNA library is prepared by ligating adapter sequences to each end of DNA fragments. DNA

fragments with adapters are then attached to a solid support (glass slide) that contains oligonucleotide sequences complementary to a ligated adapter. These serve as templates for repeated amplification synthesis reactions on a glass slide. The oligonucleotides on the slide are arranged so that after repeated amplification cycles, clonal "clusters" of DNA sequences are created. Millions of parallel clusters can be supported by each glass slide (Figure 1.2). Each cycle of sequence interrogation consists of a single-base extension with a modified DNA polymerase and a mixture of four nucleotides. Each of the four nucleotides has a fluorescent label and a reversible terminator. Once the fluorescent signal of added nucleotide is recorded, and the terminator is cleaved so the next base can bind. The reactions are repeated for several rounds (Adessi, 2000; Turcatti et al., 2008).

1.6. Overview of Thesis

The development of different methods for the capture and quantitative measurement of analytes from complex sample matrices has been discussed in the following chapters of the dissertation. Chapter 2 discusses the techniques for isolation, enrichment, and purification of trace pollen-free DNA for authentication of filtered honey. Chapter 3 discusses the next-generation sequencing-based authentication of high-value manuka honey. Chapter 4 discusses the construction of a novel iPCR reporter for ultrasensitive detection of protein biomarkers. Finally, Chapter 5 highlights the conclusions of this research and provides a perspective for future applications.

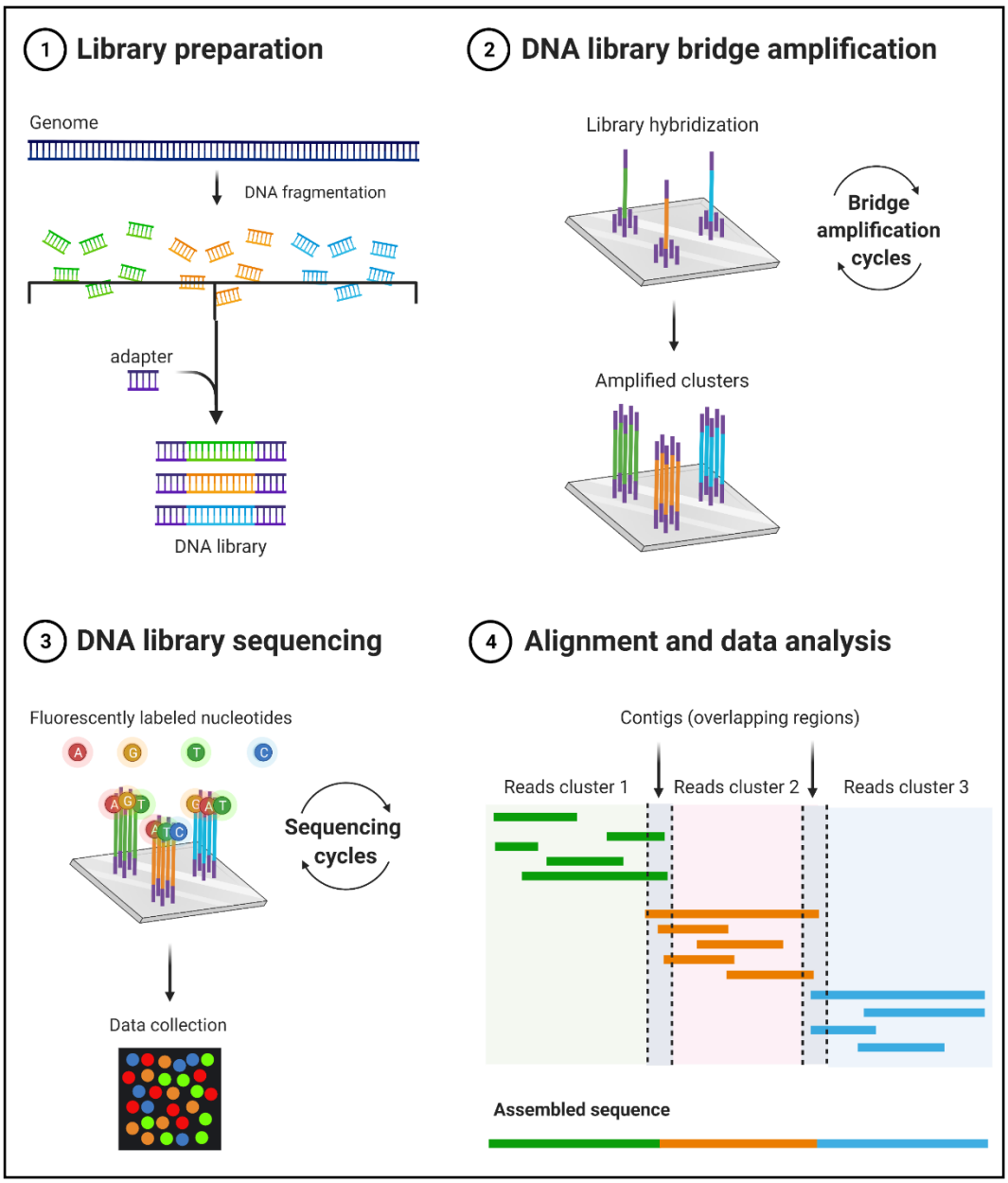


Figure 1.2. Steps involved in NGS platform developed by Illumina.

Illustration created with BioRender.com

2. Enrichment and purification of trace pollen-free DNA for authentication of honey

2.1. Introduction

With population growth and increasing demand for honey, there has been an increase in cases of honey mislabeling and adulteration (Fakhlaei et al., 2020), such that honey is the third most common target for adulteration, as reported by the U.S. Pharmacopeia's Food Fraud Database (García, 2018; Moore et al., 2012). The most common forms of honey adulteration include dilution with low-grade honey or sugar solution, filtering to remove pollen to thwart source identification, adding pollen of more-remunerative plants, and transshipment to mask true origins. Pollen microscopy (melissopalynology), a conventional method to study plant pollen found in honey (Bryant & Jones, 2001; Jones & Bryant, Jr, 2004; Petersen & Bryant, 2016), is time-consuming and fails to identify honey samples that have low pollen content or that have been filtered. Alternatively, researchers have used other analytical techniques like inductively coupled plasma-mass spectrometry (Batista et al., 2012; Chudzinska & Baralkiewicz, 2010; Zhou et al., 2018), nuclear magnetic resonance spectroscopy (Boffo et al., 2012; Ohmenhaeuser et al., 2013), near-infrared spectroscopy (Zhu et al., 2010), and gas chromatography (Aliferis et al., 2010), to authenticate the origins of honey. However, these techniques generally lack specificity and depend on specialized, high-complexity instrumentation (Soares et al., 2017).

Great progress is being made in DNA-based food authentication using next-generation sequencing (NGS) (Haynes et al., 2019). Several researchers have demonstrated the potential of pollen DNA to identify honey's botanical and geographical origins by targeting DNA barcode regions of plant DNA (Galimberti et al., 2014; Jain et al., 2013; Lalhmangaihi et al.,

2014; Richardson et al., 2015). The most extensively studied DNA barcodes in plants are three regions of the chloroplast genome (*rbcL*, *matK*, and *trnH-psbA*) and the nuclear ribosomal ITS (Bell et al., 2016). Most work has focused on pollen DNA and does not address the identification of honey samples that have been filtered. One exception is the work of Prosser and Hebert (Prosser & Hebert, 2017), who exploited the use of DNA barcoding of three regions (ITS2, rbcLa, and COI) to identify the botanical and entomological origins of seven honey types. They examined ITS2 from pollen material, but a shorter segment of plastid marker not necessarily from pollen (rbcLa, 162 bp) also was examined to identify trace and/or degraded plant DNA in liquid honey. They were able to identify both plant and insect sources in five out of seven samples. However, the liquid honey (1 ml of direct sample) used for studying the degraded plant DNA was neither centrifuged nor filtered to remove any pollens present before DNA extraction.

In this work, we tested three methods for capture of soluble pollen-free DNA (Figure 2.1): (i) Q Sepharose anion-exchange resin, (ii) ceramic hydroxyapatite (CHT) type I, and (iii) anti-dsDNA antibodies coupled to magnetic microspheres. The internal transcribed spacer 2 (ITS2) region of plant nuclear ribosomal DNA (nrDNA) was targeted because it is present in multiple copies in the plant genome (Álvarez & Wendel, 2003) and can be used to identify many plants at the genus or even species level (Chen et al., 2010; Han et al., 2013; Hollingsworth, 2011; Yao et al., 2010). We have tested and demonstrated the use of these techniques to establish the recovery and amplification of soluble DNA from filtered honey and the close correlation between pollen DNA and pollen-free DNA from three different honey samples originating from countries (United States of America, Greece, and Argentina).

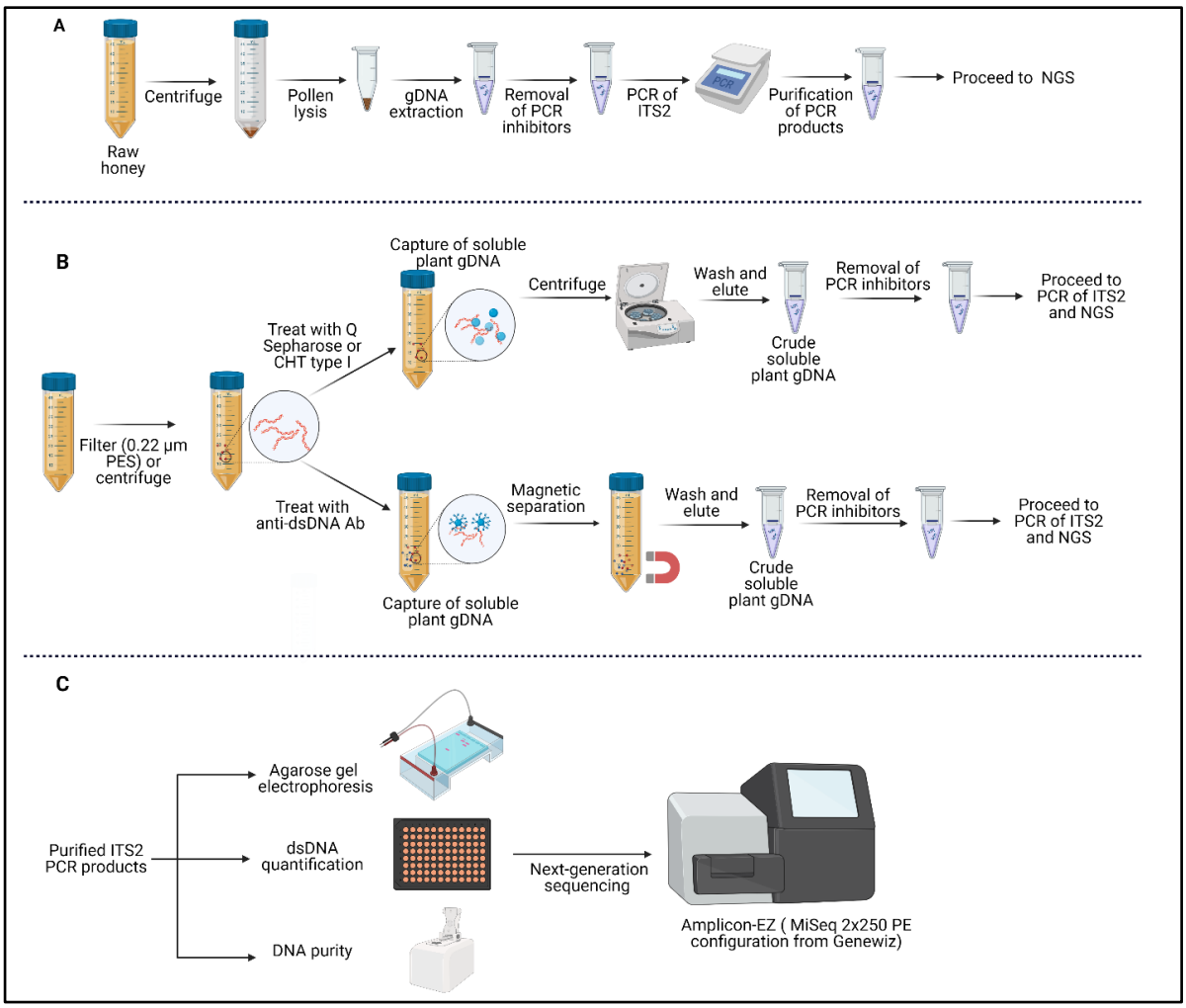


Figure 2.1. Schematic for the capture of pollen and pollen-free plant DNA from honey. A) Isolation and extraction of pollen DNA from a raw honey sample. B) Enrichment and purification of pollen-free DNA from the filtered honey sample. C) Steps for preparing samples for NGS

Illustration created with BioRender.com

2.2. Materials and Methods

2.2.1. Reagents

Anti-dsDNA monoclonal antibody (35I9 DNA; ab27156) was purchased from Abcam (Cambridge, Massachusetts). Promag amine beads (3.10 µm, PMA3N) were from Bangs Laboratories, Inc. (Fishers, Indiana). Zeba™ spin desalting columns (40K MWCO, 0.5 mL, 87766), AminoLink™ Reductant sodium cyanoborohydride (44892), and SYBR™ Safe DNA

Gel Stain (S33102) were purchased from ThermoFisher Scientific (Carlsbad, California). Amicon Ultra-0.5 centrifugal filters (100 kDa, UFC510096) were from MilliporeSigma (Burlington, Massachusetts). Phosphate-buffered saline (PBS) tablets, pH 7.4, were from Takara Bio USA Inc. (Mountain View, California).

MilliporeSigma™ Steriflip™ Sterile Disposable Vacuum Filter Units (polyethersulfone membrane, 0.22 µm, SCGP00525) and molecular biology grade ethanol (BP2818500) were purchased from Fisher Scientific (Hanover Park, Illinois). CHT™ ceramic hydroxyapatite, Type I (40 µm particle size) was from Bio-Rad (Hercules, California). Q Sepharose® Fast Flow (wet bead size 45-165 µm, preswollen in 20% ethanol, Q1126-100ML), Nuclease-Free Water, for Molecular Biology (W4502), Glass beads (425-600 µm, G9268), Corning® 96-well Black Flat Bottom Polystyrene NBS Microplate (3991), Trizma® base (T6066), Hydroxylamine hydrochloride (159417), Glycine (G7126), and Ethylenediaminetetraacetic acid disodium salt dihydrate (EDTA, molecular biology grade, E5134) were purchased from Sigma Aldrich (St. Louis, Missouri).

Buffer PB (19066), Buffer PE (concentrate, 100 ml, 19065), DNeasy Plant Mini Kit (69104) QIAquick Spin Columns (28115), QIAquick® PCR Purification Kit (28104), Nuclease-Free Water (129117), and QIAGEN Proteinase K (19131) were purchased from Qiagen (Germantown, Maryland). Q5® Hot Start High-Fidelity 2X Master Mix (M0494S), Gel loading buffer Purple (6X, B7024S), and Proteinase K (Molecular Biology Grade, P8107S) were purchased from New England BioLabs Inc. (Ipswich, Massachusetts).

Plant ITS2 primers used in this study were as previously reported by Chen *et al.* (Chen et al., 2010) Forward primer (20 nt, 5’-ATG CGA TAC TTG GTG TGA AT -3’) and Reverse primer (21 nt, 5’-GAC GCT TCT CCA GAC TAC AAT-3’) were purchased from Integrated

DNA Technologies, Inc. (Coralville, Iowa). Mx3000P optical strip tubes (401428) and Mx3000P optical strip caps (401425) were from Agilent Technologies, Inc. (Santa Clara, California). Eppendorf DNA LoBind Tubes, 2.0 ml, PCR clean, colorless (4043-1048) were purchased from USA Scientific, Inc. (Orlando, Florida). Agarose Med EEO (A1035) was from U.S. Biological Life sciences. QuantiFluor® dsDNA System (E2670) was purchased from Promega (Madison, Wisconsin).

2.2.2. Honey samples

Three different commercial raw honey samples purchased from local grocery stores were selected for this study (Table 2.1). Each honey sample was processed to obtain plant gDNA from its pollen, and soluble plant DNA (or pollen-free DNA) released from broken pollen from the liquid portion of the honey.

Table 2.1. Details of commercial honey samples used in this study

Brand name	Type of Honey	Country of origin
Kelley's Texas honey	Natural raw and unfiltered	USA
Attiki Pure Greek honey	Raw	Greece
Isabella Miel Pura De Abejas	-	Argentina

2.2.3. Extraction of plant gDNA from pollen isolated from honey

We developed a method to isolate plant genomic DNA (gDNA) from pollen based on that of Soares *et al.* (Soares et al., 2015). Approximately 15 g raw honey sample was made up to 50 g using nuclease-free water in a sterile 50 ml centrifuge tube. The diluted honey sample was then heated for 15 min in a water bath at 56 ^{o}C to allow homogeneous mixing. The tubes were centrifuged at 4000 *g* for 30 min at room temperature, the supernatant discarded, and the pellet containing pollen transferred to a sterile 2 ml microcentrifuge tube. The pellet was

washed with 2 ml nuclease-free water and recentrifuged at 4000 *g* for 15 min, and the pollen pellet was resuspended in 100 µl nuclease-free water.

The isolated pollens were pulverized by vortexing for 2 min at high speed with 7-8 sterile glass beads, and the disrupted pollens were transferred to a new sterile 2 ml microcentrifuge tube. At this stage, the samples can be stored at -20 ^{o}C until the next step. For efficient extraction of plant gDNA, at least 100 mg (wet weight) of the pollen pellet was processed for each sample. For honey samples with lower pollen content, two separate 15 g honey samples were processed as described above, and pooled.

Plant gDNA was extracted using Qiagen's DNeasy plant mini kit. The pollen pellet (100 mg) was treated with 400 µl Buffer AP1 from the kit and 25 µl of Proteinase K (20 mg/ml). The pellet was vortexed at medium speed for efficient lysis and to prevent shearing of gDNA. The treated pellet was incubated at 56 ^{o}C for 10 min, then cooled for 2 min at room temperature, and 4 µl RNase A (100 mg/ml; provided in the kit) was added. The tube was mixed by inverting 3-4 times and then incubated at 65 ^{o}C for 10 min. The remainder of the procedure was per the manufacturer's instructions, with gDNA eluted from DNeasy mini spin column in 200 µl of Buffer AE.

The resulting crude extract of plant gDNA was further purified using Qiagen's QIAquick PCR purification kit to remove PCR inhibitors. Briefly, 200 µl of plant gDNA in Buffer AE from the previous step was mixed with 1000 µl Buffer PB by aspirating 7-8 times using a microtip. The remainder of the procedure was per the manufacturer's instructions with gDNA was eluted from silica column in 50 µl of Buffer EB (10 mM Tris·Cl, pH 8.5; provided in the kit). The purified plant gDNA was then stored at -20 ^{o}C until further use, avoiding repeated freeze-thaw cycles and vortexing.

2.2.4. Methods for the enrichment of trace pollen-free DNA from honey

We tested three methods for capturing and enriching the traces of pollen-free DNA present in low-pollen-content or filtered honey samples. All buffers used for these in-house methods were prepared using nuclease-free water to prevent degradation of isolated gDNA.

a) Batch adsorption on an anion-exchanger (Q Sepharose)

Approximately 15 g of raw honey was diluted to 50 g with 150 mM NaCl in 20 mM Tris, pH 8.42 in a 50 ml sterile centrifuge tube. The diluted honey sample was heated for 15 min at 56 ^{o}C to allow homogeneous mixing. The honey was cooled, and the pH was adjusted to 8.5, followed by filtration using a sterile disposable vacuum filter unit (PES membrane, 0.22 µm). We modified the process for samples with low pollen content (such as Greece and Argentina used in this study), diluted honey after heat treatment and pH adjustment, was centrifuged at 4000 *g* for 30 min to separate out the pollen. The filtered honey sample or the supernatant obtained from a centrifuged honey sample was contacted with an anion exchange adsorbent to capture pollen-free DNA, as described below.

Briefly, Q Sepharose® Fast Flow resin was first uniformly resuspended by swirling, and 500 µl of resin suspension was then pipetted into a 15 ml sterile centrifuge tube. The resin was washed with 10 ml of 150 mM NaCl in 20 mM Tris, pH 8.42 by centrifugation at 2000 g for 20 min. The supernatant was discarded, and the settled resin was resuspended in 1.5 ml of 150 mM NaCl in 20 mM Tris, pH 8.42, to obtain 30% resin slurry (v/v). The 30% resin suspension was then added to the 50 ml tube containing the filtered honey sample (50 g) and placed on a rotator for 1 h at room temperature (28 rpm, Model #RT50, Cole-Parmer, Vernon Hills, Illinois).

An aliquot of 12 ml from a 50 ml tube of liquid treated with the resin was transferred to a 15 ml sterile tube and centrifuged at 2000 *g* for 10 min. The supernatant was discarded, and the resin was collected. This step was repeated till all resin was collected in the same 15 ml tube. The resin was washed twice with 5 ml of 400 mM NaCl in 20 mM Tris, pH 8.42 by centrifugation at 2000 *g* for 10 min. Finally, to elute the captured plant gDNA from the resin, 1.5 ml of 2 M NaCl in 20 mM Tris, pH 8.42 ("elution buffer") was added to the washed resin. The resin suspended in the elution buffer was transferred to a 2 ml sterile tube and incubated on a rotator for 30 min at room temperature (28 rpm). The resin was centrifuged at 2000 *g* for 10 min to yield supernatant containing the eluted pollen-free DNA.

The supernatant (approximately 1.5 ml) containing the isolated pollen-free DNA was then transferred to a new 15 ml tube and mixed with 7.5 ml of Buffer PB by aspirating gently 5-6 times. This mixture was then concentrated using a DNeasy mini spin column from Qiagen's DNeasy plant mini kit by centrifugation at 10,000 *g* for 1 min. The plant gDNA bound to the silica column was then washed with 500 µl Buffer AW2 by centrifugation at 10,000 *g* for 1 min. A second wash of AW2 was repeated at 20,000 *g* for 2 min. Finally, pollen-free DNA was eluted in 200 µl of Buffer AE and was stored overnight at -20 ^{o}C until the next step.

The resulting crude extract of pollen-free DNA was further purified using Qiagen's QIAquick PCR purification kit to remove PCR-inhibitory components. Briefly, 200 µl of pollen-free DNA present in Buffer AE was mixed with 1000 µl Buffer PB by aspirating 7-8 times. The rest of the steps for purification of DNA were as per the manufacturer's instructions, and finally, pollen-free DNA was eluted in 50 µl of Buffer EB (10 mM Tris·Cl,

pH 8.5; provided in the kit) and stored at -20 ºC until the next step. Repeated freeze-thaw cycles and vortexing of extracted DNA were avoided to prevent degradation of DNA.

b) Batch adsorption on ceramic hydroxyapatite (CHT) type I

Approximately 15 g of raw honey sample was diluted to 50 g using 1 mM EDTA and 0.5 M NaCl in 10 mM $NaPO_4$, pH 7.0 in a sterile 50 ml centrifuge tube. The diluted honey sample was heated for 15 min at 56 ºC to allow homogeneous mixing. The honey was cooled, and the pH was adjusted to 7.5, followed by filtration using a sterile disposable vacuum filter unit (PES membrane, 0.22 µm). The filtered honey sample was treated with CHT for the capture of soluble DNA, as described below.

Briefly, 500 mg CHT adsorbent was first weighed into a sterile 15 ml centrifuge tube. The adsorbent was washed with 10 ml of 1 mM EDTA in 10 mM $NaPO_4$, pH 7.0, by centrifugation at 750 *g* for 5 min. The supernatant was discarded, and the settled adsorbent was then resuspended in 1.5 ml of 1 mM EDTA in 10 mM $NaPO_4$, pH 7.0, to obtain 30% adsorbent slurry (w/v). The 30% adsorbent slurry was then added to the 50 ml tube containing the filtered honey sample and kept on a rotator for 1 h at room temperature (28 rpm).

A 12 ml sample from the 50 ml tube treated with the adsorbent was transferred to a 15 ml sterile tube and centrifuged at 750 *g* for 2 min. The supernatant was discarded, and the adsorbent was collected. This step was repeated till all the adsorbent material was collected in the same 15 ml tube. The adsorbent was washed twice with 5 ml of 1 mM EDTA in 10 mM $NaPO_4$, pH 7.0, by centrifugation at 750 *g* for 2 min. Finally, to elute the captured pollen-free DNA from the adsorbent, 1.5 ml of 1 mM EDTA in 400 mM $NaPO_4$, pH 7.0 ("elution buffer") was added to the washed adsorbent. The adsorbent suspended in the elution buffer

was transferred to a 2 ml sterile tube and incubated on a rotator for 30 min at room temperature (28 rpm). The tube was centrifuged at 750 *g* for 2 min to yield the supernatant containing the eluted pollen-free DNA.

The supernatant (approximately 1.5 ml) containing the isolated pollen-free DNA was then transferred to a new 15 ml tube and mixed with 7.5 ml of Buffer PB by aspirating gently 5-6 times. This mixture was concentrated on Qiagen's QIAquick spin column to promote DNA binding by centrifugation at 17,000 *g* for 1 min. The pollen-free DNA bound to the silica column was then washed using 750 µl Buffer PE by centrifugation. The column was centrifuged again to remove any traces of Buffer PE. The pollen-free DNA was eluted in 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5). We repeated the silica column treatment by mixing 50 µl eluted pollen-free DNA with 250 µl of Buffer PB. The rest of the steps were as described above after buffer PB. Finally, the pollen-free DNA was eluted in 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) and stored at -20 ^{o}C until the next step.

c) Capture by anti-dsDNA antibodies coupled to magnetic microspheres

Anti-dsDNA antibodies were coupled to amine magnetic microspheres using periodate-based carbohydrate oxidation as described in our previous publication (Goux et al., 2018). Briefly, 35 µl of anti-dsDNA antibody stock (0.91 mg/ml) was transferred into 100 mM sodium acetate buffer (pH 5.4) using a Zeba column (40K, 0.5 ml) and then mixed with 5 µl of 0.1 M $NaIO_4$. The tube containing this mixture (protected from light by aluminum foil) was incubated on a rotator (28 rpm) for 30 min at room temperature. The aldehyde-activated antibodies were purified and concentrated using a 100 kDa Amicon Ultra centrifugal filter in

200 mM sodium carbonate buffer (pH 9.6). The recovered antibody stock was diluted to 100 µl at 50 µg/ml in 200 mM sodium carbonate buffer (pH 9.6) and kept on ice until conjugation.

In another tube, 100 µl Promag amine microspheres (3.1 µm; 1×10^8 particles) were washed three times using a magnetic stand and resuspended in 100 µl of 200 mM sodium carbonate buffer (pH 9.6). The 100 µl of washed particles were mixed with 100 µl of oxidized antibody preparation and incubated on a rotator for 2 h at room temperature. After incubation, 5 µl of 5 M $NaCNBH_3$ was added to the reaction and incubated on a rotator for 30 min at room temperature. Unreacted aldehydes were quenched with 75 µl of 1 M hydroxylamine. The antibody-functionalized magnetic particles were separated and washed three times using phosphate-buffered saline (PBS; pH 7.4) using a magnetic stand. Finally, anti-dsDNA antibodies coupled to magnetic particles were resuspended in 100 µl PBS, pH 7.4, and stored at 4 ^{o}C until further use.

Approximately 15 g of raw honey was diluted to 50 g with 25 mM Tris, pH 8.42 in a 50 ml sterile centrifuge tube. The diluted honey sample was heated for 15 min at 56 ^{o}C to allow homogeneous mixing. The honey was cooled, and the pH was adjusted to 8.5, followed by filtration using a sterile disposable vacuum filter unit (PES membrane, 0.22 µm). The filtered honey sample was then treated with anti-dsDNA antibodies coupled to magnetic particles for the capture of soluble DNA, as described below.

Briefly, 20 µl anti-dsDNA antibodies coupled to magnetic particles in PBS, pH 7.4, were washed thrice, resuspended in 25 mM Tris, pH 8.42, and then added to the 50 ml tube containing the filtered honey sample. The tube was then kept on a rotator for 1 h at room temperature (28 rpm). A 2 ml aliquot from the 50 ml tube treated with the particles was concentrated in a 2 ml tube using magnetic separation till all particles from the 50 ml tube

were collected. The particles were washed twice with 5 ml of 25 mM Tris, pH 8.42 by magnetic separation. Finally, 50 µl of 100 mM glycine (pH 3) was added to elute the captured pollen-free DNA from the anti-dsDNA antibody-coupled magnetic particles. The 50 µl eluted pollen-free DNA was immediately transferred to another 2 ml tube containing 5 µl of 2 M Tris, pH 8.42.

The 55 µl of eluted pollen-free DNA was then mixed with 250 µl of Buffer PB by aspirating 5-6 times. This mixture was concentrated on a Qiagen's QIAquick spin column to promote DNA binding by centrifugation at 17,000 g for 1 min. The pollen-free DNA bound to the silica column was then washed using 750 µl Buffer PE by centrifugation. The column was centrifuged again to remove any traces of Buffer PE. The pollen-free DNA was eluted in 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5). We repeated the silica column treatment by mixing 50 µl eluted pollen-free DNA with 250 µl of Buffer PB. The rest of the steps were as described above after buffer PB. Finally, the pollen-free DNA was eluted in 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) and stored at -20 ^{o}C until the next step.

2.2.5. PCR amplification of the ITS2 region barcode region

The pollen DNA and pollen-free DNA isolated from the honey samples were quantified using the QuantiFluor dsDNA system prior to PCR. The purity of the isolated DNA was assessed by examining A260/A280 ratio on the Nanodrop instrument. The polymerase chain reaction (PCR) amplification was carried out in a total reaction volume of 50 µl containing 25 µl Q5® High-Fidelity 2X Master Mix, 2.5 µl of 10 µM of each primer, 2 µl of pollen DNA template (approximately 10-50 ng for most pollen samples) or 10 µl of pollen-free DNA template and nuclease-free water to a final volume of 50 µl. All PCR

reaction tubes were assembled in a PCR hood (Air Clean 600 PCR workstation). PCR was performed in an MJ Mini thermal cycler (Bio-Rad Laboratories, Hercules, California) using the following program: (i) initial denaturation at 98 ^{o}C for 30 sec; (ii) 40 cycles of 98 ^{o}C for 10 sec, 62 ^{o}C for 30 sec and 72 ^{o}C for 1 min; and (iii) final extension at 72 ^{o}C for 5 min.

The amplified ITS2 products were then purified using a Qiagen QIAquick PCR purification kit as per the manufacturer's instructions. Finally, PCR products were eluted in 50 µl of Buffer EB (10 mM Tris·Cl, pH 8.5; provided in the kit) and stored at -20 ^{o}C until further use. The PCR products were analyzed by1.5% agarose gel electrophoresis stained using SYBR™ safe DNA gel stain, and by Nanodrop absorbance measurements. The concentration of purified PCR products was determined using the QuantiFluor® dsDNA System.

We applied the following criteria to purified PCR products to be sent for amplicon-based NGS analysis: i) concentration of products normalized to 20 ng/µl; ii) at least 500 ng of DNA present, and iii) DNA purity index (A260/A280) 1.8-2.0. Typically, two reactions (each 50 µl) were pooled to obtain sufficient DNA for Genewiz Amplicon-EZ analysis (2 x 250 bp Illumina sequencing).

2.2.6. Data analysis

The credit for establishing the bioinformatics pipeline for the honey project goes to Dr. Aniko Sabo and Dr. Jay R T. Adolacion. The details of the pipeline are explained below. Each of the three honey samples was processed to study its pollen DNA and pollen-free DNA content (two or three replicates), resulting in 19 sample sets. Table 2.2 lists the codes assigned for each of the samples during bioinformatics analysis.

Table 2.2. NGS or sample codes assigned for 19 datasets of pollen DNA and pollen-free DNA

Type of DNA	Method of DNA capture and replicate number	Sample code
Sample codes for Kelley's Texas honey (USA)		
Pollen DNA	DNeasy Plant Mini Kit (n=3)	H15-1
		H15-2
		H15-3
Pollen-free DNA	Q Sepharose (n=3)	25D-1
		25D-2
		25D-3
	CHT type I (n=3)	25C-1
		25C-2
		25C-3
	Anti-dsDNA coupled to magnetic particles (n=2)	25A-T1
		25A-T2
Sample codes for Attiki pure Greek honey (Greece)		
Pollen DNA	DNeasy Plant Mini Kit (n=2)	H75-T2
		H75-T3
Pollen-free DNA	Q Sepharose (n=2)	177D-T1
		177D-T2
Sample codes for Isabella Miel Pura De Abejas (Argentina)		
Pollen DNA	DNeasy Plant Mini Kit (n=2)	H58-T2
		H58-T3
Pollen-free DNA	Q Sepharose (n=2)	187D-T1
		187D-T2

The raw FASTQ files received from Genewiz were analyzed using a bioinformatics pipeline as shown in Figure 2.2 and adapted from the DADA2 ITS Pipeline Workflow, and the workflow for Microbiome Data Analysis (Callahan, 2021; Callahan et al., 2016). All raw FASTQ files received were random mixtures of forward and reverse reads, and therefore the first step of the analysis was to segregate forward and reverse reads into R1_001 and R2_001 files, respectively. The first 20 or 21 bases were used to identify whether a read was a forward read or a reverse read based on its sequence matching with the primer sequences. A cutoff of 5 for the maximum number of mismatches was selected to retain the maximum number of unique reads mapped to the forward and reverse primers.

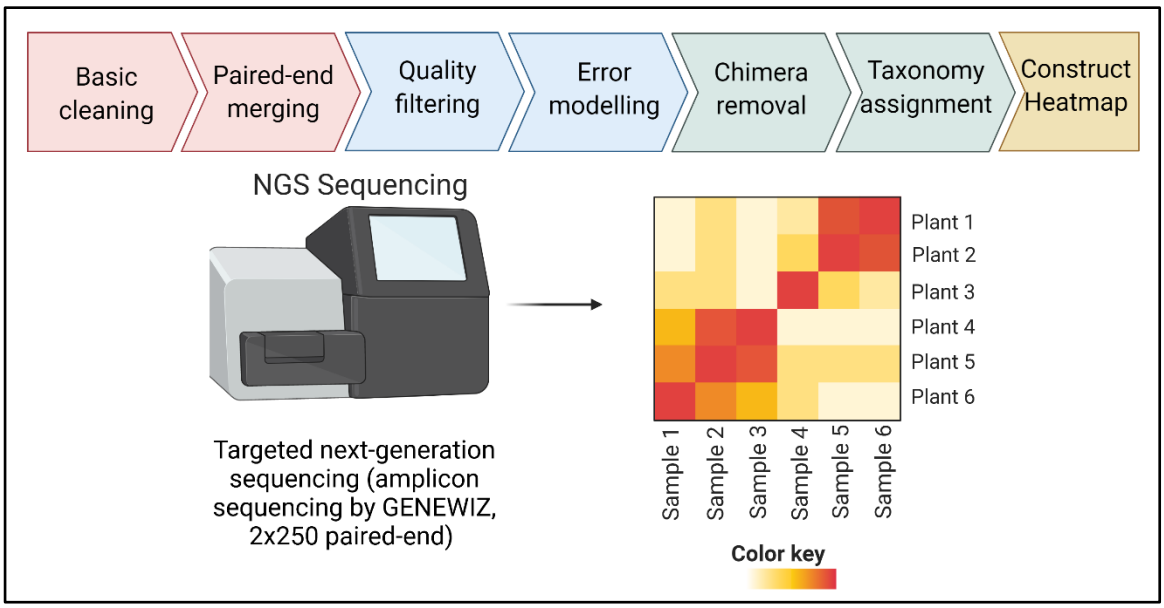


Figure 2.2. Schematic for processing DNA sequences for assigning plant taxa

Illustration created with BioRender.com

The next step was to stitch the paired-end reads using NGmerge (Gaspar, 2018) at a minimum overlap of 20 nt, allowing for a maximum of 5 nt or fewer mismatches. The merged reads containing both forward and reverse primers were considered full reads and spanned the entire ITS2 sequence. On the other hand, some ITS2 amplified regions exceeded the maximum effective read length of the sequencing platform, which was 480 nt (2×250 read length and subtracting 20 nt minimum overlap). In the cases where the reads failed to merge, we retained the forward reads only as they were on average higher quality than the reverse reads. Reads that successfully merged and forward reads that failed to merge were further processed.

Reads were filtered further to facilitate DADA2 (Callahan et al., 2016) error modeling for amplicon sequence variants (ASVs). All reads with ambiguous bases (Ns), reads with bases with quality scores below ten, and all reads with a cumulative expected error greater than three for merged reads or greater than four for forward reads were filtered. The filtering

parameters were chosen as such to refine error modeling for the DADA2 algorithm downstream of the analysis by improving read quality while retaining as many reads as possible. Any unusually short (<240 nt) and chimeric ASVs were removed as well.

Unique ASVs having more than two reads were searched against NCBI's nucleotide database (blastn) to identify the source organism. Results were restricted to the top ten sequences producing significant alignments and limited to records that include Viridiplantae (taxid:33090). Each ASV was assigned a species based on the top blast hit. To further refine the taxa assigned to the ASVs, ASVs with top blast hits that satisfied at least one of the following criteria were filtered out for low quality: i) percent identity in the alignment <85, ii) alignment length less than 1/3 of the length of the ASV (query) iii) alignment is less than 150 bp long, iv) blast bit score <200, and v) species name included a match to "environmental sample", "*Vachellia jacquemontii* (sequencing artifact deposited at NCBI)" or matched "N/A".

Heatmaps were constructed after expressing read counts in CPM (counts per million) from taxa assigned to the ASVs to establish a correlation between the relative abundance of pollen DNA sequences and pollen-free DNA sequences at the species level. This was achieved by converting the counts of each taxonomic label into fractions of the total sample size, then multiplied by one million, rounding off the resulting values to the nearest whole number.

Proportional relative abundance of plant DNA sequences identified at genus level obtained from pollen DNA, and pollen-free DNA of three commercial honey samples was then plotted as a bar plot. All counts were aggregated based on the taxonomic label's genus and expressed as fractions of the total sample size in the bar plot.

2.3. Results and Discussion

2.3.1. Enrichment and isolation of pollen DNA and pollen-free DNA

The concentrations of pollen DNA isolated from honey samples of USA, Greece, and Argentina were found to be 8.7 ng/µl, 7.1 ng/µl, and 1.98 ng/µl, respectively, with DNA purity ratio A260/A280 between 1.48 and 1.90 by Nanodrop. The concentration of pollen-free DNA isolated using Q Sepharose from three samples varied between 0.9 - 2.8 ng/µl, and DNA purity ratios A260/A280 were lower, between 1.39 - 1.50 by Nanodrop. The concentration of pollen-free DNA isolated using anti-dsDNA coupled magnetic particles and CHT was too low to be quantified using the QuantiFluor dsDNA kit.

2.3.2. Amplification of ITS2

The ITS2 region of plant DNA was successfully amplified from pollen DNA and pollen-free DNA using a primer pair published by Chen *et al.* (Chen et al., 2010). The ITS2 region in plants varies in length from ~180-390 bp (Moorhouse-Gann et al., 2018; Timpano et al., 2020). The forward and reverse primers anneal to the conserved regions of the 5.8S (~85 bp upstream of ITS2) and 26S (~142 bp downstream of ITS2) genes. Accordingly, ITS2 amplicon lengths ranging from 100 bp to 700 bp were observed, as shown in Figure 2.3 (for honey from the USA). A similar pattern of ITS2 PCR products between pollen DNA and pollen-free DNA captured by Q Sepharose was observed on an agarose gel. The yield of ITS2 PCR products obtained from USA honey pollen DNA was 106 ng/µl. The yields of ITS2 PCR products obtained from soluble DNA using Q Sepharose, CHT, and anti-dsDNA antibody coupled to magnetic particles were 99.6 ng/µl, 25 ng/µl, and 12.4 ng/µl, respectively. The low

concentrations of the ITS2 PCR products obtained using CHT and anti-dsDNA antibody made them difficult to observe on the gel.

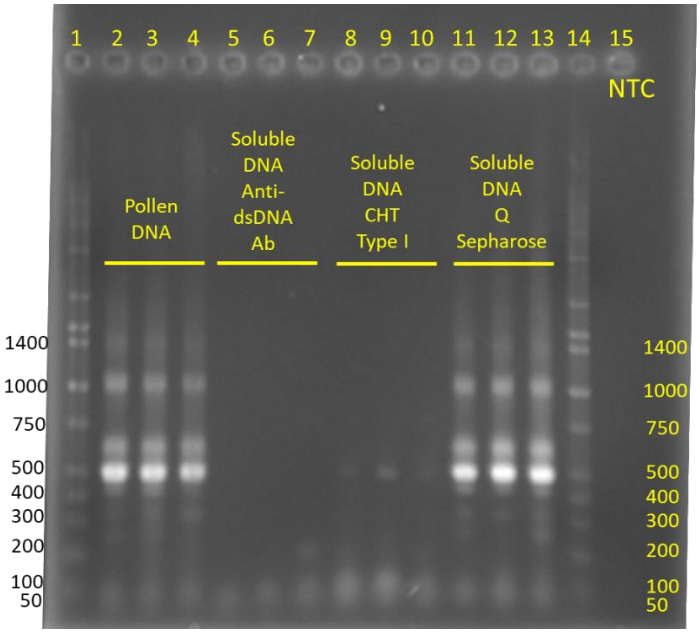


Figure 2.3. Agarose gel electrophoresis of ITS2 from pollen DNA and pollen-free DNA of Kelley's Texas honey (n=3). Lane 1 and 14: DNA ladder. Lanes 2 to 4: ITS2 amplified from pollen DNA of raw honey (H15). Lanes 5 to 13: ITS2 amplified from pollen-free DNA of filtered honey captured using three methods- a) anti-dsDNA Ab coupled to magnetic particles (25A), b) CHT type I (25C), and c) Q Sepharose (25D). Lane 15: No-template control.

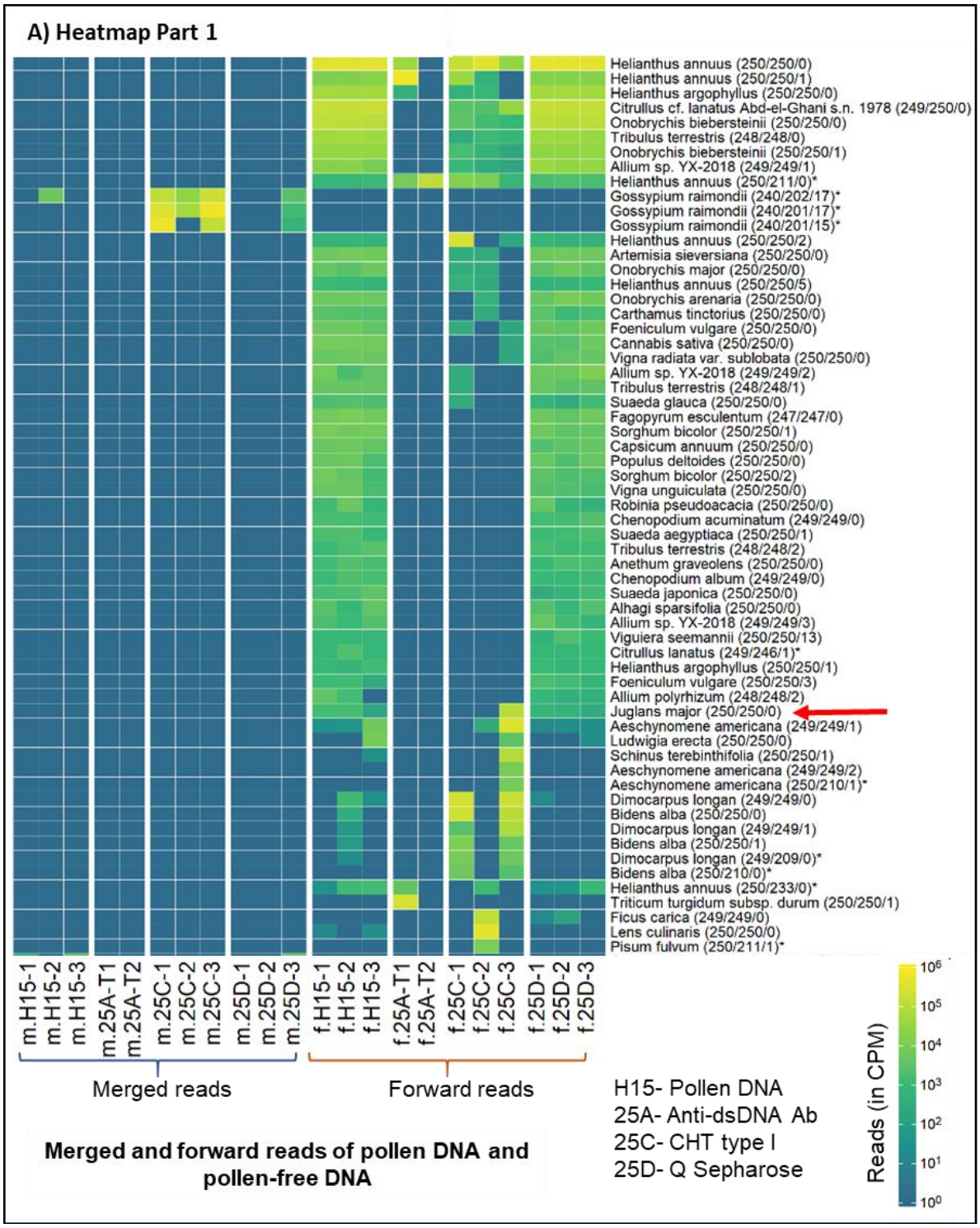
2.3.3. Data analysis for NGS of honey sample from the USA

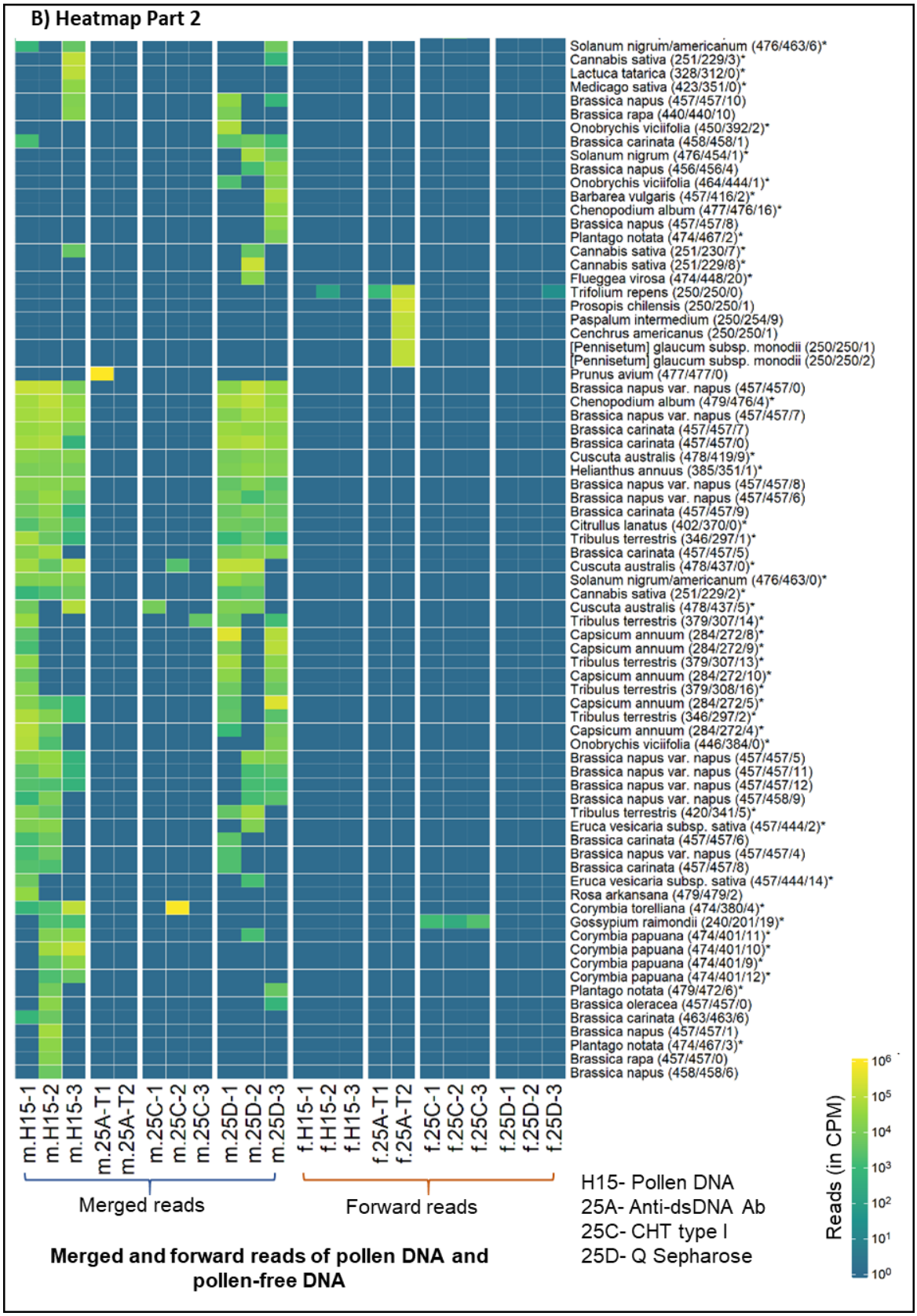
The same amount of starting honey sample (15 g) was processed to capture pollen DNA and pollen-free DNA and facilitate direct comparison between DNA sequences obtained from NGS of pollen DNA and pollen-free DNA. The ITS2 products obtained from pollen DNA and pollen-free DNA were then sequenced, and plant species identified were analyzed by plotting a heatmap (Figure 2.4) to understand the abundances of plant species observed across all datasets. ASVs classified to species level and having 5000 or more reads were plotted in the heatmap. The heatmap was split into two parts labeled as A and B of Figure 2.4 to make the labels of taxa assigned legible. We were interested to see if NGS data obtained for replicates of methods used for isolation of pollen DNA and pollen-free DNA showed the

same plant species. We were also interested in studying whether the methods were successful in establishing the correlation between the plant species observed in pollen DNA and pollen-free DNA.

We detected the same plant species when comparing the three replicates of pollen DNA (H15) with each other. Similarly, three replicates of pollen-free DNA captured by Q Sepharose (25D) agreed well with each other and with pollen DNA. This confirmed the reproducibility and robustness of the method to capture the pollen DNA and pollen-free DNA by Q Sepharose every time when a fraction of the sample was processed. However, not many plant species or much diversity was captured by isolation methods involving the use of anti-dsDNA antibodies and CHT. One of the reasons we speculate upon for poor isolation of pollen-free DNA by anti-dsDNA antibodies might be steric interference of large plant gDNA blocking the epitope sites of anti-dsDNA antibodies coupled to magnetic particles. So, not all sites are available for capturing trace pollen-free DNA from filtered honey samples. In case of CHT, which is commonly used for the isolation of DNA (Hilbrig & Freitag, 2012), the poor isolation of pollen-free DNA might be due to other interfering proteins bound to plant DNA.

Figure 2.4. Heatmap comparing the DNA sequences obtained from NGS of pollen DNA and pollen-free DNA of Kelley's Texas honey (USA). A) Heatmap Part 1. B) Heatmap Part 2. Rows of the heatmap correspond to plant species observed across the sample sets. Columns correspond to sample types (H15 represents three replicates of pollen DNA; 25A, 25C, and 25D represent two or three replicates of pollen-free DNA captured by anti-dsDNA antibody, CHT, and Q Sepharose). Columns starting with "m" are the merged reads, whereas columns starting with "f" are forward reads. Three numbers are printed after each taxonomic assignment (rows). The first number refers to the ASV length, the second number indicates the portion of the ASV length that mapped to a GenBank sequence, and the third number is the number of mismatches. Taxonomic assignments where only a fraction of the ASV sequences matched to the NCBI nt database are marked with asterisks.


A) Heatmap Part 1
Helianthus annuus (250/250/0)
Helianthus annuus (250/250/1)
Helianthus argophyllus (250/250/0)
Citrullus cf. lanatus Abd-el-Ghani s.n. 1978 (249/250/0)
Onobrychis biebersteinii (250/250/0)
Tribulus terrestris (248/248/0)
Onobrychis biebersteinii (250/250/1)
Allium sp. YX-2018 (249/249/1)
Helianthus annuus (250/211/0)*
Gossypium raimondii (240/202/17)*
Gossypium raimondii (240/201/17)*
Gossypium raimondii (240/201/15)*
Helianthus annuus (250/250/2)
Artemisia sieversiana (250/250/0)
Onobrychis major (250/250/0)
Helianthus annuus (250/250/5)
Onobrychis arenaria (250/250/0)
Carthamus tinctorius (250/250/0)
Foeniculum vulgare (250/250/0)
Cannabis sativa (250/250/0)
Vigna radiata var. sublobata (250/250/0)
Allium sp. YX-2018 (249/249/2)
Tribulus terrestris (248/248/1)
Suaeda glauca (250/250/0)
Fagopyrum esculentum (247/247/0)
Sorghum bicolor (250/250/1)
Capsicum annuum (250/250/0)
Populus deltoides (250/250/0)
Sorghum bicolor (250/250/2)
Vigna unguiculata (250/250/0)
Robinia pseudoacacia (250/250/0)
Chenopodium acuminatum (249/249/0)
Suaeda aegyptiaca (250/250/1)
Tribulus terrestris (248/248/2)
Anethum graveolens (250/250/0)
Chenopodium album (249/249/0)
Suaeda japonica (250/250/0)
Alhagi sparsifolia (250/250/0)
Allium sp. YX-2018 (249/249/3)
Viguiera seemannii (250/250/13)
Citrullus lanatus (249/246/1)*
Helianthus argophyllus (250/250/1)
Foeniculum vulgare (250/250/3)
Allium polyrhizum (248/248/2)
Juglans major (250/250/0)
Aeschynomene americana (249/249/1)
Ludwigia erecta (250/250/0)
Schinus terebinthifolia (250/250/1)
Aeschynomene americana (249/249/2)
Aeschynomene americana (250/210/1)*
Dimocarpus longan (249/249/0)
Bidens alba (250/250/0)
Dimocarpus longan (249/249/1)
Bidens alba (250/250/1)
Dimocarpus longan (249/209/0)*
Bidens alba (250/210/0)*
Helianthus annuus (250/233/0)*
Triticum turgidum subsp. durum (250/250/1)
Ficus carica (249/249/0)
Lens culinaris (250/250/0)
Pisum fulvum (250/211/1)*
m.H15-1
m.H15-2
m.H15-3
m.25A-T1
m.25A-T2
m.25C-1
m.25C-2
m.25C-3
m.25D-1
m.25D-2
m.25D-3
f.H15-1
f.H15-2
f.H15-3
f.25A-T1
f.25A-T2
f.25C-1
f.25C-2
f.25C-3
f.25D-1
f.25D-2
f.25D-3
Merged reads
Forward reads
H15- Pollen DNA
25A- Anti-dsDNA Ab
25C- CHT type I
25D- Q Sepharose
Reads (in CPM)
10^6
10^5
10^4
10^3
10^2
10^1
10^0
Merged and forward reads of pollen DNA and pollen-free DNA


B) Heatmap Part 2
Solanum nigrum/americanum (476/463/6)*
Cannabis sativa (251/229/3)*
Lactuca tatarica (328/312/0)*
Medicago sativa (423/351/0)*
Brassica napus (457/457/10)
Brassica rapa (440/440/10)
Onobrychis viciifolia (450/392/2)*
Brassica carinata (458/458/1)
Solanum nigrum (476/454/1)*
Brassica napus (456/456/4)
Onobrychis viciifolia (464/444/1)*
Barbarea vulgaris (457/416/2)*
Chenopodium album (477/476/16)*
Brassica napus (457/457/8)
Plantago notata (474/467/2)*
Cannabis sativa (251/230/7)*
Cannabis sativa (251/229/8)*
Flueggea virosa (474/448/20)*
Trifolium repens (250/250/0)
Prosopis chilensis (250/250/1)
Paspalum intermedium (250/254/9)
Cenchrus americanus (250/250/1)
[Pennisetum] glaucum subsp. monodii (250/250/1)
[Pennisetum] glaucum subsp. monodii (250/250/2)
Prunus avium (477/477/0)
Brassica napus var. napus (457/457/0)
Chenopodium album (479/476/4)*
Brassica napus var. napus (457/457/7)
Brassica carinata (457/457/7)
Brassica carinata (457/457/0)
Cuscuta australis (478/419/9)*
Helianthus annuus (385/351/1)*
Brassica napus var. napus (457/457/8)
Brassica napus var. napus (457/457/6)
Brassica carinata (457/457/9)
Citrullus lanatus (402/370/0)*
Tribulus terrestris (346/297/1)*
Brassica carinata (457/457/5)
Cuscuta australis (478/437/0)*
Solanum nigrum/americanum (476/463/0)*
Cannabis sativa (251/229/2)*
Cuscuta australis (478/437/5)*
Tribulus terrestris (379/307/14)*
Capsicum annuum (284/272/8)*
Capsicum annuum (284/272/9)*
Tribulus terrestris (379/307/13)*
Capsicum annuum (284/272/10)*
Tribulus terrestris (379/308/16)*
Capsicum annuum (284/272/5)*
Tribulus terrestris (346/297/2)*
Capsicum annuum (284/272/4)*
Onobrychis viciifolia (446/384/0)*
Brassica napus var. napus (457/457/5)
Brassica napus var. napus (457/457/11)
Brassica napus var. napus (457/457/12)
Brassica napus var. napus (457/458/9)
Tribulus terrestris (420/341/5)*
Eruca vesicaria subsp. sativa (457/444/2)*
Brassica carinata (457/457/6)
Brassica napus var. napus (457/457/4)
Brassica carinata (457/457/8)
Eruca vesicaria subsp. sativa (457/444/14)*
Rosa arkansana (479/479/2)
Corymbia torelliana (474/380/4)*
Gossypium raimondii (240/201/19)*
Corymbia papuana (474/401/11)*
Corymbia papuana (474/401/10)*
Corymbia papuana (474/401/9)*
Corymbia papuana (474/401/12)*
Plantago notata (479/472/6)*
Brassica oleracea (457/457/0)
Brassica carinata (463/463/6)
Brassica napus (457/457/1)
Plantago notata (474/467/3)*
Brassica rapa (457/457/0)
Brassica napus (458/458/6)
m.H15-1
m.H15-2
m.H15-3
m.25A-T1
m.25A-T2
m.25C-1
m.25C-2
m.25C-3
m.25D-1
m.25D-2
m.25D-3
f.H15-1
f.H15-2
f.H15-3
f.25A-T1
f.25A-T2
f.25C-1
f.25C-2
f.25C-3
f.25D-1
f.25D-2
f.25D-3
Merged reads
Forward reads
H15- Pollen DNA
25A- Anti-dsDNA Ab
25C- CHT type I
25D- Q Sepharose
Reads (in CPM)
10^6
10^5
10^4
10^3
10^2
10^1
10^0
Merged and forward reads of pollen DNA and pollen-free DNA

The plant species *Julgans major* or Arizona walnut (USDA, NRCS, 2021), known to occur in USA (Texas), was observed in forward reads of three replicates of pollen DNA and pollen-free DNA captured by Q Sepharose and in one replicate of CHT (pointed out with a red arrow in part A of Figure 2.4). The cosmopolitan plant species such as *Helianthus annus* (sunflower) and *Citrullus sp.* (watermelon), commonly cultivated in every country were observed in most of the sample sets. Honeybees are often attracted to common plant species like *Brassica napus* (rapeseed), *Helianthus annuus* (sunflower), *Allium sp.* (ornamental onions), and *Onobrychis sp.* (sainfoins) for collecting pollen (Farkas & Zajácz, 2007). This might be the reason for the presence of the above plants in the sample.

The bar plot (Figure 2.5) shows the proportional abundances of the top twenty plant DNA sequences identified at the genus level across the pollen and pollen-free DNA sample sets. 115 different genera were observed across 22 sample sets. This plot also showed the same output as in the heatmap representation constructed for studying the plant species that were observed in pollen DNA and pollen-free DNA by three methods. We were interested in identifying the most abundant genera observed as ITS2 is known to classify at the genus level strongly.

As expected, we saw similar abundance ratios of the top twenty genera between the replicates of the merged and forward reads for pollen DNA (H15) and pollen-free DNA by Q Sepharose (25D). We saw a strong correlation between genera captured in pollen DNA (H15) and pollen-free DNA by Q Sepharose (25D). However, the replicates of isolation methods involving the use of anti-dsDNA antibodies and CHT showed poor diversity and were not as reproducible. The only genus *Prunus* was observed in one of the replicates of merged reads of pollen-free DNA captured by anti-dsDNA antibodies (m25A-T1). This output was consistent

with the output of the heatmap. Genus *Helianthus* was the top hit observed in all sets of forward reads of pollen DNA and pollen-free DNA captured by three methods.

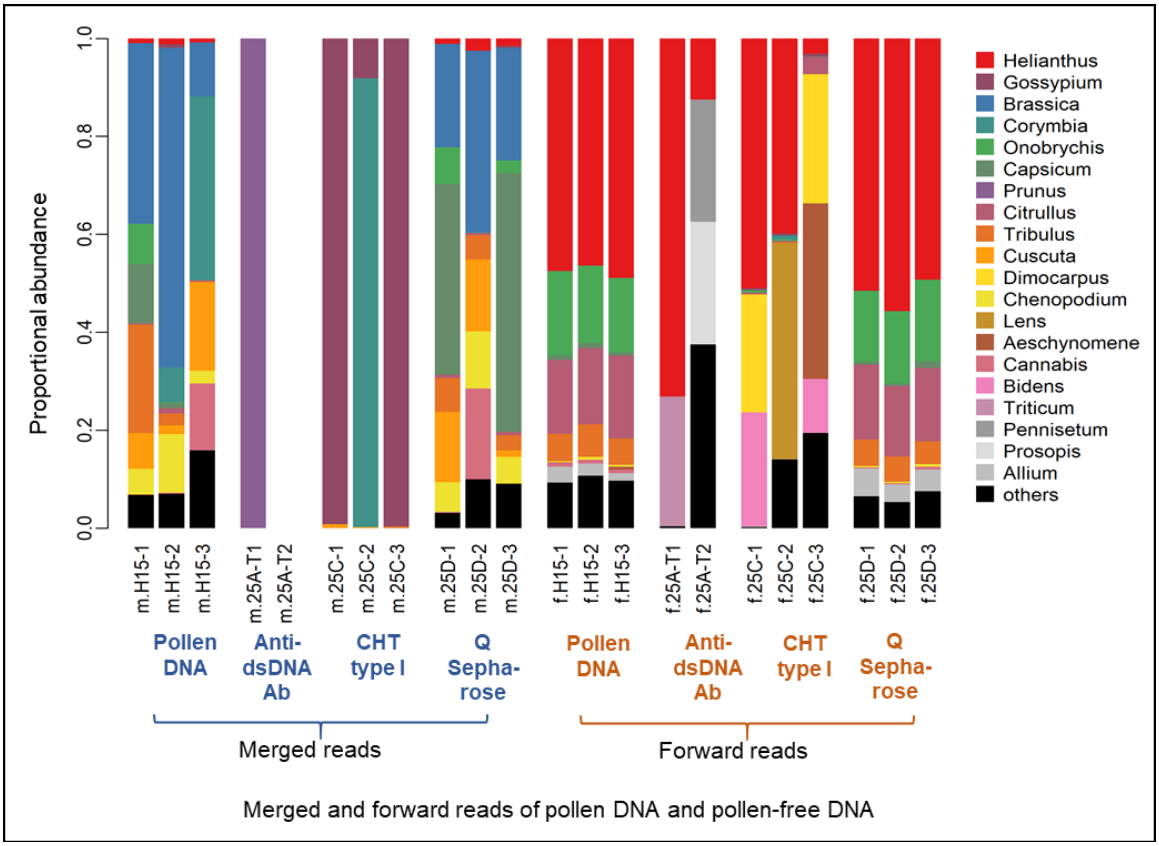


Figure 2.5. Bar plot indicating proportional richness and abundance of top twenty plant genera obtained from NGS of pollen DNA and pollen-free DNA of Kelley's Texas honey, USA. Columns correspond to sample types (H15 represents three replicates of pollen DNA; 25A, 25C, and 25D represent two or three replicates of pollen-free DNA captured by anti-dsDNA antibody, CHT, and Q Sepharose). Columns starting with "m" are the merged reads, whereas columns starting with "f" are forward reads.

2.3.4. Data analysis for NGS of honey samples from Greece and Argentina

Based on these results, Q Sepharose was more preferred to capture the pollen-free DNA from filtered samples, and we selected this method for further testing. Several researchers have published papers using Q Sepharose to purify DNA in other settings, including us (Murphy et al., 2003). Additional raw honey samples H75 (Greece) and H58 (Argentina) were treated with Q Sepharose to isolate pollen-free DNA. The yields of ITS2

PCR products obtained from pollen DNA and pollen-free DNA from the honey sample from Greece were 57 ng/µl and 19 ng/µl, respectively. The yields of ITS2 PCR products obtained from pollen DNA and pollen-free DNA from the honey sample from Argentina were 80 ng/µl and 72 ng/µl, respectively.

A similar pattern of ITS2 PCR amplicons was observed between pollen DNA and pollen-free DNA captured by Q Sepharose for honey from both Greece and Argentina on agarose gel electrophoresis with a distinct band of ITS2 PCR product appearing at 500 bp, as shown in Figure 2.6. The ITS2 PCR products from each sample set (pollen DNA and pollen-free DNA) were sequenced twice to cover the maximum species observed in the sample and establish the correlation between pollen DNA and pollen-free DNA. At the same time, we wanted to gather information about any artifacts introduced during the sequencing process or any bias among specific DNA sequences during the sequencing run.

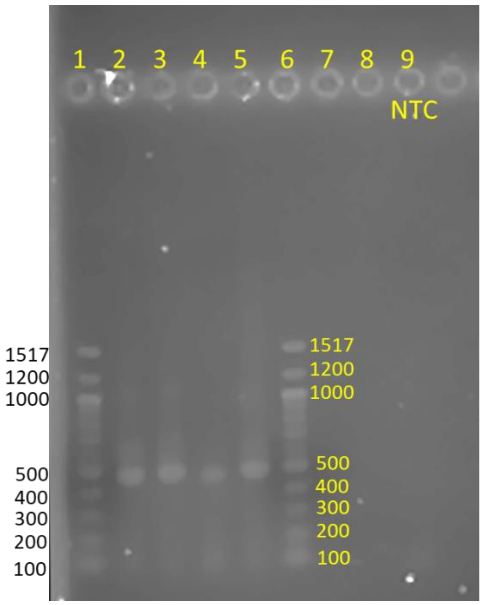


Figure 2.6. Agarose gel electrophoresis of ITS2 from pollen DNA and pollen-free DNA of honey samples from Greece and Argentina. Lane 1 and 6: DNA ladder. Lanes 2 and 3: ITS2 amplified from Pollen DNA of Greece (H75) and Argentina (H58) respectively. Lanes 4 and 5: ITS2 amplified from pollen-free DNA captured by Q Sepharose of Greece (177D) and Argentina (187D) respectively. Lane 9: No-template control.

A bar plot was generated to study the abundances of the top twenty plant genera obtained from NGS of pollen DNA (H75) and pollen-free DNA by Q Sepharose (177D) sample sets from Greece (Figure 2.7). A total of 132 different genera were observed, from which the twenty most abundant genera were plotted. We saw a similarity in the abundances of genera observed between replicates of both pollen DNA and pollen-free DNA. This observation firmly concluded that no artifact or bias was introduced during sequencing. Additionally, we saw a close correlation between the genera observed in reads of pollen DNA and pollen-free DNA. However, not all genera of pollen DNA were observed in the dataset of pollen-free DNA. We speculate this might be due to differential release of DNA as pollen-free DNA from broken pollen of different plants (Bolick & Vogel, 1992).

The genus *Polygonum* was the most abundant plant genus observed in the merged reads of pollen DNA and pollen-free DNA (Figure 2.7). The majority of the merged reads of pollen-free DNA were from the Genus *Polygonum*. *Erica* was the third most abundant genus found to be common between forward reads of pollen DNA and pollen-free DNA. *Erica manipuliflora* (heather) is known to exist in Albania, Cyprus, Greece, Italy, Kriti, Lebanon, Syria, Sicily, Turkey, and the former Yugoslavia. Another Greece-specific plant species *Quercus coccifera* (kermes oak), was observed in forward reads of both pollen DNA and pollen-free DNA of the Greek honey (POWO, 2021).

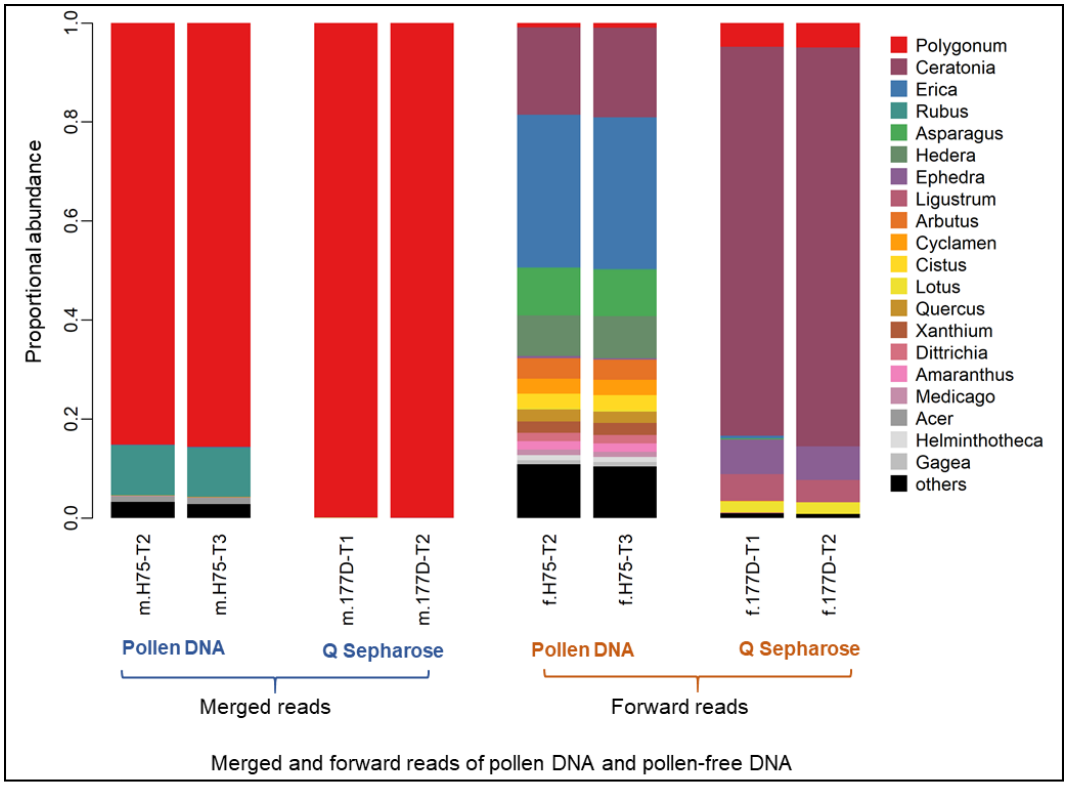


Figure 2.7. Bar plot indicating proportional richness and abundance of top twenty plant genera obtained from NGS of pollen DNA and pollen-free DNA of Attiki Pure Greek honey (Greece). Columns correspond to sample types (H75 represents two replicates of pollen DNA and 177D represents two replicates of pollen-free DNA captured by Q Sepharose). Columns starting with "m" are the merged reads, whereas columns starting with "f" are forward reads.

A bar plot (Figure 2.8) was generated to study the abundances of the top twenty plant genera identified in Argentine honey by NGS of pollen DNA (H58) and pollen-free DNA isolated by Q Sepharose (187D). A total of 108 different genera were observed, of which the twenty most abundant genera were plotted. As with the previous sample from Greece, we saw similar abundances of genera between replicates of both pollen DNA and pollen-free DNA, indicating no artifacts during sequencing. Additionally, we saw a strong correlation between the genera observed in pollen DNA and pollen-free DNA.

Amongst these 108 genera, the top hit was for Genus *Lotus* and *Eucalyptus* (Figure 2.8) in both merged and forward reads of pollen DNA and pollen-free DNA. *Lotus* and *Eucalyptus* are known to occur in Argentina as introduced species (POWO, 2021).

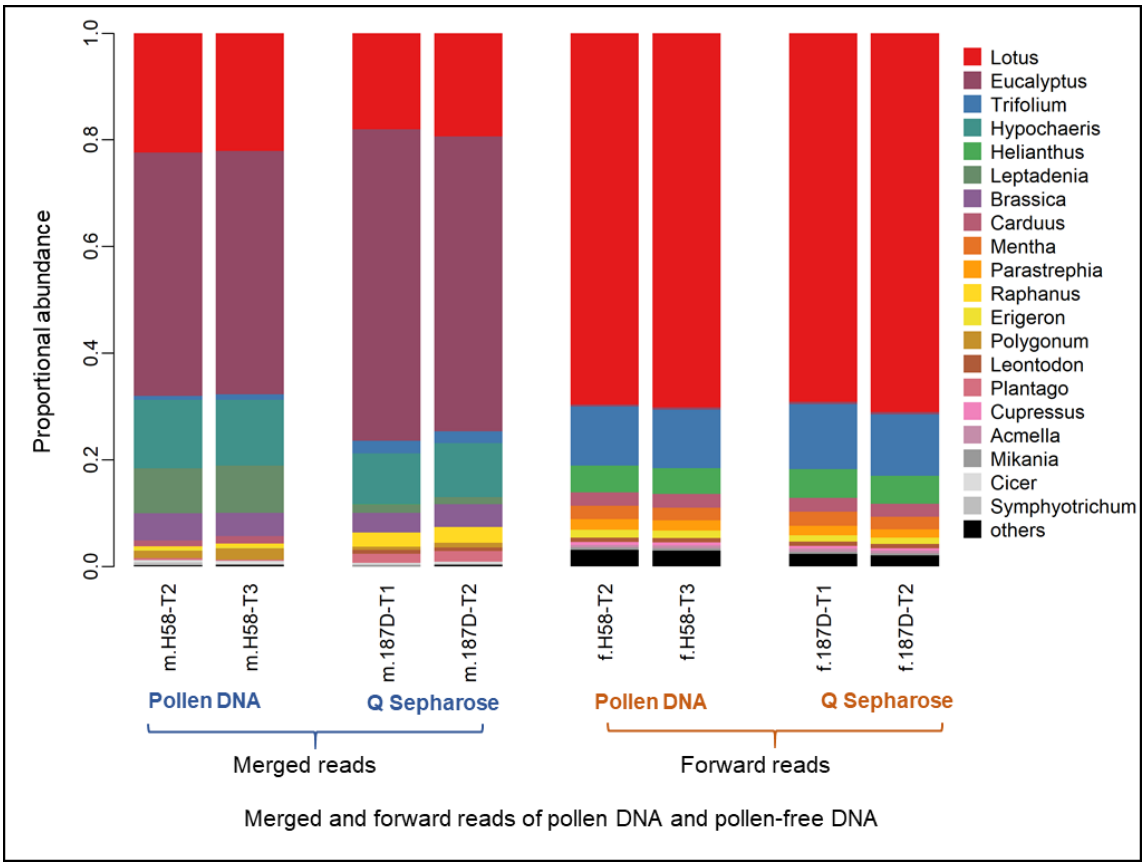


Figure 2.8. Bar plot indicating proportional richness and abundance of top twenty plant genera obtained from NGS of pollen DNA and pollen-free DNA of Isabella Miel Pura De Abejas (Argentina). Columns correspond to sample types (H58 represents two replicates of pollen DNA and 187D represents two replicates of pollen-free DNA captured by Q Sepharose). Columns starting with "m" are the merged reads, whereas columns starting with "f" are forward reads.

3. Authentication of manuka honey by targeted next-generation sequencing

3.1. Introduction

Manuka honey contains pollen of the signature plant, the manuka myrtle tree (*Leptospermum scoparium*), which occurs throughout New Zealand and eastern Australia (El-Senduny et al., 2021; McDonald et al., 2018). Manuka honey is widely consumed for its reputed medicinal properties, including antioxidant and antimicrobial activities. Methylglyoxal (MGO) formed from dihydroxyacetone (DHA) is the major constituent contributing to the antibacterial activity of manuka honey (Johnston et al., 2018; Smallfield et al., 2018). The 'unique manuka factor' (UMF) assigned to manuka honey corresponds to the equivalent of the phenol concentration showing antibacterial activity (Burns et al., 2018; Kato et al., 2014; Wallace et al., 2010). The market price of manuka honey is influenced by high values of UMF, or MGO content. Around 11,000 tons of honey are produced each year in New Zealand, of which ca. 7,000 tons are exported as manuka honey (Burns et al., 2018; Zhou et al., 2018).

Currently, honey sourced in New Zealand is classified as manuka honey only if it passes tests for a combination of five attributes developed by the Ministry for Primary Industries (MPI) of New Zealand (Table 1). The four chemical markers found in manuka honey are 3-phenyllactic acid, 2′-methoxyacetophenone, 2-methoxybenzoic acid, and 4-hydroxyphenyllacetic acid, which are quantified using LC-MS/MS (MPI, 2017a). The DNA marker (an unidentified sequence from *Leptospermum scoparium*) is determined by multiplex qPCR, which helps to differentiate between manuka honey and kanuka honey (MPI, 2017b).

Table 3.1. The five attributes and specified limits to characterize monofloral and multifloral manuka honey by MPI (MPI, 2017a; MPI, 2017b)

Compounds	Monofloral	Multifloral
3-Phenyllactic acid	≥ 400 mg/kg	≥ 20 mg/kg but < 400 mg/kg
2′-Methoxyacetophenone	≥5 mg/kg	≥ 5 mg/kg
2-Methoxybenzoic acid	≥1 mg/kg	≥ 1 mg/kg
4-Hydroxyphenyllactic acid	≥1 mg/kg	≥ 1 mg/kg
Manuka DNA sequence PCR	< C_q 36 (approximately 3 fg/µl)	< C_q 36 (approximately 3 fg/µl)

Kanuka honey contains pollen of the kanuka plant (*Kunzea ericoides*) belonging to the same Myrtaceae family as manuka. The kanuka plant is morphologically very similar to manuka, and they have overlapping flowering times (Semprini et al., 2019; Smallfield et al., 2018). The qPCR criteria set by MPI rely on the use of a commercial ManKan™ qPCR kit to amplify unspecified sequences (MPI, 2017b). The kit was validated by testing its specificity with over 800 honey samples collected from different floral sources and geographic locations. Although the kit is specific for *Leptospermum scoparium*, this approach is vulnerable to counterfeiting by the addition of the amplicon it produces, in a manner similar to recent incidents of adulteration by the addition of chemical constituents (MGO and DHA) of manuka (MPI, 2019). Moreover, the qPCR assay also gives no insight into the diversity of other plants found in manuka honeys.

In this work, we have developed an NGS-based approach to authenticate manuka samples based on all plants contributing to their pollen DNA content (Figure 3.1). We used a method based on the work of Soares *et al.* (Soares et al., 2015) to isolate plant genomic DNA (gDNA) from pollen obtained from manuka honey samples (n=21). The internal transcribed spacer 2 (ITS2) region of plant nuclear ribosomal DNA (nrDNA) was targeted because it is present in multiple copies in the plant genome and can identify many plants to the genus or

even species level. We developed an in-house bioinformatics pipeline to study other plant ITS2 sequences that co-occur with the manuka plant in honey pollen DNA. This NGS-based approach opens a new door to authenticate the botanical origins of manuka samples.

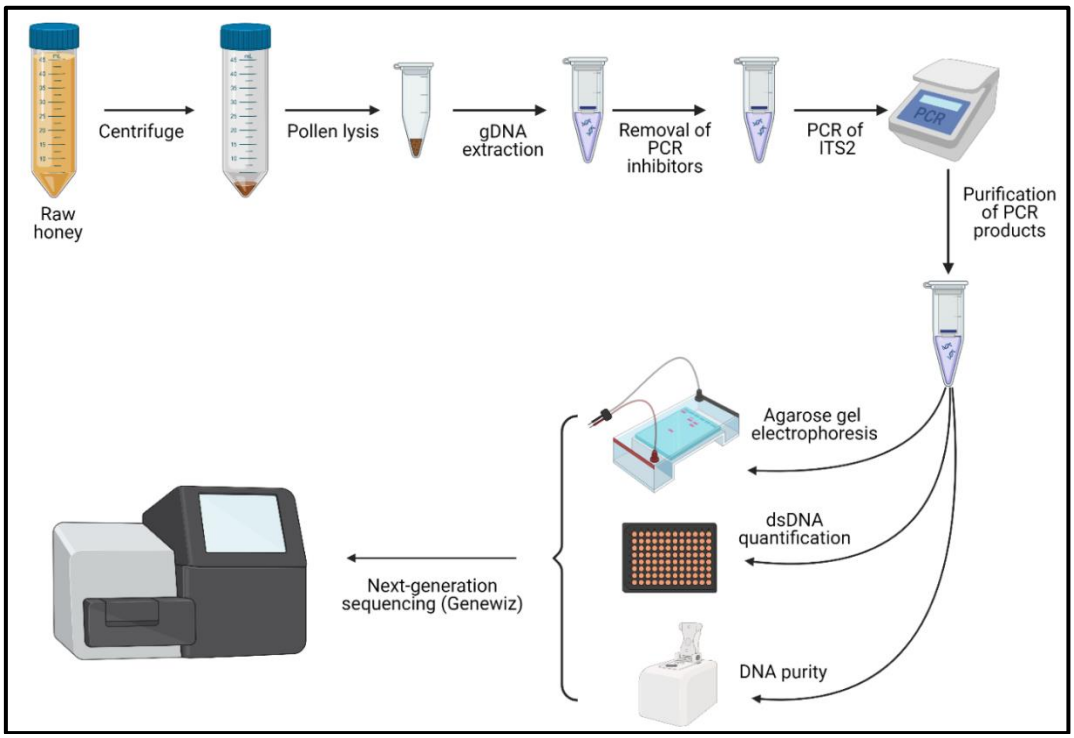


Figure 3.1. Schematic for isolation and next-generation sequencing of pollen DNA obtained from Manuka samples.

Illustration created with BioRender.com

3.2. Materials and Methods

3.2.1. Reagents

Molecular biology grade ethanol (BP2818500) was purchased from Fisher Scientific (Hanover Park, Illinois). Nuclease-Free Water, for Molecular Biology (W4502), Glass beads (425-600 µm, G9268), Corning® 96-well Black Flat Bottom Polystyrene NBS Microplate (3991), Trizma® base (T6066), and Ethylenediaminetetraacetic acid disodium salt dihydrate (EDTA, molecular biology grade, E5134) were purchased from Sigma Aldrich (St. Louis,

Missouri). DNeasy Plant Mini Kit (69104) QIAquick Spin Columns (28115), QIAquick® PCR Purification Kit (28104), Nuclease-Free Water (129117), and QIAGEN Proteinase K (19131) were purchased from Qiagen (Germantown, Maryland). Q5® Hot Start High-Fidelity 2X Master Mix (M0494S), Gel loading buffer Purple (6X, B7024S), and Proteinase K (Molecular Biology Grade, P8107S) were purchased from New England BioLabs Inc. (Ipswich, Massachusetts). Plant ITS2 primers used in this study were as previously reported by Chen et al. (Chen et al., 2010): Forward primer (20 nt, 5'-ATG CGA TAC TTG GTG TGA AT -3') and Reverse primer (21 nt, 5'-GAC GCT TCT CCA GAC TAC AAT-3') were purchased from Integrated DNA Technologies, Inc. (Coralville, Iowa). Mx3000P optical strip tubes (401428) and Mx3000P optical strip caps (401425) were from Agilent Technologies, Inc. (Santa Clara, California). Eppendorf DNA LoBind Tubes, 2.0 ml, PCR clean, colorless (4043-1048) were purchased from USA Scientific, Inc. (Orlando, Florida). Agarose Med EEO (A1035) was from U.S. Biological Life sciences. SYBR™ Safe DNA Gel Stain (S33102) were purchased from ThermoFisher Scientific (Carlsbad, California). QuantiFluor® dsDNA System (E2670) was purchased from Promega (Madison, Wisconsin).

3.2.2. Honey samples

Twenty-one different manuka samples selected for this study were purchased from local grocery stores and online (Table 3.2).

Table 3.2. Details of Manuka samples studied in this research and codes assigned for NGS analysis

NGS Sample Code	Qty (g)	Cost ($)	Approximate Cost ($) / (g)
H19	325	22.95	0.07
H42	500	36.99	0.07
H79	500	15.00	0.03
H80	250	45.00	0.18
H81	250	14.99	0.06
H82	250	64.99	0.26
H83	500	37.99	0.08
H84	250	17.70	0.07
H85	110	12.75	0.12
H86	250	26.99	0.11
H87	500	49.99	0.10
H88	500	146.99	0.29
H89	250	24.87	0.10
H90	250	24.99	0.10
H105	500	29.95	0.06
H109	500	15.99	0.03
H111	500	100.00	0.20
H117	250	23.95	0.10
H118	250	39.99	0.16
H123	250	85.00	0.34
H150	1000	76.99	0.08

3.2.3. Extraction and purification of plant gDNA from pollen isolated from honey

We developed a method to isolate plant gDNA from pollen based on the work of Soares *et al.* (Soares et al., 2015). The protocol for extraction is same as described in section 2.2.3 of chapter 2.

3.2.4. PCR amplification of ITS2 region

Pollen DNA isolated from each manuka honey sample was quantified using the QuantiFluor dsDNA system prior to PCR. Polymerase chain reaction (PCR) amplification was carried out in a total reaction volume of 50 µl containing 25 µl Q5® High-Fidelity 2X

Master Mix, 2.5 µl of 10 µM of each primer, 2 µl of pollen DNA template (approximately 1-50 ng for most samples) and nuclease-free water to a final volume of 50 µl. All PCR reaction tubes were assembled in a PCR hood (Air Clean 600 PCR workstation). PCR was performed in an MJ Mini thermal cycler (Bio-Rad Laboratories, Hercules, California) using the following program: (i) initial denaturation at 98 ^{o}C for 30 sec; (ii) 40 cycles of 98 ^{o}C for 10 sec, 62 ^{o}C for 30 sec and 72 ^{o}C for 1 min; and (iii) final extension at 72 ^{o}C for 5 min.

The plant ITS2 amplicons were then purified using a Qiagen QIAquick PCR purification kit as per the manufacturer's instructions. Finally, ITS2 PCR products were eluted in 50 µl of Buffer EB (10 mM Tris·Cl, pH 8.5; provided in the kit) and stored at -20 ^{o}C until further use. The PCR products were analyzed by1.5% agarose gel electrophoresis stained using SYBR™ safe DNA gel stain, and by Nanodrop absorbance measurements. The concentration of the purified PCR products was determined using the QuantiFluor dsDNA System.

We applied the following criteria to purified PCR products to be sent for amplicon-based NGS analysis: i) concentration of products normalized to 20 ng/µl; ii) at least 500 ng of DNA present, and iii) DNA purity index (A260/A280) 1.8-2.0. Typically, two reactions (each 50 µl) were pooled to obtain sufficient DNA for Genewiz Amplicon-EZ analysis (2 x 250 bp Illumina sequencing).

3.2.5. Data analysis

Twenty-one honey samples were processed to study the pollen DNA, resulting in twenty-one datasets. The raw FASTQ files received from Genewiz were analyzed using a bioinformatics pipeline adapted from the DADA2 ITS Pipeline Workflow and the workflow

for Microbiome Data Analysis (Callahan, 2021; Callahan et al., 2016). All raw FASTQ files received were random mixtures of forward and reverse reads, and therefore the first step of the analysis was to segregate forward and reverse reads into R1_001 and R2_001 files, respectively. The first 20 or 21 bases were used to identify whether a read was a forward read or a reverse read based on its sequence matching with the primer sequences. A cutoff of 5 for the maximum number of mismatches was selected to retain the maximum number of unique reads mapped to the forward and reverse primers.

The next step was to stitch the paired-end reads using NGmerge (Gaspar, 2018) at a minimum overlap of 20 nt, allowing for a maximum of 5 nt or fewer mismatches. The merged reads containing both forward and reverse primers were considered full reads and spanned the entire ITS2 sequence. On the other hand, some ITS2 amplified regions exceeded the maximum effective read length of the sequencing platform, which was 480 nt (2×250 read length and subtracting 20 nt minimum overlap). In the cases where the reads failed to merge, we retained the forward reads only as they were on average higher quality than the reverse reads. Reads that successfully merged and forward reads that failed to merge were further processed.

Reads were filtered further to facilitate DADA2 (Callahan et al., 2016) error modeling for amplicon sequence variants (ASVs). All reads with ambiguous bases (Ns), reads with bases with quality scores below ten, and all reads with a cumulative expected error greater than three for merged reads or greater than four for forward reads were filtered. The filtering parameters were chosen as such to refine error modeling for the DADA2 algorithm downstream of the analysis by improving read quality while retaining as many reads as possible. Any unusually short (<240 nt) and chimeric ASVs were removed as well.

Unique ASVs having more than two reads were searched against NCBI's nucleotide database (blastn) to identify the source organism. Results were restricted to the top ten sequences producing significant alignments and limited to records that include Viridiplantae (taxid:33090). Each ASV was assigned a species based on the top blast hit. To further refine the taxa assigned to the ASVs, ASVs with top blast hits that satisfied at least one of the following criteria were filtered out for low quality: i) percent identity in the alignment <85, ii) alignment length less than 1/3 of the length of the ASV (query) iii) alignment is less than 150 bp long, iv) blast bit score <200, and v) species name included a match to "environmental sample" or "*Vachellia jacquemontii"* (a sequencing artifact deposited at NCBI) or was "N/A".

After the assignment of taxa to unique ASVs, the percentage of manuka reads out of total plant reads was calculated to compare the presence and absence of manuka plant (*Leptospermum scoparium)* across 21 samples. Samples showing the presence of *Leptospermum scoparium* were selected to plot the proportional relative abundance of sequences identified at genus level obtained from pollen DNA of true manuka samples as a bar plot. All counts were aggregated based on the taxonomic label's genus and expressed as fractions of the total sample size in the bar plot.

3.3. Results and Discussion

3.3.1. Isolation of pollen DNA and amplification of ITS2

The concentrations of pollen DNA template isolated from 21 samples were found to be 0.5 to 3 ng/µl, with DNA purity ratio A260/A280 between 1.30 to 1.50. The plant ITS2 region was successfully amplified from pollen DNA isolated from twenty-one honey samples using a primer pair published by Chen *et al.* (Chen et al., 2010). The ITS2 region in plants

varies from ~180-390 bp (Moorhouse-Gann et al., 2018; Timpano et al., 2020). The forward and reverse primers anneal in the conserved regions of 5.8S (~85 bp upstream of ITS2) and 26S (~142 bp downstream of ITS2). Most of the amplified ITS2 products showed a faint band of 500 bp with a smear of other DNA products. This smearing of samples might be due to the poor quality of isolated plant gDNA from pollen by the standard protocol.

We tried to improve the quality of isolated plant gDNA and amplification ITS2 by employing different strategies. We tried to re-isolate plant gDNA from fresh pollen sample, using three different DNA extraction kits from Qiagen: DNeasy plant mini kit, QIAamp DNA blood mini kit, and QIAamp DNA stool kit. We also tried a different ITS2 primer pair as published by Prosser *et al.* (Prosser & Hebert, 2017), in combination with another high-fidelity DNA polymerase (Platinum™ SuperFi II PCR Master Mix sold by Thermo Fisher Scientific). No improvement in the quality of PCR product was obtained. Of note, a response report by Ministry for Primary Industries (MPI) of New Zealand stated that the presence of high levels of 5-hydroxymethylfurfural (HMF) in honey could also interfere in the isolation of the plant DNA marker and subsequent failure of PCR reaction (MPI, 2017c). HMF is formed from the degradation of sugar through the Maillard reaction during food processing or long storage of honey (Shapla et al., 2018) and is often used as indicator of honey quality and age. So, we concluded that high HMF content of manuka samples could be one possible reason in isolating the plant gDNA. The yield of ITS2 PCR products obtained from pollen DNA of the twenty-one samples varied between 30-70 ng/µl.

3.3.2. Sequencing analysis

Raw reads obtained from NGS of twenty-one samples varied between 5,000-106,000 reads per sample. After applying the QC filtering parameters as discussed above (Section 3.2.5), we detected ITS2 reads matching manuka (*Leptospermum scoparium*) in nine out of 21 samples, as a varying fraction of total reads (Table 3.3). We set 0.001% cut off i.e., the sample should have at least one manuka read out of 1000 reads (after QC filtering) of total DNA to identify the sample as a true manuka sample.

Table 3.3. Total reads vs reads of manuka plants in 21 different manuka samples

NGS code/ Sample	Manuka reads	Total reads	Percentage of Manuka reads (%)	Approximate Cost ($) / (g)
H89	379	8884	4.27	0.1
H88	3	148	2.03	0.29
H86	469	24329	1.93	0.11
H90	3	852	0.35	0.1
H83	3	1033	0.29	0.08
H105	12	4175	0.29	0.06
H150	1	1311	0.08	0.08
H87	2	6542	0.03	0.1
H85	1	9598	0.01	0.12
H19	0	87499	0.00	0.07
H42	0	2473	0.00	0.07
H79	0	13865	0.00	0.03
H80	0	7292	0.00	0.18
H81	0	3582	0.00	0.06
H82	0	605	0.00	0.26
H84	0	394	0.00	0.07
H109	0	2491	0.00	0.03
H111	0	69	0.00	0.2
H117	0	2586	0.00	0.1
H118	0	224	0.00	0.16
H123	0	167	0.00	0.34

The highest fraction of manuka reads was detected in sample H89, with 4.27% (379/8884) of reads originating from manuka. Samples H86 had about 1.93% reads originating from manuka. Samples H83 and H105 had 0.29% reads from manuka but varying total reads. The smallest fraction of manuka reads (1/9598) was observed in sample H85. Sample H88 had just 3 reads originating from manuka with 148 total reads.

Sample H19 had a total of 87,499 reads passing filtering criteria, but none of these reads matched manuka honey. Instead, we found reads matching kanuka plant (*Kunzea ericoides*), a species closely related to manuka, putting into question the origin of such honey sample labeled as manuka honey. Sample H84 also contained reads matching kanuka plant (*Kunzea ericoides*) but was labeled as manuka honey.

Samples such as H79 and H80 had more than 5000 total reads, but no reads of manuka were observed. On the other hand, samples H82, H84, H111, H118, and H123 had total reads between only 69 and 605 with no manuka reads. Such samples need to be carefully examined to avoid false classification of a sample as a fake manuka sample.

The nine samples (H83, H85, H86, H87, H88, H89, H90, H105, and H150) showing manuka reads were selected for creating a bar plot to study the biodiversity observed with the manuka plant in those honey samples. We found approximately 102 different genera in these nine manuka samples, including *Leptospermum*. We then selected the top twenty genera from these 101 to plot proportional abundances of the plant DNA sequences identified at the genus level DNA along with *Leptospermum* sample set (Figure 3.2).

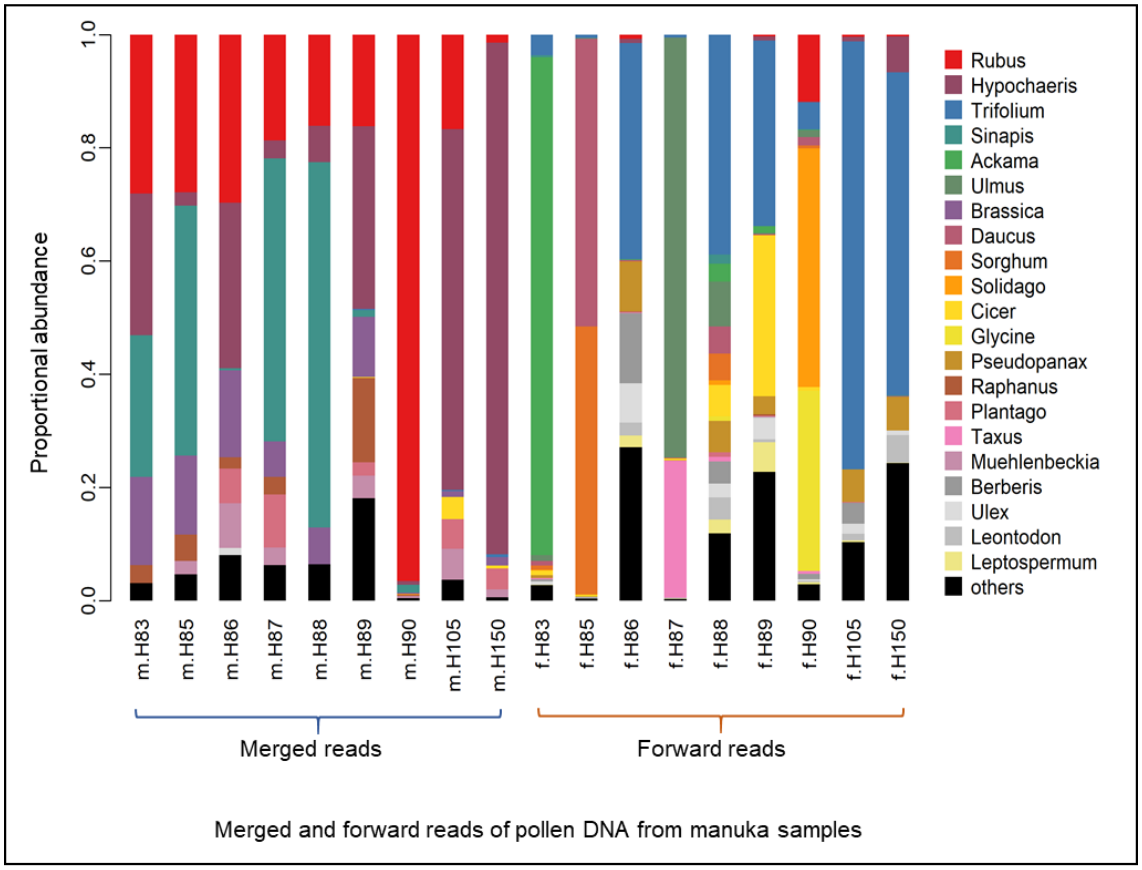


Figure 3.2. Bar plot indicating proportional richness and abundance of top twenty genera along with Leptospermum obtained from NGS of pollen of nine manuka samples. Columns correspond to sample sets of nine manuka honeys. Columns starting with "m" are the merged reads, whereas columns starting with "f" are forward reads.

Genus *Ackama*, dominantly found in New Zealand and Australia, was found in six (forward reads of H83, H86, H88, H89, H90, and H150) out of nine samples. Genera of cosmopolitan plants like *Rubus* (berries), *Hypochaeris*, *Trifolium* (clover), *Daucus*, *Solidago*, and *Muehlenbeckia* are also known to occur to New Zealand either as native or introduced species (POWO, 2021).

4. Neutral DNA-avidin nanoparticles as ultrasensitive reporters in immuno-PCR

The contents of this chapter is published as Chavan, D., Chen, H., Crum, M., Vu, B., Safari, M., Smith, M., Vekilov, P., Conrad, J. C., Kourentzi, K., & Willson, R. C., (2020). Neutral DNA-avidin nanoparticles as ultrasensitive reporters in immuno-PCR. Analyst, 145(14), 4942–4949. https://doi.org/10.1039/d0an00134a.

4.1. Introduction

The need for ultrasensitive protein detection has challenged the scientific community for many years, with a notable advance being the introduction of radio-immunoassay by Berson and Yalow in 1959 (Berson & Yalow, 1959). The gold standard for detecting protein molecules has been ELISA (enzyme-linked immunosorbent assay) in which an analyte is captured on the surface of a microplate well by immobilized antibodies and recognized by an antibody conjugated to a signal-generating enzyme reporter. Various technical innovations (e.g., miniaturization (Roman et al., 2011), single-molecule counting (Rissin et al., 2010), microfluidics and automation (Ye et al., 2018), engineered reporters (Corrie & Plebanski, 2018; Cui et al., 2018), and substrates (Simon & Ezan, 2017)) have improved the performance of immunoassays. Of particular note is immuno-PCR (iPCR; introduced by Sano et al. in 1992 (Sano et al., 1992)), which combines the versatility and specificity of antibody recognition in immunoassays with the exponential signal-amplifying power of PCR, promising a wide dynamic range and dramatically-enhanced sensitivity (Greenwood et al., 2015). Immuno-PCR uses an antibody conjugated to an amplifiable DNA reporter which can be detected very sensitively by PCR, but its great promise has been compromised by various

technical difficulties (Adler et al., 2008). First, naked DNA molecules non-specifically bind to various surfaces (Cai & Yang, 2002; Kan et al., 2015; Shi et al., 2015) and biomolecules (Afek & Lukatsky, 2012; Ganguly et al., 2012; Sun et al., 2016), increasing iPCR background signal. Second, iPCR requires often-complicated preparation of specific DNA-antibody conjugates (Chen et al., 2018; Greenwood et al., 2015; Zhang et al., 2013). To address these challenges, a variety of alternative biological or chemical nanostructures, including liposomes (He et al., 2012) and bacteriophage virus nanoparticles (Brasino & Cha, 2017; Litvinov et al., 2014; Liu et al., 2014) have been explored in an effort to "shield" the DNA reporters and reduce non-specific binding.

We have previously explored M13 bacteriophage as a reporter in iPCR (Litvinov et al., 2014). Although the no-target background was greatly reduced, we found weak dependence of the signal on analyte concentration, likely due to steric interference among the large viral particles. Another drawback to using naturally-occurring DNA reporter sequences (e.g., M13 gDNA) in iPCR assays is their possible adventitious presence in biological samples. Inspired by an alternative immunoassay reporter with low nonspecific binding, a protein-DNA core-shell nanoparticle (Abud et al., 2019; Morpurgo et al., 2004, 2012; Pignatto et al., 2010), in which avidin and polyethylene glycol (PEG) are used to condense and stabilize plasmid DNA, we incorporated multiple, *de novo* designed, repetitive PCR templates into the plasmid DNA and enhanced the PCR detectability of these custom-designed nanoparticles (Figure 4.1). We also demonstrated the use of these custom-designed iPCR reporter nanoparticles in the detection of human chorionic gonadotropin (hCG), a glycoprotein hormone and a novel biomarker for pregnancy (Canfield et al., 1987) and

testicular cancer (Lempiäinen et al., 2008). We were able to quantitate hCG at concentrations as low as 660 fM using our iPCR reporter nanoparticles and standard laboratory equipment.

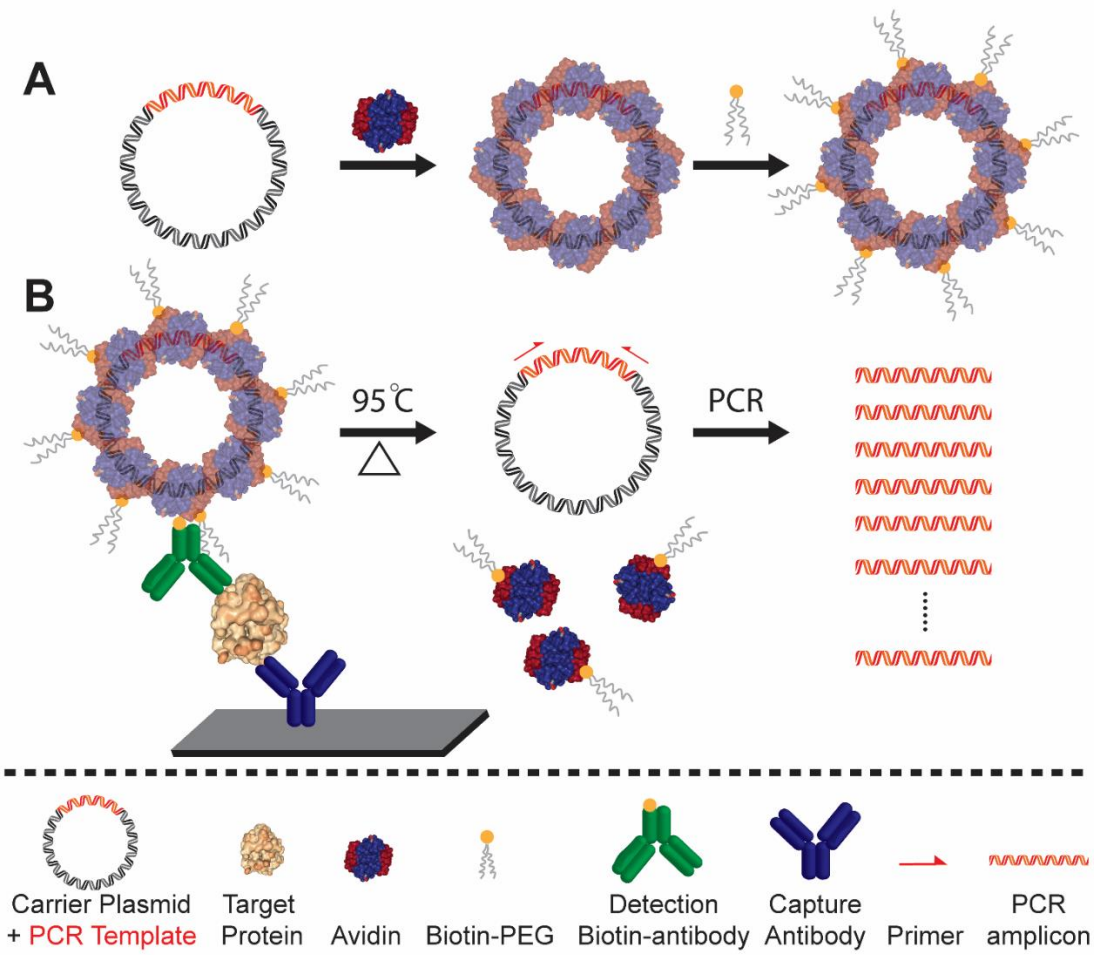


Figure 4.1. Schematic of the immuno-nanoparticle PCR assay. A) Assembly of DNA-avidin core-shell nanoparticles. DNA plasmids carrying the synthetic PCR template are sequentially assembled with avidin and biotin-polyethylene glycol (PEG). B) Workflow of immuno-nanoparticle PCR. Target protein molecules are captured by a capture antibody and detected with nanoparticles via a DTT-cleavable-biotin-linked detection antibody. The captured nanoparticles are disassembled by heat to expose the PCR template for PCR amplification (not to scale).

4.2. Materials and Methods

4.2.1. Reagents

Synthetic DNA was from Integrated DNA Technologies, Inc. (Coralville, Iowa). Avidin (434401), Pierce™ premium grade Sulfo-NHS-SS-Biotin (PG82077), 4'-hydroxyazobenzene-2-carboxylic acid (HABA, 28010), Zeba™ spin desalting columns (40 K MWCO, 0.5 mL, 87766), Dithiothreitol (DTT, R0861), and MediSorp clear flat-bottom immuno nonsterile 96-well plates, 400µL, (467320) were purchased from ThermoFisher Scientific. Two-arm PEG-Biotin (10 kDa, PG2A-BN-10K) was from Nanocs (Boston, Massachusetts). Amicon ultra-0.5 centrifugal filter unit (100 kDa, UFC510096), bovine serum albumin (BSA, A7906), and human chorionic gonadotropin (hCG; CG10-1VL, using the conversion factor 9.28 IU/µg from the 3rd International Standard) were from Millipore Sigma (Burlington, Massachusetts). Healthy human (male) serum was obtained from Gulf Coast Regional Blood Center, Houston, Texas 77054. Bovine serum albumin (IgG free, BSA-BAF-SMP) from Rocky Mountain Biologicals, Inc. (Missoula, Montana). Anti-hCG beta chain mAb, clone 2 (monoclonal, ABBCG-0402), and Goat anti-hCG alpha chain (polyclonal, ABACG-0500) were from Arista Biologicals, Inc. (Allentown, Pennsylvania). Phosphate-buffered saline (PBS) tablets, pH 7.4, were from Takara Bio USA Inc. (Mountainview, CA). Tween® 20, Molecular Biology Grade (H5152) was from Promega (Madison, Wisconsin). Mx3000P optical strip tubes (401428), Mx3000P optical strip caps (401425), and Brilliant III ultra-fast SYBR QPCR master mix (600882) were from Agilent Technologies, Inc. (Santa Clara, California).

A synthetic DNA template and primers were designed as previously reported (Chen et al., 2018). Briefly, the 79-bp synthetic template 5'-

TGCTGCGAGAGTATTATCTTGCACCTTATGCTACCGTGATTCATCCAGTCTCATCG TGAAACAGACGTACTACTACCTG-3' and the 20 nt primers were designed for both minimum similarity to any reported natural DNA sequence and optimal PCR conditions with high annealing temperature (60℃) and short extension time (30 sec). DNA primers were (Forward: 5'-CAGGTAGTAGTACGTCTGTT-3', Reverse: 5'-GTGCTGCGAGAGTATTATCT-3'). QIAprep Spin Miniprep Kit (27106) was from Qiagen Inc (Germantown, Maryland).

4.2.2. Construction of multi-template plasmid DNA

Plasmids containing one to seven repeats of the specific 79-bp PCR target were constructed in pBC, a cloning plasmid with high copy number and chloramphenicol resistance for easy preparation and selection, as follows. For the template, the following oligos were annealed to generate the specific dsDNA template. Oligo F: 5'-TCAGGTAGTAGTACGTCTGTTTCACGATGAGACTGGATGAATCACGGTAGCATAA GGTGCAAGATAATACTCTCGCAGCAGAGCTCATGTCACCAT-3' and Oligo R: 5'-CTAGATGGTGACATGAGCTCTGCTGCGAGAGTATTATCTTGCACCTTATGCTACC GTGATTCATCCAGTCTCATCGTGAAACAGACGTACTACTACCTGAAGCT-3'.

Once annealed, the dsDNA template has, on its R oligo's 5' end, four overhang nucleotides (5'-GATC-3') that complement the *Xba*I restriction site, and on its 3' end, four overhang nucleotides (3'-AGCT-5') that complement the *Sac*I restriction site. However, this 3'end is not recognized by the *Sac*I enzyme since the restriction site was mutated during construction (AAGCT instead of GAGCT; this T is shown in red bold). In addition, nine

nucleotides upstream of the *Xba*I site, a non-altered *Sac*I restriction site, was placed (Figure 4.2).

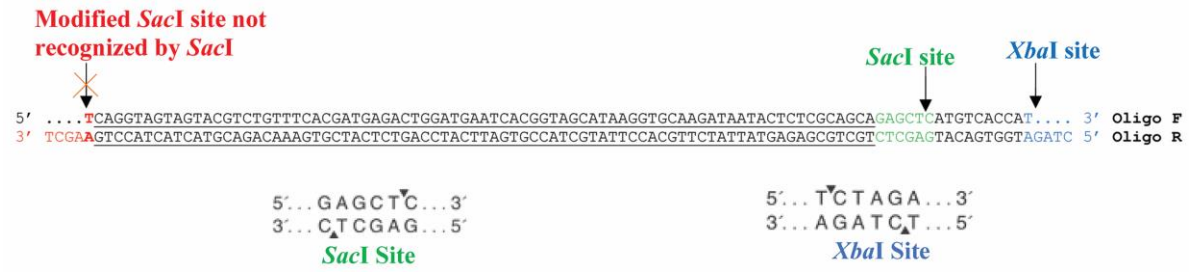


Figure 4.2. Nucleotide sequence of the annealed dsDNA template and restriction enzyme recognition sequences. The cohesive end annealing to the compatible SacI site on the digested plasmid pBC is shown in red. Once ligated to the plasmid, the recognition sequence for SacI is abolished by a thymine-to-cytosine nucleotide change (shown in red bold). The cohesive end which anneals to the compatible, conserved XbaI site on the digested plasmid pBC is shown in blue. Shown in green is an extra SacI recognition sequence included in the dsDNA template to be used for the introduction of additional template repeats. The 79-bp PCR reporter sequence is shown in black and underlined.

The pBC plasmid was linearized with *Sac*I and *Xba*I enzymes, then mixed with and ligated to the dsDNA template. To introduce the second repeat of the target sequence, the plasmid carrying one repeat of template was linearized with *Sac*I and *Xba*I enzymes and again ligated to the dsDNA template. This was done sequentially until all 7 repeats were inserted into the plasmid. Plasmids were transformed into *E. coli Top10 F'* chemical competent (F'[lacI^q Tn10(tet^R)] mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 deoR nupG recA1 araD139 Δ(ara-leu)7697 galU galK rpsL(Str^R) endA1 λ$^-$). The size of the plasmid with one repeat of the template was 3,500 bp, and the size of plasmid increased by 85 bp for every additional repeat of the template introduced. We have deposited the plasmid with four repeats of the PCR target into the Addgene repository (#127380).

E. coli cells harboring the plasmids were grown in LB broth supplemented with 25 µg/ml chloramphenicol at 37 ^{o}C for 14 h with shaking at 200 rpm. Plasmids were then isolated using the QIAprep Spin Miniprep Kit and were eluted in 10 mM Tris·Cl, pH 8.5 as per manufacturer's instructions.

Plasmid DNA was further purified by ethanol precipitation to reduce the ionic strength of the DNA solution, as salts could interfere with assembly of nanoparticles. To precipitate the DNA, 20 µl of 3M sodium acetate (pH 5.2) and 400 µl of ice-cold molecular biology grade absolute ethanol were added to 200 µl of plasmid DNA. The mixture was incubated at -20 ^{o}C for 1 h and centrifuged at 14,550 *g* for 15 min, and the supernatant was discarded. The pellet was then washed with 500 µl of 70% ethanol and allowed to air-dry with the tube inverted. The DNA pellet was finally resuspended in 50 µl sterile water (typically to 160-200 ng/µl) and stored at -20 ^{o}C until use.

4.2.3. Construction of DNA-avidin nanoparticles

For construction of DNA-avidin nanoparticles coated with PEG-biotin, plasmid DNA was first diluted with deionized water in a sterile microcentrifuge tube to $1x10^{12}$ DNA copies/ml based on Nanodrop A260 absorbance value. The volume of avidin and PEG-biotin mixed with the plasmid DNA was varied with the size of plasmid as discussed below. Both DNA and avidin pre-diluted in deionized water (6.1 mg/ml) were kept on ice for 15 min and plasmid DNA was then added to the avidin at a ratio of one avidin molecule for every 4 bp DNA, after which the volume was adjusted to 980 µl using deionized water. The sample was immediately vortexed for 30 sec and then allowed to incubate for 1 h at room temperature on a rotator (40 rpm, Model #RT50, Cole-Parmer, Vernon Hills, Illinois). After incubation, 10

kDa 2-arm PEG-biotin (4 mg/ml) in deionized water was added to the DNA-avidin mixture. PEG-biotin offered was 30% of the avidin biotin binding sites (7.8×10^{14} and 8.4×10^{14} biotin binding sites/ml for 1×10^{12} DNA copies/ml of plasmid 1 and plasmid 4, respectively). The final volume of the mixture was adjusted to 1 ml using deionized water. The mixture was incubated for 24 h at 4 ^{o}C on a rotator, then divided between two Amicon Ultra-0.5, 100 kDa membrane filters and centrifuged at 10,000 *g* for 15 min to remove free avidin and PEG-biotin. Approximately 20 µl of DNA-avidin nanoparticle suspension was then recovered per filter by centrifugation at 1,000 *g* for 2 min, and the final volume was made up to 50 µl using deionized water. The two samples were pooled to give 100 µl stock solution of reporter particles and was stored at 4 ^{o}C.

As an example, to construct DNA-avidin nanoparticles based on the plasmid DNA with 4 repeats of template (3,755 bp), 25 µl of plasmid DNA (167.2 µg/ml, 4×10^{13} DNA copies/ml) was mixed with 17 µl of avidin (6.1 mg/ml) and 3.5 µl of PEG-Biotin (4 mg/ml) and the final volume was adjusted to 1 ml using deionized water. This 1 ml particle suspension, after 24 h incubation, was then filtered using two Amicon membrane filters (100 kDa) as described above, and finally pooled together to give 100 µl of stock solution of Particle 4.

4.2.4. qPCR of plasmid DNA and DNA-avidin nanoparticles

Plasmid DNA or DNA-avidin nanoparticles were serially diluted in sterile water from 5 million DNA copies/reaction to five DNA copies/reaction and subjected to qPCR (Agilent Mx3005P qPCR System). For setting up qPCR reactions, 10 µl of template was mixed with 10 µL 2x qPCR Master Mix (containing 1 µM primers) and DNA was amplified using the

following PCR conditions: 1 cycle at 95 °C for 10 min, then 50 cycles of 95 °C for 15 sec and 60 °C for 30 sec (Chen et al., 2018).

4.2.5. Nanoparticle Tracking Analysis (NTA)

The size and concentration of DNA-avidin nanoparticles present in the filtered stock solution were determined using nanoparticle tracking analysis (NTA). The filtered stock solution of DNA-avidin nanoparticles was diluted 100 times with sterile 0.22 µm- filtered DI water. A NanoSight LM14 microscope (NanoSight Ltd) equipped with a 20x lens (NA. 0.4) and a 532 nm laser was employed to monitor the diffusional mobility of individual DNA–avidin nanoparticles at 10 °C. Approximately 500 µl of 100X diluted DNA–avidin nanoparticle solution was injected into the NanoSight cuvette with holdup volume of 300 µl. To prevent overheating of samples due to laser irradiation, data was collected within 5 min of sample injection. A sensitive CMOS camera (Model C11440-50B, Hamamatsu Photonics K.K., Japan) supplied by NanoSight was used to record the particles' Brownian motion at 24 fps. The accompanying software package (NTA. Version 2.3 Build 0025) was used to generate particle trajectories. Three movies showing distinct fields of view were collected for each sample for over 30 s using manual shutter and gain adjustments.

4.2.6. Zeta potential

The zeta potential of plasmid DNA and DNA-avidin nanoparticles diluted in sterile water to 5×10^7 DNA copies/ml and 5×10^7 nanoparticles/ml was determined by running 10 cycles at 23 ºC using a Nicomp 380 ζ-potential analyzer, calibrated using a zeta potential transfer standard (-42 mV ± 4.2 mV, Malvern Instruments Ltd., UK).

4.2.7. Antibody biotinylation

Polyclonal Goat anti-hCG alpha chain antibody was mixed with Pierce™ premium grade Sulfo-NHS-SS-Biotin (DTT-cleavable biotin) at a 1:20 mole ratio and incubated on ice for 2 h. Biotinylated antibody was then separated from unbound biotin using Zeba™ spin desalting columns (40K). Using the HABA assay the biotinylation ratio was determined to be between 4.11-5.33 biotin molecules per antibody. Biotinylated antibody was stored in PBS (pH 7.4) at 4 ^{o}C. The DTT-cleavable biotinylated detection antibody allowed release of the bound DNA-avidin nanoparticles in the PCR assay, as described below. We used a similar DTT-cleavable biotinylated detection antibody in a previous publication (Chen et al., 2018).

4.2.8. DNA-avidin nanoparticle-based iPCR

Wells of a 96 well plate were charged with 100 µl of 10 µg/ml anti-hCG beta chain monoclonal antibody in PBS, pH 7.4, incubated overnight at 4 ^{o}C, blocked with 300 µl PBS containing 3% BSA for 2 h at 25 ^{o}C, and washed thrice with PBS + 0.1% Tween 20 (HydroFlex microplate washer, Tecan, Co., Männedorf, Switzerland). 100 µl of hCG diluted from 10 ng/ml to 1 pg/ml in PBS containing 1% BSA was added to the wells (triplicates) and incubated for 1.5 h at 25 ^{o}C. For no-hCG control (triplicates), 100 µl of PBS containing 1% BSA was added in the wells and incubated for 1.5 h at 25 ^{o}C. The remining assay steps were the same for all wells, as described below.

Wells were then washed thrice with PBS + 0.1% Tween 20. 100 µl of 100 ng/ml biotinylated detection antibody (anti-hCG alpha chain Ab conjugated with DTT-cleavable biotin) was added and incubated for 1.5 h at 25 ^{o}C. Wells were washed thrice with PBS containing 0.1% Tween 20. 100 µl of DNA-avidin nanoparticles based on plasmid with four

copies of template ("Particle 4"; 5×10^7 particles/ml) diluted in PBS + 2% BSA was added and incubated overnight at 4 ^{o}C. Wells were washed five times with PBS + 0.1% Tween 20. Bound particles were released by adding 100 µl 50 mM DTT and incubated for 2 h at 25°C (as per manufacturer's instructions). 10 µl sample from each well was mixed with 10 µL 2x qPCR Master Mix (containing 1 µM primers) and DNA was then amplified using qPCR (1 cycle at 95 °C for 10 min, then 50 cycles of 95 °C for 15 sec and 60 °C for 30 sec).

4.3. Results and Discussion

4.3.1. Design and preparation of multi-template PCR-amplifiable DNA reporter

Using the NCBI Nucleotide Blast (megablast algorithm, optimized for highly similar sequences), no significant similarity was found between the NCBI nucleotide collection database and our 79-bp synthetic DNA template, indicating that our *de novo* designed PCR amplifiable DNA reporter is not present in any sequenced biological specimen. We then constructed plasmid DNAs containing one to seven repeats of this reporter and tested the hypothesis that including more repeats of the amplifiable reporter template in the plasmid DNA would increase detectability. As shown in Figure 4.3, C_t decreased substantially with each additional repeat from one to four, with diminishing returns for additional repeats up to seven. The C_t value of 5×10^5 DNA copies of plasmid with one repeat and four repeats of template corresponded were 18.22 and 13.26 respectively. Based on these results we chose the plasmids that contain one and four repeats of the template for construction of the DNA-avidin reporter nanoparticles.

We confirmed the cloning of the repeats of the reporter sequence in the pBC plasmid by gel electrophoresis of the pBC plasmids containing one to seven repeats of the reporter

sequence, digested with restriction enzymes *Kpn*I and *Bpm*I. We also confirmed the insertion of the repetitive sequences by Sanger sequencing of the various plasmids containing the repeats with primers M13R and M13F.

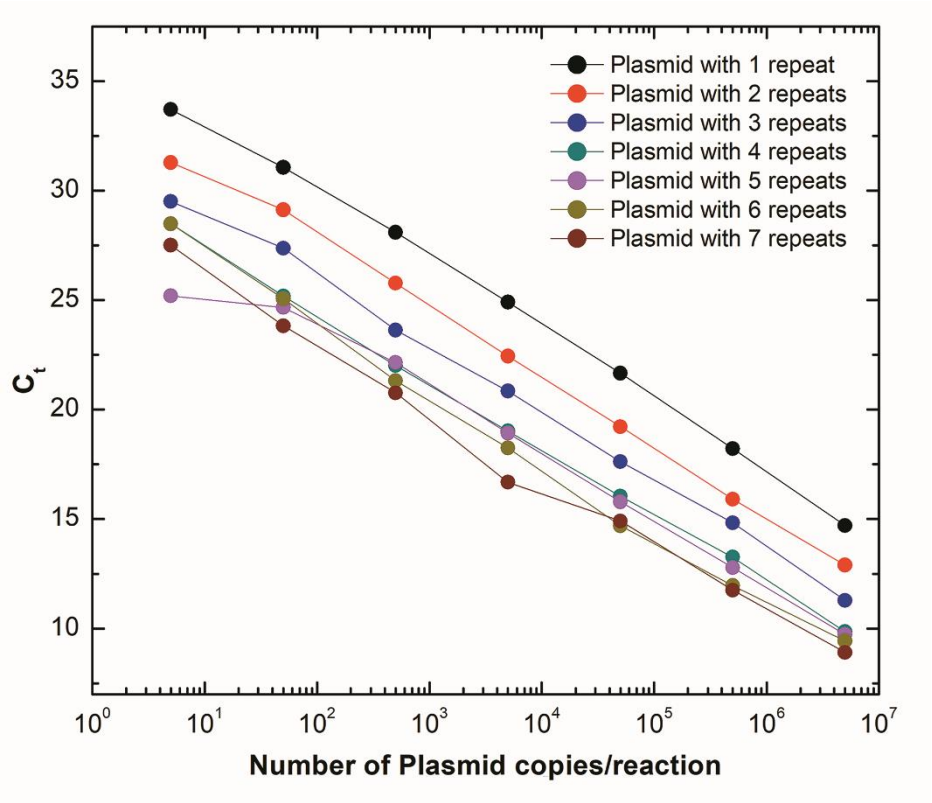


Figure 4.3. qPCR standard curves of plasmid DNA constructs containing one to seven repeats of target template. The plasmid with no template showed C_t over 35.

4.3.2. Construction and Characterization of DNA-avidin nanoparticles

Polycationic compaction agents, (e.g., spermidine (Gosule & Schellman, 1976)) bind the major or minor grooves of dsDNA, neutralizing its charge and reducing its volume by four to six orders of magnitude (Golan et al., 1999). In vivo, they function to package genomic DNA, e.g., into sperm (Balhorn, 1982). We have previously reported the use of compaction agents such as spermine and spermidine for the condensation and selective purification of DNA (Murphy et al., 2003; Vu et al., 2012). Avidin, a 68-kDa, very cationic biotin-binding

glycoprotein found in chicken egg white (Green, 1990), has been shown to condense DNA through high-affinity interactions with DNA without impairing avidin's biotin-binding activity, leading to the formation of 120 nm nanoparticles of toroidal shape (Morpurgo et al., 2004). Selective PEGylation through the available biotin-binding sites in avidin further stabilizes the nanoparticles, allowing their use as ELISA-based immunodetection reporters with low non-specific binding (Morpurgo et al., 2012; Pignatto et al., 2010).

Our DNA-avidin nanoparticles were constructed through condensation and neutralization of plasmid DNA containing 1 ("Particle 1") or 4 ("Particle 4") repeats of target template with avidin, and coating with PEG-biotin. We mixed the plasmid DNA containing one or four repeats of template with avidin such that there was one avidin molecule present for every 4 bp of DNA. This was followed by the addition of a hydrophilic 2-arm, 10 kDa biotin-PEG polymer such that the PEG-biotin occupied 30% of avidin biotin-binding sites. These DNA-avidin nanoparticles were found to be stable for 6 months when stored in water at 4 ^{o}C. As shown in Figure 4.4, the C_t values of 5×10^5 copies of Particles 1 and 4 were 20.79 and 14.64, respectively. Samples with 5×10^5 DNA copies of plasmid DNA with 4 repeats of template or of Particle 4 derived from that plasmid gave similar C_t values (13.26 and 14.64, respectively), suggesting that one nanoparticle contained one condensed plasmid DNA molecule.

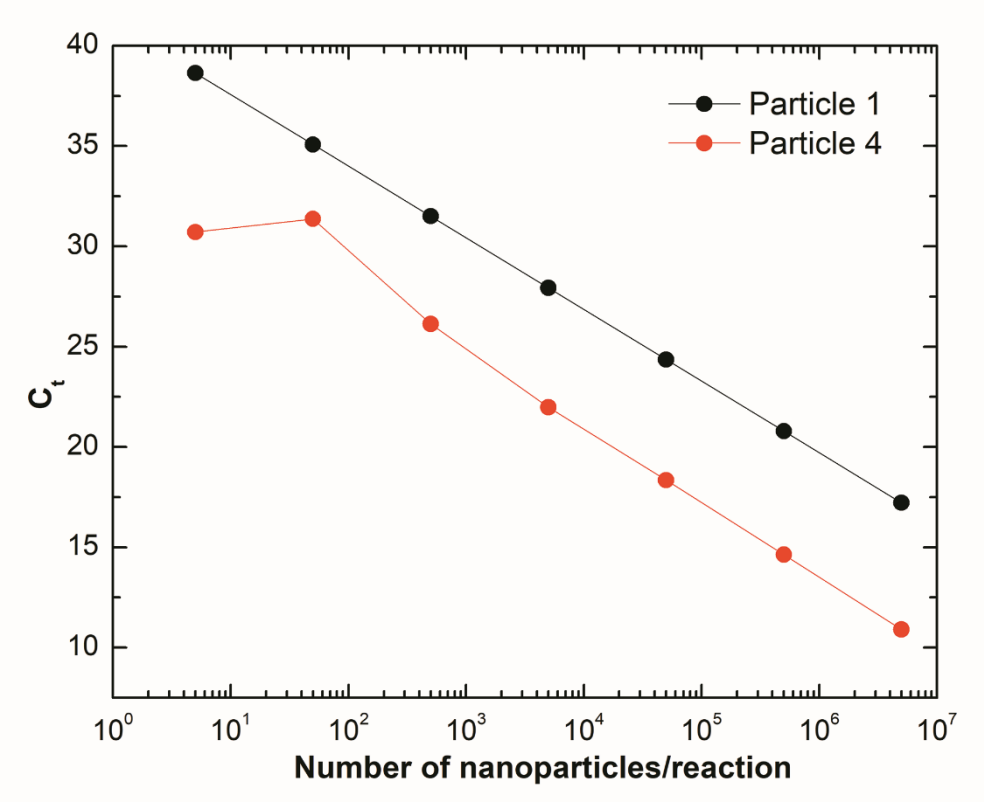


Figure 4.4. qPCR standard curves of DNA-avidin nanoparticles (n=1). Particle 1 (solid black) and Particle 4 (solid red)

4.3.3. Characterization of DNA-avidin nanoparticles

A. Nanoparticle tracking analysis (NTA)

Nanoparticle sizes and concentration were estimated from the collected nanoparticle tracking analysis movies for Particle 1 and Particle 4. As shown in Figure 4.5, the average sizes of Particle 1 and Particle 4 were found to be 109 ± 3.8 nm, and 95 ± 3.7 nm respectively. The approximate undiluted stock concentrations of Particles 1 and 4 were found to be 6.7×10^{10} particles/ml and 7.2×10^{10} particles/ml.

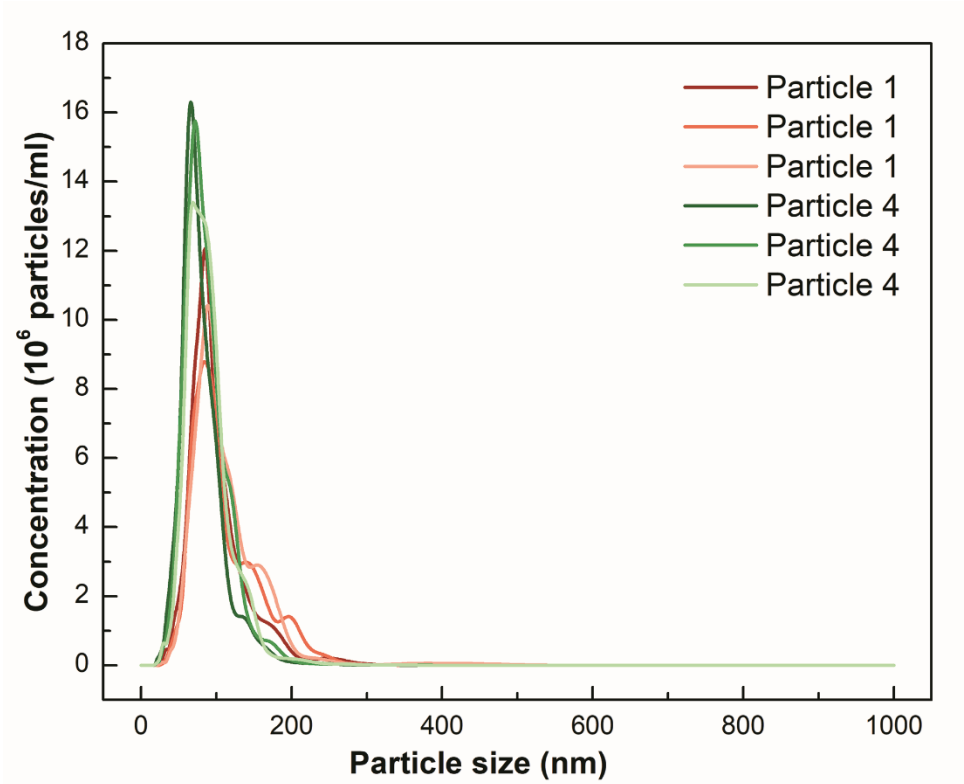


Figure 4.5. Nanoparticle tracking analysis of DNA-avidin nanoparticles. Three distinct fields of view were observed for calculating size and concentration of both Particle 1 and Particle 4. Curves shown in shades of red correspond to Particle 1 stock solution diluted 100-fold with water. The average size and undiluted concentration of Particle 1 were found to be 109 ± 3.8 nm and 6.7×10^{10} particles/ml respectively. Curves shown in shades of green correspond to Particle 4 stock solution diluted 100-fold with water. The average size and undiluted concentration of Particle 4 were found to be 95 ± 3.7 nm and 7.2×10^{10} particles/ml respectively.

B. Zeta Potential

The zeta potential of Particle 4 was found to be 1.17 ± 1.4 mV in contrast to plasmid DNA with 4 repeats of template (-5.2 ± 2.4 mV), indicating avidin largely neutralized the negatively charged phosphate groups of DNA.

4.3.4. DNA-avidin nanoparticle-based iPCR assay

We investigated the feasibility of nanoparticle-based iPCR by detecting human chorionic gonadotropin (hCG). hCG is an ideal model protein, commonly used to validate novel assay technologies, as it is extensively studied and many anti-hCG antibodies are commercially available.

To demonstrate detection of hCG, we immobilized monoclonal antibodies recognizing the hCG beta-chain in the wells of a microplate. Biotinylated (DTT-cleavable biotin) detection antibodies (anti-hCG alpha chain mAb) were added, followed by the addition of Particle 4 DNA-avidin nanoparticles. As shown in Figure 4.6, -Delta C_t increased monotonically with the concentration of hCG; we estimated the limit of detection (LOD) at 25 pg/ml (signal higher than the no-hCG control plus three times the standard deviation of the no-hCG control; 660 fM; 100 μl sample volume). The data from iPCR assay of hCG using Particle 4 was then analyzed using a nonlinear regression (logistic regression), five-parameter (5PL) fit to determine the dynamic range of the assay. The dynamic range of protein quantification for the assay was found to be 25 pg/ml to 10,000 pg/ml. The LOD of our iPCR platform (6×10^{-17} moles in 100 μl sample volume or 25 pg/ml) is close to that of a previously published study (Hendrickson et al., 1995) having LOD of 10^{-17} moles in a 50 μl sample volume for hCG.

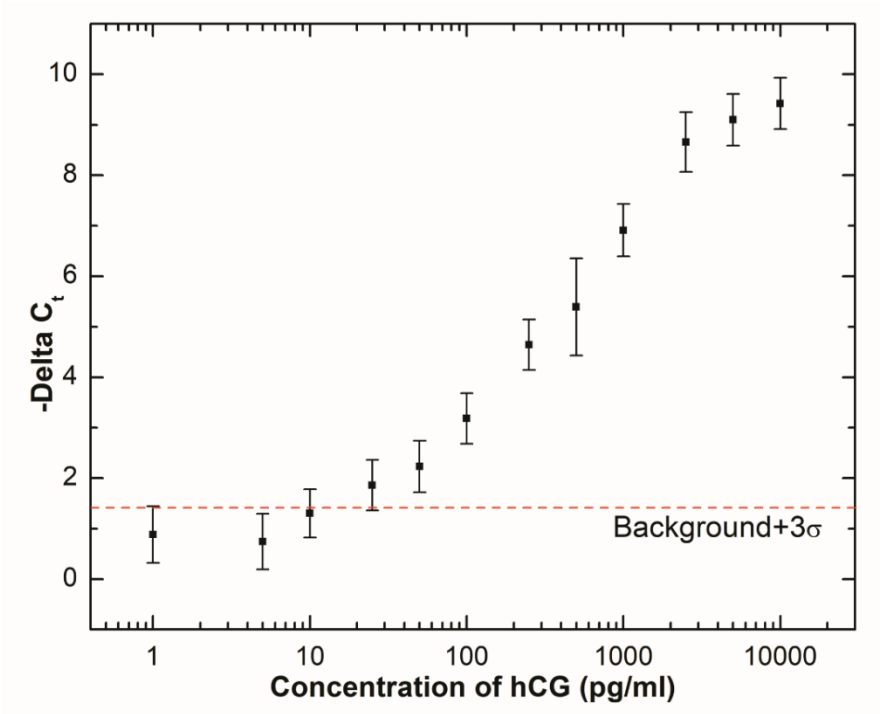


Figure 4.6. Quantification of hCG spiked in PBS +1% BSA using DNA-avidin nanoparticle (with four repeats of template)-based iPCR (n=3, error bars ±1 SD; non-template control gave no C_t). The dashed red line is the detection threshold of the assay, which is defined as the average -delta C_t value of the no-hCG control plus three times the standard deviation of the no-hCG control. A standard approach was used to estimate the Limit of Detection as the lowest analyte concentration that gave a signal clearly distinguishable from the detection threshold. The Limit of Detection was estimated at 25 pg/ml hCG (660 fM).

We then tested the assay with a complex matrix like 25% human serum. We spiked different concentrations of hCG ranging from 10 pg/ml to 1000 pg/ml in 25% human serum (100% serum diluted to 25% in PBS containing 1% IgG-free BSA). We found the limit of detection to be 50 pg/ml for hCG spiked in 25% human serum. We have also compared the reproducibility of our iPCR platform with different batches of Particle 4 as immuno-reporter for detection of hCG spiked in PBS+1% BSA, which showed similar sensitivity (Figure 4.7).

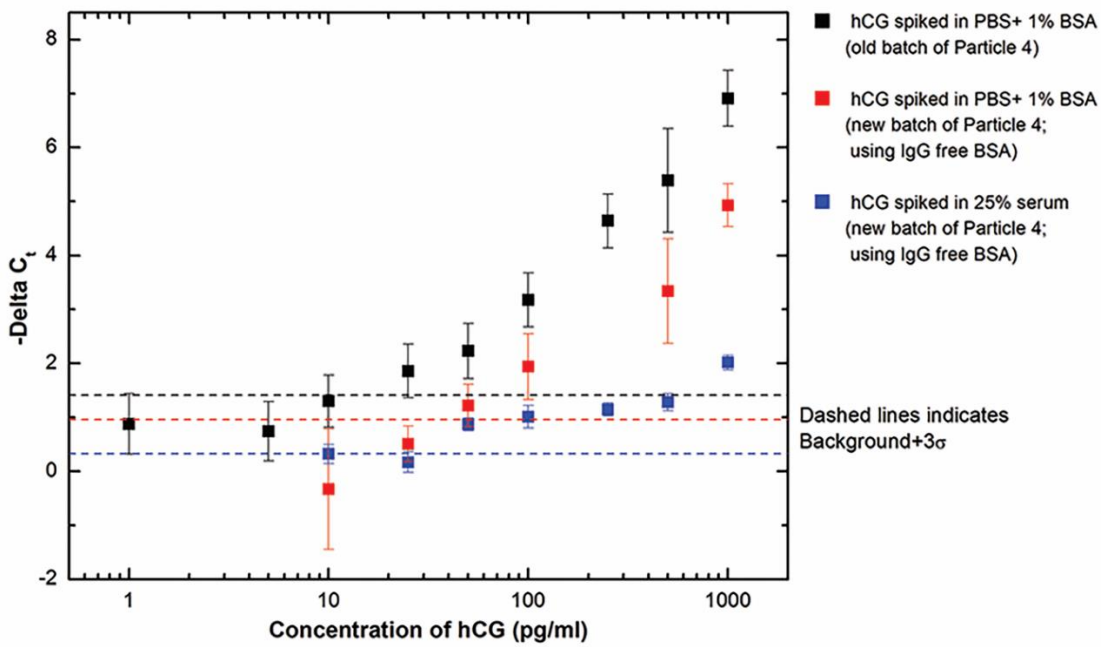


Figure 4.7. Comparison of quantification of hCG spiked in PBS +1% BSA and 25% human serum using DNA-avidin nanoparticle (with four repeats of template)-based iPCR (n=3, error bars ±1 SD; non-template control gave no C_t). Solid black squares (using old batch of Particle 4; data in Figure 4.6) and red squares (using new batch of Particle 4) correspond to hCG spiked in PBS+1% BSA and have LOD of 25 pg/ml and 50 pg/ml, respectively. Alternatively, solid blue squares correspond to hCG spiked in 25% human serum and has a LOD of 50 pg/ml. The dashed lines (black, red and blue) are the detection threshold of the assay for respective conditions described above, which is defined as the average -delta C_t value of the no-hCG control plus three times the standard deviation of the no-hCG control.

5. Conclusions and Future work

5.1. Enrichment and purification of trace pollen-free DNA for authentication of honey

In this work, we have developed and tested three different strategies to enrich and isolate trace pollen-free DNA from honey samples filtered or centrifuged to remove the pollens. The pollen-free DNA captured by the three methods (Q Sepharose, CHT, and anti-dsDNA antibody coupled magnetic particles) was then compared with pollen DNA of the same honey to establish the correlation and authenticity of honey samples. A strong correlation was observed between DNA sequences obtained from NGS of pollen DNA and pollen-free DNA by Q Sepharose on three different commercial honey samples. CHT and anti-dsDNA antibodies showed poor recovery and diversity of captured plant DNA sequences. This might be due to steric interference of large plant genomic DNA resulting in poor capture efficiency. Another reason might be other DNA-binding proteins found in honey that can interfere with the binding of DNA to adsorbents. Country-specific plant DNA sequences (*Erica manipuliflora* and *Quercus coccifera*) were observed in pollen DNA and pollen-free DNA isolated from the honey sample of Greece.

The presence of cosmopolitan plant species such as *Helianthus annus* (sunflower) and *Citrullus sp.* (watermelon) which are commonly cultivated in every country, were also observed in two out of three samples. Such occurrences of common plants can complicate the authentication of honey samples, specifically for honey samples with very low pollen content or comparing samples of two neighboring countries like the USA and Argentina. In such cases, the deep sequencing approach can be applied to a larger quantity of sample. We developed methods using a 15 g honey sample. The probability of capturing trace plant DNA sequences can be increased by increasing the sample size to 100 g.

Some concerns for the method developed in this research can be if samples are spiked with DNA sequences of more lucrative plant in the liquid portion of honey. Then, identifying the true origins of such samples can be difficult. Another, prospective problem can be if honey is first treated with nucleases and then spiked with DNA of favorable plants. In such cases, metagenomic analysis of honey samples can help in providing a more picture of DNA originating from plants, honeybee, fungi, viruses, and the gut bacteria of honeybee (Bovo et al., 2018).

This preliminary work can be further developed to obtain DNA sequences by targeting honey samples of different countries of the world. DNA sequences obtained from such research will increase the richness of the public DNA database and help to link occurrences of source plants worldwide. Thus, by blending efforts in DNA purification and sequencing, we have established techniques that will help mitigate fraud associated with filtered honey and indirectly aid in providing authentic and safe food for consumers.

5.2. Authentication of manuka honey by targeted next-generation sequencing

We report NGS and bioinformatic analysis of 21 manuka samples. We tested the authenticity of these 21 manuka honey by confirming the presence of ITS2 reads corresponding to *Leptospermum scoparium* (a validated DNA marker of manuka honey). Nine out of twenty-one samples passed the stringency set in our study for authenticating manuka samples and showed the presence of manuka reads. One out of nine samples, H88, needs to be carefully examined as it had 148 total reads (after QC filtering) with 3 reads originating from manuka. Genus *Ackama*, pre-dominantly found only in New Zealand and Australia, was found in five (forward reads of H83, H86, H89, H90, and H150) out of nine samples. Such samples can be confidently assigned as true manuka samples as we have the information of other plants known to co-occur with the manuka plant also were observed in these samples.

Manuka samples which failed the stringency needs to be carefully examined to avoid false classification of such samples as fake manuka sample. Samples with total reads between 100-1,000 (after QC filtering) need to be carefully examined to avoid false classification as a fake manuka sample. The same sample can be sequenced multiple times to obtain maximum reads to assign as a true or fake manuka sample.

The methods developed for pollen-free DNA can also be combined here to identify the plant DNA sequences in the liquid portion of manuka honey to identify if the sample was spiked with the pollen of the manuka plant. We believe this NGS-based approach if combined with the current commercial kit used by the Ministry for Primary Industries (MPI) of New Zealand, will help to increase further the standards set for classifying honey as manuka honey.

5.3. Neutral DNA-avidin nanoparticles as ultrasensitive reporters in immuno-PCR

In summary, we have demonstrated an ultra-sensitive iPCR platform using novel ultra-detectable, reduced-nonspecific binding DNA-avidin nanoparticles. The nanoparticles carry multiple repeats of a de novo designed synthetic PCR amplifiable DNA sequence for enhanced detectability and are modified with hydrophilic PEG for reduced non-specific binding, one of the major problems in traditional iPCR formats with naked DNA. Traditional iPCR universal platforms require prior preparation of antibody-DNA oligo conjugates using either thiol-maleimide or biotin-streptavidin chemistry (Chang et al., 2016; Malou & Raoult, 2011). However, covalent conjugation of DNA to antibodies can affect the affinity of the antibody (Brasino & Cha, 2017). Additionally, the tetrameric structure of the avidin and streptavidin results in the generation of a heterogeneous pool of DNA-antibody conjugates, thereby affecting the robustness of iPCR assay (Sano et al., 1993). An additional potential advantage of our iPCR platform based on DNA-avidin nanoparticles is the homogeneity of the reporter nanoparticle, which have a single copy of condensed plasmid DNA with four repeats of the reporter DNA template. Our nanoparticle reporters are relatively easily prepared, and provide a generic, readily-customizable platform for the detection of proteins for which high-affinity antibodies exist. This technology could readily be applied to other protein targets, including microbial antigens (Mehta et al., 2014), cytokines, tumor markers (Assumpção & da Silva, 2016), and anti-drug antibodies induced by biologic therapeutics (Partridge et al., 2016) to precisely quantitate target analytes at ultra-low levels.

6. BIBLIOGRAPHY

Abud, J. E., Santamaría, C. G., Oggero, M., & Rodriguez, H. A. (2019). Methodological aspects of Universal immuno-PCR on standard tubes. *Analytical Biochemistry*, *570*, 56–61.

Adessi, C. (2000). Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Research*, *28*, e87.

Adler, M., Wacker, R., & Niemeyer, C. M. (2008). Sensitivity by combination: Immuno-PCR and related technologies. *Analyst*, *133*, 702–718.

Afek, A., & Lukatsky, D. B. (2012). Nonspecific protein-DNA binding is widespread in the yeast genome. *Biophysical Journal*, *102*, 1881–1888.

Ali, N., Rampazzo, R. de C. P., Costa, A. D. T., & Krieger, M. A. (2017). Current nucleic acid extraction methods and their implications to point-of-care diagnostics. *BioMed Research International*, *2017*, 1–13.

Aliferis, K. A., Tarantilis, P. A., Harizanis, P. C., & Alissandrakis, E. (2010). Botanical discrimination and classification of honey samples applying gas chromatography/mass spectrometry fingerprinting of headspace volatile compounds. *Food Chemistry*, *121*, 856–862.

Álvarez, I., & Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, *29*, 417–434.

Assumpção, A. L. F. V., & da Silva, R. C. (2016). Immuno-PCR in cancer and non-cancer related diseases: a review. *Veterinary Quarterly*, *36*, 63–70.

Balhorn, R. (1982). A model for the structure of chromatin in mammalian sperm. *Journal of Cell Biology*, *93*, 298–305.

Barba, M., Czosnek, H., & Hadidi, A. (2014). Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses*, *6*, 106–136.

Batista, B. L., da Silva, L. R. S., Rocha, B. A., Rodrigues, J. L., Berretta-Silva, A. A., Bonates, T. O., Gomes, V. S. D., Barbosa, R. M., & Barbosa, F. (2012). Multi-element determination in Brazilian honey samples by inductively coupled plasma mass spectrometry and estimation of geographic origin with data mining techniques. *Food Research International*, *49*, 209–215.

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.

Berson, S. A., & Yalow, R. S. (1959). Quantitative aspects of the reaction between insulin and insulin-binding antibody. *Journal of Clinical Investigation*, *38*, 1996–2016.

Boffo, E. F., Tavares, L. A., Tobias, A. C. T., Ferreira, M. M. C., & Ferreira, A. G. (2012). Identification of components of Brazilian honey by ^{1}H NMR and classification of its botanical origin by chemometric methods. *LWT - Food Science and Technology*, *49*, 55–63.

Bolick, M. R., & Vogel, S. (1992). Breaking strengths of pollen grain walls. *Plant Systematics and Evolution*, *181*, 171–178.

Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. L., Wertheim-Van Dillen, P. M., & van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology*, *28*, 495–503.

Bovo, S., Ribani, A., Utzeri, V. J., Schiavo, G., Bertolini, F., & Fontanesi, L. (2018). Shotgun metagenomics of honey DNA: Evaluation of a methodological approach to describe a multi-kingdom honey bee derived environmental DNA signature. *PLoS ONE*, *13*, e0205575.

Brasino, M., & Cha, J. N. (2017). Real-time femtomolar detection of cancer biomarkers from photoconjugated antibody– phage constructs. *Analyst*, *142*, 91–97.

Bryant, V. M., & Jones, G. D. (2001). The R-values of honey: pollen coefficients. *Palynology*, *25*, 11–28.

Burns, D. T., Dillon, A., Warren, J., & Walker, M. J. (2018). A critical review of the factors available for the identification and determination of mānuka honey. *Food Analytical Methods*, *11*, 1561–1567.

Cai, X. E., & Yang, J. (2002). Molecular forces for the binding and condensation of DNA molecules. *Biophysical Journal*, *82*, 357–365.

Callahan, B. J. (2021). *DADA2 ITS Pipeline Workflow (1.8)*. https://benjjneb.github.io/dada2/ITS_workflow.html

Callahan, Ben J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J., & Holmes, S. P. (2016). Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F1000Research*, *5*, 1492.

Callahan, Benjamin J, McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*, 581–583.

Canfield, R. E., O'Connor, J. F., Birken, S., Krichevsky, A., & Wilcox, A. J. (1987). Development of an assay for a biomarker of pregnancy and early fetal loss. *Environmental Health Perspectives*, *74*, 57–66.

Chen, H., Crum, M., Chavan, D., Vu, B., Kourentzi, K., & Willson, R. C. (2018). Nanoparticle-based proximity ligation assay for ultrasensitive, quantitative detection of protein biomarkers. *ACS Applied Materials and Interfaces*, *10*, 31845–31849.

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*, e8613.

Chen, Y., Zhu, Z., & Yu, Y. (2014). Novel methodologies in analysis of small molecule biomarkers and living cells. *Tumor Biology*, *35*, 9469–9477.

Chudzinska, M., & Baralkiewicz, D. (2010). Estimation of honey authenticity by multielements characteristics using inductively coupled plasma-mass spectrometry (ICP-MS) combined with chemometrics. *Food and Chemical Toxicology*, *48*, 284–290.

Corrie, S. R., & Plebanski, M. (2018). The emerging role of nanomaterials in immunological sensing — a brief review. *Molecular Immunology*, *98*, 28–35.

Cui, X., Liu, Y., Hu, D., Qian, W., Tin, C., Sun, D., Chen, W., & Lam, R. H. W. (2018). A fluorescent microbead-based microfluidic immunoassay chip for immune cell cytokine secretion quantification. *Lab on a Chip*, *18*, 522–531.

El-Senduny, F. F., Hegazi, N. M., Abd Elghani, G. E., & Farag, M. A. (2021). Manuka honey, a unique mono-floral honey. A comprehensive review of its bioactives, metabolism, action mechanisms, and therapeutic merits. *Food Bioscience*, *42*, 101038.

Fakhlaei, R., Selamat, J., Khatib, A., Razis, A. F. A., Sukor, R., Ahmad, S., & Babadi, A. A. (2020). The toxic impact of honey adulteration: A review. *Foods*, *9*, 1538.

Farkas, Á., & Zajácz, E. (2007). Nectar production for the Hungarian honey industry. *The European Journal of Plant Science and Biotechnology*, *1*, 125–151.

Galimberti, A., de Mattia, F., Bruni, I., Scaccabarozzi, D., Sandionigi, A., Barbuto, M., Casiraghi, M., & Labra, M. (2014). A DNA barcoding approach to characterize pollen collected by honeybees. *PLoS ONE*, *9*, e109363.

Ganguly, A., Rajdev, P., Williams, S. M., & Chatterji, D. (2012). Nonspecific interaction between DNA and protein allows for cooperativity: A case study with mycobacterium DNA binding protein. *Journal of Physical Chemistry B*, *116*, 621–632.

García, N. L. (2018). The current situation on the international honey market. *Bee World*, *95*, 89–94.

Gaspar, J. M. (2018). NGmerge: Merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinformatics*, *19*, 1–9.

Giovannini, R., & Freitag, R. (2001). Comparison of different types of ceramic hydroxyapatite for the chromatographic separation of plasmid DNA and a recombinant anti-Rhesus D antibody. *Bioseparation*, *9*, 359–368.

Golan, R., Pietrasanta, L. I., Hsieh, W., & Hansma, H. G. (1999). DNA toroids: stages in condensation. *Biochemistry*, *38*, 14069–14076.

Gosule, L. C., & Schellman, J. A. (1976). Compact form of DNA induced by spermidine. *Nature*, *259*, 333–335.

Goux, H. J., Chavan, D., Crum, M., Kourentzi, K., & Willson, R. C. (2018). *Akkermansia muciniphila* as a model case for the development of an improved quantitative RPA microbiome assay. *Frontiers in Cellular and Infection Microbiology*, *8*, 237.

Green, N. (1990). Avidin and streptavidin. *Methods in Enzymology*, *184*, 51–67.

Greenwood, C., Ruff, D., Kirvell, S., Johnson, G., Dhillon, H. S., & Bustin, S. A. (2015). Proximity assays for sensitive quantification of proteins. *Biomolecular Detection and Quantification*, *4*, 10–16.

Han, J., Zhu, Y., Chen, X., Liao, B., Yao, H., Song, J., Chen, S., & Meng, F. (2013). The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS. *BioMed Research International*, *2013*, 1–7.

Haynes, E., Jimenez, E., Pardo, M. A., & Helyar, S. J. (2019). The future of NGS (next generation sequencing) analysis in testing food authenticity. *Food Control*, *101*, 134–143.

He, J., Evers, D. L., O'Leary, T. J., & Mason, J. T. (2012). Immunoliposome-PCR: a generic ultrasensitive quantitative antigen detection system. *Journal of Nanobiotechnology*, *10*, 26.

Hendrickson, E. R., Truby, T. M. H., Joerger, R. D., Majarian, W. R., & Ebersole, R. C. (1995). High sensitivity multianalyte immunoassay using covalent DNA-labeled antibodies and polymerase chain reaction. *Nucleic Acids Research*, *23*, 522–529.

Hilbrig, F., & Freitag, R. (2012). Isolation and purification of recombinant proteins, antibodies and plasmid DNA with hydroxyapatite chromatography. *Biotechnology Journal*, *7*, 90–102.

Hollingsworth, P. M. (2011). Refining the DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, *108*, 19451–19452.

Hood, L. E., Hunkapiller, M. W., & Smith, L. M. (1987). Automated DNA sequencing and analysis of the human genome. *Genomics*, *1*, 201–212.

Jain, S. A., Jesus, F. T. de, Marchioro, G. M., & Araújo, E. D. (2013). Extraction of DNA from honey and its amplification by PCR for botanical identification. *Food Sci. Technol.*, *33*, 753–756.

Janssen, K. P. F., Knez, K., Spasic, D., & Lammertyn, J. (2013). Nucleic acids for ultra-sensitive protein detection. *Sensors*, *13*, 1353–1384.

Johnston, M., McBride, M., Dahiya, D., Owusu-Apenten, R., & Singh Nigam, P. (2018). Antibacterial activity of Manuka honey and its components: An overview. *AIMS Microbiology*, *4*, 655–664.

Jones, G. D., & Bryant, Jr, V. M. (2004). The use of ETOH for the dilution of honey. *Grana*, *43*, 174–182.

Kan, Y., Tan, Q., Wu, G., Si, W., & Chen, Y. (2015). Study of DNA adsorption on mica surfaces using a surface force apparatus. *Scientific Reports*, *5*, 8442.

Kato, Y., Fujinaka, R., Ishisaka, A., Nitta, Y., Kitamoto, N., & Takimoto, Y. (2014). Plausible authentication of manuka honey and related products by measuring leptosperin with methyl syringate. *Journal of Agricultural and Food Chemistry*, *62*, 6400–6407.

Lalhmangaihi, R., Ghatak, S., Laha, R., Gurusubramanian, G., & Senthil Kumar, N. (2014). Protocol for optimal quality and quantity pollen DNA isolation from Honey samples. *Journal of Biomolecular Techniques*, *25*, 92–95.

Lempiäinen, A., Stenman, U.-H., Blomqvist, C., & Hotakainen, K. (2008). Free β-subunit of human chorionic gonadotropin in serum is a diagnostically sensitive marker of seminomatous testicular cancer. *Clinical Chemistry*, *54*, 1840–1843.

Litvinov, J., Hagström, A. E. V, Lopez, Y., Adhikari, M., Kourentzi, K., Strych, U., Monzon, F. A., Foster, W., Cagle, P. T., & Willson, R. C. (2014). Ultrasensitive immuno-detection using viral nanoparticles with modular assembly using genetically-directed biotinylation. *Biotechnology Letters*, *36*, 1863–1868.

Liu, X., Xu, Y., Xiong, Y., Tu, Z., Li, Y., He, Z., Qiu, Y., Fu, J., Gee, S. J., & Hammock, B. D. (2014). VHH phage-based competitive real-time immuno-polymerase chain reaction for ultrasensitive detection of ochratoxin A in cereal. *Analytical Chemistry*, *86*, 7471–7477.

Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, *74*, 560–564.

McDonald, C. M., Keeling, S. E., Brewer, M. J., & Hathaway, S. C. (2018). Using chemical and DNA marker analysis to authenticate a high-value food, manuka honey. *NPJ Science of Food*, *2*, 9.

Melzak K.A., Sherwood C.S., Turner R.F.B., & Haynes C.A. (1996). Driving forces for DNA adsorption to silica in perchlorate solutions. *Journal of Colloid and Interface Science*, *181*, 635–644.

Moore, J. C., Spink, J., & Lipp, M. (2012). Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. *Journal of Food Science*, *77*, R118–R126.

Moorhouse-Gann, R. J., Dunn, J. C., de Vere, N., Goder, M., Cole, N., Hipperson, H., & Symondson, W. O. C. (2018). New universal ITS2 primers for high-resolution herbivory analyses using DNA metabarcoding in both tropical and temperate zones. *Scientific Reports*, *8*, 8542.

Morpurgo, M., Facchin, S., Pignatto, M., Silvestri, D., Casarin, E., & Realdon, N. (2012). Characterization of multifunctional nanosystems based on the avidin-nucleic acid interaction as signal enhancers in immuno-detection. *Analytical Chemistry*, *84*, 3433–3439.

Morpurgo, M., Radu, A., Bayer, E. A., & Wilchek, M. (2004). DNA condensation by high-affinity interaction with avidin. *Journal of Molecular Recognition*, *17*, 558–566.

Moura, J. F., Delacerda, L., Sandrini, R., Borba, F. M., Castelo, D. N., Sade, E. R., Sella, S., Minozzo, J. C., Callefe, L. G., & Figueiredo, B. C. (2004). ELISA for determination of human growth hormone: recognition of helix 4 epitopes. *Journal of Biomedicine and Biotechnology*, *2004*, 143–149.

MPI (2017a). Determination of four chemical characterisation compounds in honey by liquid chromatography tandem mass spectrometry (LC-MS/MS). MPI Technical – Paper No: 2017/30.

MPI (2017b). Multiplex qPCR for detection of *Leptospermum scoparium* DNA from pollen in honey. MPI Technical – Paper No: 2017/31.

MPI (2017c). Response to submissions on MPI's proposed definition for mānuka honey. MPI Information Paper No: 2017/14.

MPI (2019). Fines totalling $372,500 imposed in landmark mānuka honey fraud case. Retrieved July 11, 2021, from https://www.mpi.govt.nz/news/media-releases/fines-totalling-372500-imposed-in-landmark-manuka-honey-fraud-case/.

Murphy, J. C., Fox, G. E., & Willson, R. C. (2003). Enhancement of anion-exchange chromatography of DNA using compaction agents. *Journal of Chromatography A*, *984*, 215–221.

Nguyen, V. K., Rihn, B., Heckel, C., Bisseret, F., Girardot, R., & Monteil, H. (1990). Enzyme immunoassay (ELISA) for detection of *Clostridium difficile* toxin B in specimens of faeces. *Journal of Medical Microbiology*, *31*, 251–257.

Ohmenhaeuser, M., Monakhova, Y. B., Kuballa, T., & Lachenmeier, D. W. (2013). Qualitative and quantitative control of honeys using NMR spectroscopy and chemometrics. *ISRN Analytical Chemistry*, *2013*, 1–9.

Petersen, S., & Bryant, V. (2016). The study of pollen and its role in the honey market. *American Bee Journal*, *151*, 591.

Pignatto, M., Realdon, N., & Morpurgo, M. (2010). Optimized avidin nucleic acid nanoassemblies by a tailored PEGylation strategy and their application as molecular amplifiers in detection. *Bioconjugate Chemistry*, *21*, 1254–1263.

POWO (2021). Plants of the world online. Facilitated by the royal botanic gardens, Kew. Retrieved 15072021. Published on the internet; http://www.plantsoftheworldonline.org/.

Prosser, S. W. J., & Hebert, P. D. N. (2017). Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding. *Food Chemistry*, *214*, 183–191.

Richardson, R. T., Lin, C.-H., Sponsler, D. B., Quijia, J. O., Goodell, K., & Johnson, R. M. (2015). Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. *Applications in Plant Sciences*, *3*, 1400066.

Rissin, D. M., Kan, C. W., Campbell, T. G., Howes, S. C., Fournier, D. R., Song, L., Piech, T., Patel, P. P., Chang, L., Rivnak, A. J., Ferrell, E. P., Randall, J. D., Provuncher, G. K., Walt, D. R., & Duffy, D. C. (2010). Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nature Biotechnology*, *28*, 595–599.

Roman, J., Qiu, J., Dornadula, G., Hamuro, L., Bakhtiar, R., & Verch, T. (2011). Application of miniaturized immunoassays to discovery pharmacokinetic bioanalysis. *Journal of Pharmacological and Toxicological Methods*, *63*, 227–235.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*, 5463–5467.

Sano, T., Smith, C., & Cantor, C. (1992). Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science*, *258*, 120–122.

Semprini, A., Singer, J., Braithwaite, I., Shortt, N., Thayabaran, D., McConnell, M., Weatherall, M., & Beasley, R. (2019). Kanuka honey versus aciclovir for the topical treatment of herpes simplex labialis: a randomised controlled trial. *BMJ Open*, *9*, e026201.

Shapla, U. M., Solayman, M., Alam, N., Khalil, M. I., & Gan, S. H. (2018). 5-Hydroxymethylfurfural (HMF) levels in honey and other food products: effects on bees and human health. *Chemistry Central Journal*, *12*, 35.

Shi, B., Shin, Y. K., Hassanali, A. A., & Singer, S. J. (2015). DNA Binding to the Silica Surface. *Journal of Physical Chemistry B*, *119*, 11030–11040.

Simon, S., & Ezan, E. (2017). Ultrasensitive bioanalysis: Current status and future trends. *Bioanalysis*, *9*, 753–764.

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of next-generation sequencing technologies. *Current Protocols in Molecular Biology*, *122*, 1–15.

Smallfield, B. M., Joyce, N. I., & van Klink, J. W. (2018). Developmental and compositional changes in Leptospermum scoparium nectar and their relevance to mānuka honey bioactives and markers. *New Zealand Journal of Botany*, *56*, 183–197.

Soares, S., Amaral, J. S., Oliveira, M. B. P. P., & Mafra, I. (2015). Improving DNA isolation from honey for the botanical origin identification. *Food Control*, *48*, 130–136.

Soares, S., Amaral, J. S., Oliveira, M. B. P. P., & Mafra, I. (2017). A comprehensive review on the main honey authentication issues: production and origin. *Comprehensive Reviews in Food Science and Food Safety*, *16*, 1072–1100.

Sun, L., Tabaka, M., Hou, S., Li, L., Burdzy, K., Aksimentiev, A., Maffeo, C., Zhang, X., & Holyst, R. (2016). The hinge region strengthens the nonspecific interaction between lac-repressor and DNA: A computer simulation study. *PLoS ONE*, *11*, 1–12.

Tan, S. C., & Yiap, B. C. (2009). DNA, RNA, and Protein Extraction: The Past and The Present. *Journal of Biomedicine and Biotechnology*, *2009*, 1–10.

Timpano, E. K., Scheible, M. K. R., & Meiklejohn, K. A. (2020). Optimization of the second internal transcribed spacer (ITS2) for characterizing land plants from soil. *PLoS ONE*, *15*, e0231436.

Turcatti, G., Romieu, A., Fedurco, M., & Tairi, A. P. (2008). A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Research*, *36*, e25.

USDA, NRCS. 2021. The PLANTS Database (http://plants.usda.gov, 07/15/2021). National Plant Data Team, Greensboro, NC USA.

Vu, B. V., Anthony, K. L., Strych, U., & Willson, R. C. (2012). Recovery of Small DNA Fragments from Serum Using Compaction Precipitation. *PLoS ONE*, *7*, e51863.

Wallace, A., Eady, S., Miles, M., Martin, H., McLachlan, A., Rodier, M., Willis, J., Scott, R., & Sutherland, J. (2010). Demonstrating the safety of manuka honey UMF® 20+ in a human clinical trial with healthy individuals. *British Journal of Nutrition*, *103*, 1023–1028.

Wu, R., & Kaiser, A. D. (1968). Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *Journal of Molecular Biology*, *35*, 523–537.

Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P., & Chen, S. (2010). Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS ONE*, *5*, e13102.

Ye, Z., Tu, J., Midde, K., Edwards, M., & Bennett, P. (2018). Singlicate analysis: should this be the default for biomarker measurements using ligand-binding assays? *Bioanalysis*, *10*, 909–912.

Zhang, H., Xu, Y., Huang, Q., Yi, C., Xiao, T., & Li, Q. (2013). Natural phage nanoparticle-mediated real-time immuno-PCR for ultrasensitive detection of protein marker. *Chemical Communications*, *49*, 3778–3780.

Zhou, X., Taylor, M. P., Salouros, H., & Prasad, S. (2018). Authenticity and geographic origin of global honeys determined using carbon isotope ratios and trace elements. *Scientific Reports*, *8*, 14639.

Zhu, X., Li, S., Shan, Y., Zhang, Z., Li, G., Su, D., & Liu, F. (2010). Detection of adulterants such as sweeteners materials in honey using near-infrared spectroscopy and chemometrics. *Journal of Food Engineering*, *101*, 92–97.