HOCHSCHULE
MITTWEIDA
UNIVERSITY OF
APPLIED SCIENCES

MASTER THESIS

Ms B.Sc.

Birgit Pannicke

Genetic pollen analysis based on dual-metabarcoding to portray honey bee foraging in different agro-environments

Mittweida, 2022

Faculty of Applied Computer Sciences and Biosciences

MASTER THESIS

Genetic pollen analysis based on dual-metabarcoding to portray honey bee foraging in different agro-environments

Author:

Ms B.Sc.

Birgit Pannicke

Study Programme:

Genomic Biotechnology

Seminar Group:

GB20wM-M

First Referee:

Prof. Dr. rer. nat. habil. Röbbe Wünschiers

Second Referee:

M.Sc. Lisa Prudnikow

Mittweida, 03.11.2022

Bibliographic Description

Pannicke, Birgit: Genetic pollen analysis based on dual-metabarcoding to portray honey bee foraging in different agro-environments. - 2022 - VIII, 83, VIII p., Mittweida, Hochschule Mittweida, Faculty of Applied Computer Sciences and Biosciences, Master Thesis, 2022

Deutscher Titel:

Genetische Pollenanalyse basierend auf dualem Metabarcoding zur Darstellung des Sammelverhaltens von Honigbienen in unterschiedlichen Agrarlandschaften

Abstract

Pollinating insects are of vital importance for the ecosystem and their drastic decline imposes severe consequences for the environment and humankind. The comprehension of their interaction networks is the first step in order to preserve these highly complex systems. For that purpose, the following study describes a protocol for the investigation of honey bee pollen samples from different agro-environmental areas by DNA extraction, PCR amplification and nanopore sequencing of the barcode regions *rbcL* and ITS. It was shown, that the most abundant species were classified consistently by both DNA barcodes, while species richness was enhanced by single-barcode detection of less abundant species. The analysis of the the different landscape variables exhibited a decline of species richness, Shannon diversity index, and species evenness with increasing organic crop area. However, sampling was only carried out in August and further investigations are suggested to display a more complete picture of honey bee foraging throughout the seasons.

Kurzbeschreibung

Bestäuberinsekten nehmen eine unverzichtbare Rolle im Ökosystem ein und ihr drastischer Rückgang bringt ernstzunehmende Konsequenzen für Mensch und Umwelt mit sich. Die Erfassung von Bestäubernetzwerken ist der erste wichtige Schritt zur Erhaltung dieser hochkomplexen Systeme. Aus diesem Grund beschreibt die folgende Arbeit ein Protokoll zur Untersuchung von Honigbienen-Pollenproben aus landwirtschaftlich unterschiedlich intensiv genutzten Landschaften, welche mithilfe von DNA-Extraktion, PCR-Amplifikation und Nanopore-Sequenzierung der Barcode-Regionen *rbcL* und ITS analysiert wurden. Es konnte gezeigt werden, dass die am häufigsten klassifizierten Arten der beiden DNA-Barcodes übereinstimmten. Weiterhin konnten seltenerere Arten häufig durch je einen der beiden Barcodes detektiert werden, was eine bessere Darstellung der Artenvielfalt ermöglichte. Die Analyse der unterschiedlichen Landschaftsvariablen ergab eine Abnahme von Artenvielfalt, Shannon Index und Gleichverteilung der Spezies mit zunehmendem Anteil an ökologischer Landwirtschaft. Allerdings muss dabei beachtet werden, dass die Probennahme nur im August stattfand und weitere Untersuchungen notwendig wären, um das Sammelverhalten von Honigbienen über einen zeitlichen Gradienten genauer abbilden zu können.

Contents

List of Figures

List of Tables

Nomenclature

accD	Acetyl-CoA Carboxylase Beta Subunit
atpB	ATP Synthase CF1 Beta Subunit
bp	Base Pairs
CBOL	Consortium for the Barcode of Life
COI or COXI	Cytochrome C Oxidase I Gene
DMSO	Dimethyl Sulfoxide
DNA	Desoxyribonucleic Acid
dsDNA	Double-Stranded DNA
EDTA	Ethylenediaminetetraacetic Acid
EU	European Union
Goe	Göttingen
Gos	Goslar
His	Histidine
kb	Kilo Bases
Leu	Leucine
matK	Maturase K
MM	Master Mix
NCBI	National Center for Biotechnology Information
Nor	Northeim
ONT	Oxford Nanopore Technologies
PCR	Polymerase Chain Reaction
psbA	Photosystem II Protein D1 Gene
Q-score	Quality Score
rbcL	RuBisCO Large Subunit Gene
rcf	Relative Centrifugal Force

rDNA	Ribosomal DNA
RNA	Ribonucleic Acid
RNase	Ribonuclease
rpm	Revolutions per Minute
RuBisCO	Ribulose-1.5-Bisphosphate Carboxylase-Oxygenase
ssDNA	Single-Stranded DNA
TAE-Buffer	Tris-Acetate-EDTA-Buffer
Tm	Melting Temperature
tRNA	Transfer RNA
trnH	tRNA-His Gene
trnL	tRNA-Leu Gene
UN	United Nations
UV	Ultraviolet
Wm	Werra-Meißner-Kreis

1 Introduction

It is beyond controversy that the Anthropocene faces one of the greatest depletions in species richness (Aslan et al., 2013; Ceballos et al., 2015). In contrast to previous biotic crises, the current one is not caused by environmental catastrophes but by humankind itself, thereby endangering biological systems it essentially depends on (Hallmann et al., 2017; Nazarevich, 2015). Among the most striking examples are pollination networks (Kluser and Peduzzi, 2007; Ollerton et al., 2011; Potts et al., 2016). No less than 75% of the worldwide leading food crops depend on animal pollination services (Breeze et al., 2016; Klein et al., 2007). The loss of honey bee pollination alone would account for a decrease of 5-8% of the global crop production (Khalifa et al., 2021). The urge for action has never been greater, however, the question evokes how to perform a change that is environmentally sustainable and persistent (Haaland et al., 2011; Urbanowicz et al., 2020).

This question is also the main focus of the *ComBee*-project, which was initiated in 2021 by the research group of Functional Agro-biodiversity at the Georg-August-University of Göttingen (Westphal et al., 2021). Since the *Green Deal* of the EU and the UN *Agenda 2030*, it is also in the interest of politics to promote sustainability strategies such as ecological agriculture. In Germany, the proportion of ecologically used areas should reach 20% by the year 2030 (BMEL, 2019). However, ecological guidelines are not easy to define, since the preservation of ecosystem networks poses complex demands (Steckel et al., 2014). For that purpose, the ambition of the *ComBee*-project is to examine biodiversity, population development, pathogen distribution, and resource utilization of different pollinator species regarding environmental conditions and land use (Westphal et al., 2021). The study thereby combines the interdisciplinary collaboration of different institutions and several modern biomonitoring methods including pollen analysis by genetic technology.

For a few decades, DNA sequencing techniques have indisputably been contributing to the process of biological conservation (Bänsch et al., 2020b; Bell et al., 2016; Lozier and Zayed, 2017). The fact that DNA could be made readable imposed entirely new perspectives on the term biodiversity since genetic information paved the way for rapid species identification, even of those unknown before (Hebert and Gregory, 2005; Valentini et al., 2009). In addition, high-throughput methods enabled the processing of data volumes greater than ever and thereby allowed the realization of large-scale metagenomic surveys (Fišer Pečnikar and Buzan, 2014; Reuter et al., 2015; Thomsen and Willerslev, 2015). One example is the analysis of pollen samples which has gained more and more in importance especially regarding pollinator monitoring effort (Bell et al., 2017a; de Vere et al., 2017; Richardson et al., 2021). Thereby, DNA metabarcoding and portable nanopore sequencing technologies are powerful tools for the investigation of such environmental samples, especially concerning the accumulating ecological challenges we are facing nowadays (Krehenwinkel et al., 2019; Ruppert et al., 2019; Thomsen and Willerslev, 2015).

Figure 1: The honey bee – one of the most important pollinator species

1.1 The Nutritional Value of Pollen

The prevalence of monocultural dominated landscapes has led to exorbitant mass flowering periods, followed by floral scarcity with the consequence of so-called nutritional mismatches for pollinating species (Pamminger et al., 2019b). Especially honey bees tend to favour mass flowering crops, which leads to a one-sided diet on the one hand and the urge for sufficient alternatives on the other, to feed the hive when mass flowering periods are over (Bänsch et al., 2020a). However, providing optimal foraging grounds for pollinators is a complex task. In the past, attention was mainly drawn on the quantity of flowering plants and later on, also the timing of flower-scarce periods was taken into account (Pamminger et al., 2019b). Different studies added a new aspect to this matter, which is the nutritional quality (Alaux et al., 2011; di Pasquale et al., 2013; Frias et al., 2016; Ruedenauer et al., 2021). The importance of this subject becomes even clearer when looking at the composition of the food source of pollinating insects. While nectar mainly serves as an energy source through carbohydrates, pollen provides a much more complex composition of nutrients that are indispensable for a healthy development of honey bees (Ruedenauer et al., 2021). These include lipids, starch, sterols, vitamins, and minerals but most importantly amino acids (di Pasquale et al., 2013; Frias et al., 2016). There are ten essential amino acids that cannot be synthesized by the bees themselves, which means that those amino acids have to be ingested either in free form or by protein digestion (Alaux et al., 2011; Huang, 2012). The availability of these nutrients affects the bees metabolism and immunity, thereby including tolerance to pathogens (e.g. bacteria, viruses, microsporidia) and sensitivity to pesticides (Alaux et al., 2011; di Pasquale et al., 2013; Frias et al., 2016). However, these effects are highly dependent on the pollen availability and quality, since different pollen types vary significantly in their nutritional value (di Pasquale et al., 2013; Ruedenauer et al., 2021).

1.2 DNA Metabarcoding

DNA metabarcoding describes a method for the taxonomic identification of mixed species samples based on specific DNA sequences (Hebert and Gregory, 2005; Valentini et al., 2009). These predominantly short sequences are referred to as DNA markers or barcodes and need to fulfil certain characteristics in order to serve their purpose. They have to be present in all species of interest, exhibit a high interspecific variability and often require conserved flanking regions in order to ensure amplification success throughout a broad range of species (Fišer Pečnikar and Buzan, 2014; Kress et al., 2005). The term 'DNA barcode' was first utilized by Arnot et al. (1993) but became popular with a publication by Hebert et al. (2003), who established the role of the *cytochrome c oxidase I* gene (COI or COXI) as a universal barcode for the taxonomic classification of animal taxa. Since then, DNA barcoding has proven to provide a multitude of new perspectives on taxonomic classification and showed broad applicability within different fields such as medicine, forensic science, phylogenetics, the food industry and biological conservation (Bell et al., 2016; Chase et al., 2005; Valentini et al., 2009).

A major advantage of DNA barcoding over classical taxonomic procedures is the distinction between species even if they exhibit no morphological differences also referred to as cryptospecies (Smith et al., 2006; Valentini et al., 2009). Furthermore, the method can be applied in order to identify species at various developmental stages, which is the case with metamorphic animals or different plant parts, e.g. pollen (Pfenninger et al., 2007; Pornon et al., 2016). Thereby, only a small amount of biological material is required (Petersen et al., 1996; Valière et al., 2003). Once extracted and processed, internationally accessible databases can serve as reference for the collected sequence data and enable the taxonomic identification uncoupled from subjective assessment of morphological attributes (Bell et al., 2017b; Merget et al., 2012; Santamaria et al., 2012). However, it turned out that COI was not suited as a barcode for plants since

it exhibited only low interspecific variation, which is why other options had to be taken into account (Kress et al., 2005). The majority of potential candidates (e.g. *matK, rbcL, trnH-psbA, trnL*) were derived from chloroplast DNA since it is a characteristic organelle for plants (Bell et al., 2016; Kress and Erickson, 2007). On the contrary, only one nuclear barcode (ITS) could be standardized (Bell et al., 2016; Chen et al., 2010). Though most studies agree on the performance of multi-locus approaches for metabarcoding assays, the best combination of barcodes is still subject of research (Milla et al., 2021; Pang et al., 2012; Wang et al., 2015a). Furthermore, current developments regarding long-read nanopore sequencing techniques will also add new aspects to that field (Leidenfrost et al., 2020; Maestri et al., 2019; Prudnikow, 2021).

1.3 Nanopore Sequencing Technology

The principle of nanopore sequencing technologies has actually been investigated since the 1980s (Deamer et al., 2016; Kasianowicz et al., 1996). However, the first applicable technology was released thirty years later by Oxford Nanopore Technologies (ONT) (Reuter et al., 2015). As the name implies, the main actor of ONT are protein nanopores, which are embedded within a synthetic membrane and represent the only connection between two different ionic compartments (*cis* and *trans*) (see Figure 2). The whole reaction chamber is solution-filled and constant voltage ensures the flow of an ion current through the nanopore (van Dijk et al., 2018). If a nucleic acid strand is drawn through the pore, the different nucleotides cause significant changes in the formerly constant ion current (Jiao and Schneeberger, 2017; Wang et al., 2015b).

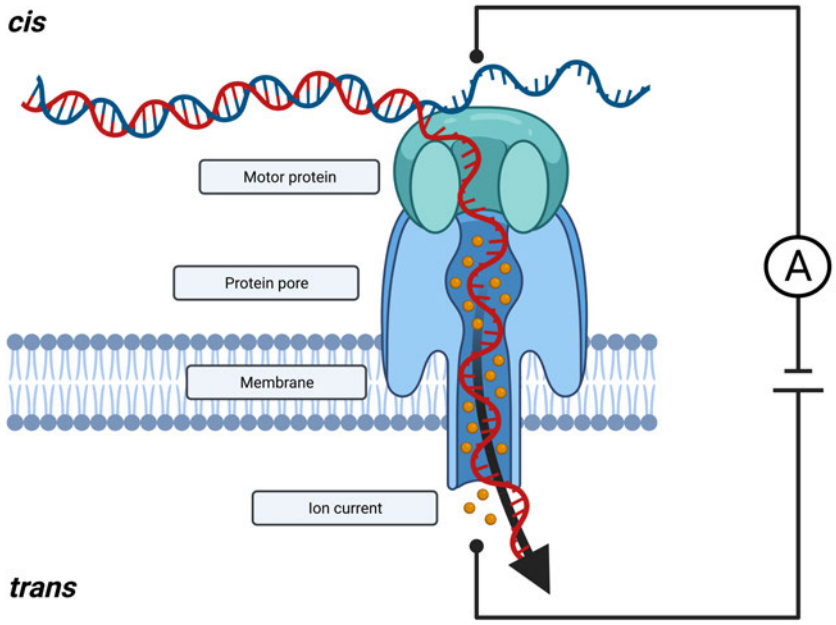


Figure 2: Principle of Oxford Nanopore Sequencing Technologies (created with *BioRender.com*)

The current shifts can be measured in real-time and are represented in a so-called *squiggle plot* (Rang et al., 2018). The translation of these different electric potentials into a nucleotide sequence is then referred to as *basecalling* and can be carried out with corresponding ONT MinKNOW software (van Dijk et al., 2018). Although the method of ONT requires far less extensive and time-consuming sample pre-treatment compared to other techniques, a library preparation is still indispensable (Reuter et al., 2015). To ensure that the nucleic acids reach the nanopore accurately and pass it with a defined translocation speed, the sequences need to be ligated to specific adapters. These consist of protein-DNA complexes thereby including an unwinding enzyme that possesses the ability to separate dsDNA into ssDNA molecules. This step is essential, since the nanopore width of 1.2 nm allows only single-stranded sequences to pass (Deamer et al., 2016; Wang et al., 2015b). Despite its simplicity in principle and application, the technology also has its downsides especially its relatively high error rate of about 10-15% in standard protocols (Li et al., 2016; van Dijk et al., 2018). Nevertheless, there is current technological progress and development of new techniques that

will lead to further improvement of nanopore sequencing in the future (Jain et al., 2015; Li et al., 2016; van Dijk et al., 2018). Although ONT techniques still lag behind other sequencing methods regarding error-rates and throughput, their read length and speed in combination with technology cost and size hold great promise for genetic investigations especially regarding environmental DNA (Krehenwinkel et al., 2019; Leidenfrost et al., 2020; Maestri et al., 2019).

Figure 3: The MinION Mk1B with applied Flongle flow cell

1.4 Choice of DNA Barcodes

In the past years, different DNA sequences have been suggested as barcode regions including genes as well as non-coding spacer sections (Chen et al., 2010; Kress and Erickson, 2007; Pang et al., 2012). Currently, *matK, rbcL, trnH-psbA, trnL* and ITS have been standardized sufficiently for the purpose of DNA metabarcoding (Bell et al., 2016; Kolter and Gemeinholzer, 2021). However, there are still extensive differences between the barcode regions especially regarding sequence length, variation levels, amplification success, and data availability (Bell et al., 2016). The majority of the sci-

entific community therefore practices multi-locus approaches with a combination of at least two different barcodes for DNA metabarcoding attempts (Bell et al., 2017a; Milla et al., 2021; Richardson et al., 2019). For this work, all of the barcode regions mentioned above were examined first by literature and database research regarding the suitability of their characteristics in context of the study's aim (see Table 1).

Table 1 Plant barcode regions

DNA Barcode	Reported Length [bp]	Number of GenBank Entries[6]
ITS	500 – 800[1]	451,505
matK	500 – 1500[2]	236,823
rbcL	1000 – 1500[3]	275,084
trnH-psbA	100 – 1000[4]	159,402
trnL	300 – 600[5]	322,125

[1](Álvarez and Wendel, 2003; Baldwin et al., 1995; Wang et al., 2015a), [2](Bell et al., 2016; Hilu and Liang, 1997; Kress and Erickson, 2007), [3](Bell et al., 2017b; Newmaster et al., 2006), [4](Bell et al., 2016; Pang et al., 2012), [5](Bell et al., 2016; Taberlet et al., 2007), [6]Accessed on 29.11.2021

Since nanopore sequencing will be a main subject, longer sequences were favoured over short ones and *trnL* with a length of only 300-600 bp was not taken into any further account (Bell et al., 2016). A more promising candidate was the *trnH-psbA* intergenic spacer as it has been recommended due to high sequence divergence levels that facilitate species identification (Bolson et al., 2015; Loera-Sánchez et al., 2020; Pang et al., 2012). However, this sequence variety is accompanied by significant length polymorphisms due to insertion, deletion, and repetitive elements (Kress and Erickson, 2007; Pang et al., 2012; Whitlock et al., 2010). These mutations lead to a variation in length of about 100-1000 bp (Bell et al., 2016; Pang et al., 2012), which can be problematic regarding nanopore sequencing, as shorter sequences are less likely read with sufficient quality (Delahaye and Nicolas, 2021). In addition, since the popularity of

trnH-psbA has developed quite recently, it possessed only half the number of GenBank entries compared to other barcode regions, which is why it was not considered for this study either. In 2009, the CBOL Plant Working Group released a publication wherein the combination of *rbcL* and *matK* was emphasized for metabarcoding purposes. However, the species discrimination only reached 75% which could already be improved with other barcode combinations in subsequent studies (Bolson et al., 2015; Loera-Sánchez et al., 2020; Pang et al., 2012). Nevertheless, both *rbcL* and *matK* represent interesting candidates when it comes to nanopore sequencing due to their comparatively long sequence length of up to 1400 bp and more (Hilu and Liang, 1997; Newmaster et al., 2006). Despite their lower discriminatory power, the two loci are still often used for barcoding approaches since they enable a reliable distinction between different plant families and genera even if not always to the species level (Cabelin and Alejandro, 2016). It is therefore adequate to use one of these barcodes as an anchor locus. Since *matK* lacks universal primers for a broad amplification across varying taxa, *rbcL* is the more suitable choice when it comes to metabarcoding (Bolson et al., 2015; Kolter and Gemeinholzer, 2021; Kress and Erickson, 2007). ITS as a spacer region is attributed to exhibit high sequence divergence but is more consistent in length than for example *trnH-psbA* (Álvarez and Wendel, 2003; Baldwin et al., 1995; Chen et al., 2010; Wang et al., 2015a). Furthermore, it comprises the highest amount on GenBank entries which ensures data availability. The choice therefore fell on the plastid gene *rbcL* and the nuclear encoded ITS region to investigate the pollen samples in this study.

1.4.1 RuBisCO Large Subunit (*rbcL*)

RbcL, which stands for *RuBisCO large subunit*, describes a gene region localized in the plastid genome (Patel and Berry, 2008). As the name implies, it is involved in the formation of RuBisCO (Ribulose-1.5-bisphosphate carboxylase-oxygenase), an indis-

pensable enzyme for the carbon fixation in photosynthetic organisms, thereby including plants (Pottier et al., 2018). The suitability of *rbcL* as a barcode region has been discussed in previous studies (Cabelin and Alejandro, 2016; CBOL Plant Working Group, 2009; Newmaster et al., 2006). Compared to other suggested plastid regions, it is one of the best characterized among them and already served as a benchmark locus for phylogenetic analysis (Maloukh et al., 2017; Savolainen and Chase, 2003). Furthermore, it has been demonstrated that the gene is easily amplifiable in the vast majority of plant genera (Cabelin and Alejandro, 2016; Kress and Erickson, 2007). Advances in primer design led to the consistent retrieval of bidirectional high-quality sequences which enhances amplification success and facilitates subsequent processes (Bolson et al., 2015; CBOL Plant Working Group, 2009). However, previous studies almost exclusively conducted short-read sequencing techniques, which is why only a smaller subsection (*rbcLa* represented in Figure 4) was amplified and examined (Cabelin and Alejandro, 2016; Laha et al., 2017; Richardson et al., 2015).

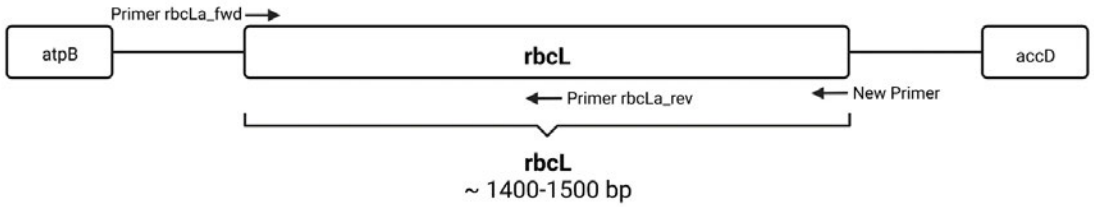


Figure 4: Schematic representation of the *rbcL* gene region in *Arabidopsis thaliana* situated between other plastid genes coding for the *ATP synthase CF1 beta subunit* (*atpB*) and the *acetyl-CoA carboxylase beta subunit* (*accD*). Arrows indicate the primer binding sites.

The attempt to use nanopore sequencing therefore necessitated the design of a universal reverse primer in order to amplify the complete gene region (see Section 5.4.1). Despite using the entire sequence for analysis, *rbcL* is not counted among the most variable barcode sequences with a species discrimination power of about 70% according to Kress and Erickson (2007) but allows the reliable assignment to plant families or genera (Cabelin and Alejandro, 2016; Richardson et al., 2021). Additionally, there

are debates whether plastid sequences can serve as barcodes for pollen analysis at all, since the scientific community is discordant on the matter whether plastid DNA is generally inherited by pollen or not (Kraaijeveld et al., 2015; Matsushima et al., 2011; Sakamoto and Takami, 2018). The reason is the active degradation process of plastid DNA during pollen development in angiosperms caused by a Mg^{2+}-dependent exonuclease (Matsushima et al., 2011; Sakamoto and Takami, 2018). In contrast to that discovery, plastid DNA including barcode sequences could be retrieved from pollen samples in multiple studies which might indicate that not all plant species are affected (Bell et al., 2017a; Galimberti et al., 2014; Richardson et al., 2021; de Vere et al., 2017). Nevertheless, these circumstances of debate and the fact that no plant barcode has yet reached the success of the COI gene lead to the necessity of a dual-barcoding approach in order to represent biodiversity as best as possible (Bell et al., 2016; Kress and Erickson, 2007).

1.4.2 Internal Transcribed Spacer (ITS)

The *Internal transcribed spacer* or short ITS is a non-coding region situated within ribosomal sequences, more precisely between the 18S- and 26S-rDNA coding section, only intermittent by the 5.8-rDNA sequence (shown in Figure 5) (Baldwin, 1992; Martinez-Seidel et al., 2020).

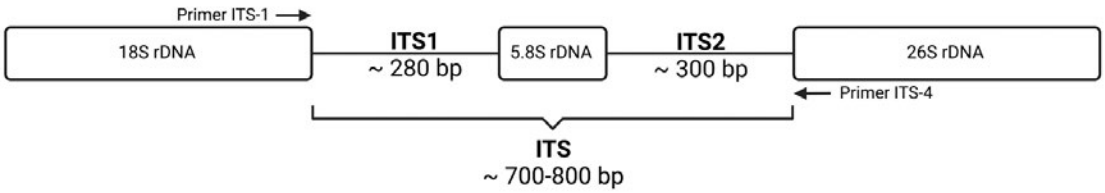


Figure 5: Schematic representation of the ITS region with primer binding sites localized within the more conserved ribosomal sequences 18S-rDNA (ITS-1) and 26S-rDNA (ITS-4) respectively.

Compared to other suggested barcode regions, its most salient characteristic is that

ITS is encoded within the nucleus and not the plastid genome as all other established barcodes for plants (Bell et al., 2016). As an intron region it exhibits high sequence divergence, leading to over 90% successful species identification, which is why Chen et al. (2010) suggested to use an even smaller section of the original barcode, namely the ITS2 region. Compared to ITS, ITS2 only counts about 300 bp on average (Bell et al., 2016; Wang et al., 2015a) instead of approximately 700 bp (Álvarez and Wendel, 2003; Baldwin et al., 1995; Wang et al., 2015a). This was also the reason why ITS2 was favoured over the complete ITS region since next generation sequencing techniques were only able to perform short-read sequencing (Chen et al., 2010; Wang et al., 2015a). However, sequencing methods have evolved rapidly during the last years and Oxford Nanopore Technologies pointed out new perspectives regarding long-read sequencing (Jiao and Schneeberger, 2017; Reuter et al., 2015). Furthermore, it had been questioned whether ITS2 should be favoured at all over ITS1 since the first section of the internal transcribed spacer actually seems to exhibit more variability than the second one (Wang et al., 2015a). These considerations led to the decision to use the entire ITS sequence in this work.

2 Aim of Study

This work aims to portray the plant-pollinator interactions in different agro-environmental landscapes by analyzing honey bee pollen baskets. Applied methods include DNA extraction, PCR amplification, nanopore sequencing, and DNA metabarcoding as presented in Figure 6. The process thereby includes the examination of the DNeasy® Plant Mini Kit and the NucleoSpin® Food Kit to develop a protocol for sufficient DNA extraction. Furthermore, two DNA barcodes will be evaluated regarding their value for plant species classification: the *Internal transcribed spacer* (ITS) and the *RuBisCO large subunit* (*rbcL*). Furthermore, a new primer will be designed for the full-length reverse amplification of the *rbcL* sequence. After identifying the plant species from the pollen samples, the outcome will be tested against three different landscape variables: the proportion of organic crop area, the percentage of semi-natural habitats, and the annual flowerfield coverage. This study will investigate the question whether these agro-environmental variables have a positive effect on species richness, abundance, and diversity of the pollen collected by honey bees.

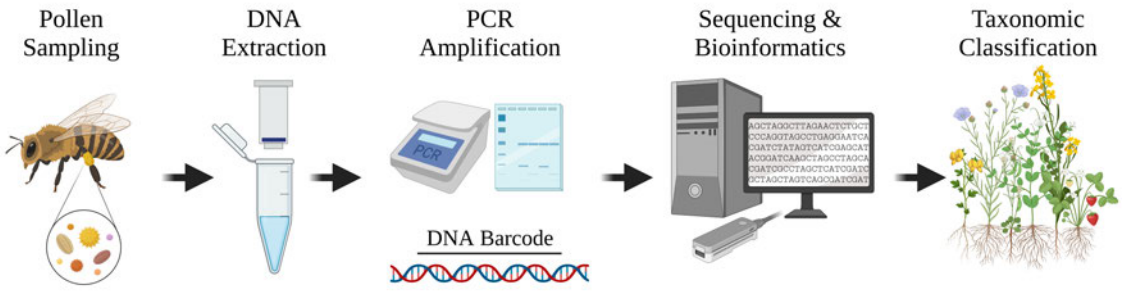


Figure 6: Illustration of the work process (created with *BioRender.com*)

3 Material

3.1 Test Animals

Apis mellifera colonies based on *Apis mellifera carnica* queens, freely mated with *A. m. carnica* and *Buckfast* drones

3.2 Chemical Reagents and Solutions

Table 2 Chemical reagents and solutions

Reagent	Manufacturing Company	Identification Number
Certified™ Molecular Biology Agarose	Bio-Rad Laboratories Inc. (Hercules, USA)	CONTROL: BRD00117
Ethanol 99.5%	Carl Roth GmbH + Co. KG (Karlsruhe, Germany)	LOT: 345233474
Ethanol 96% (p.a., undenaturated)	Carl Roth GmbH + Co. KG (Karlsruhe, Germany)	LOT: 211307928
Nuclease-free water	Carl Roth GmbH + Co. KG (Karlsruhe, Germany)	LOT: 337262174
Orange G	Carl Roth GmbH + Co. KG (Karlsruhe, Germany)	—
Rotiphorese® 50x TAE-Buffer	Carl Roth GmbH + Co. KG (Karlsruhe, Germany)	LOT: 059280422
SYBR® Safe DNA Gel Stain (10,000x in DMSO)	Thermo Fisher Scientific Inc. (Waltham, USA)	LOT: 1933715

3.3 Consumables and Kits

Table 3 Consumables and kits

Article	Manufacturing Company	Identification Number
AMPure XP beads	Beckman Coulter Inc. (Indianapolis, USA)	LOT: 18440500
DNeasy® Plant Mini Kit	Qiagen (Venlo, Netherlands)	LOT:169012674
Flongle Flow Cells FLO-FLG001	Oxford Nanopore Technologies (Oxford, UK)	LOT: 33000494
Flow Cell Priming Kit EXP-FLP002	Oxford Nanopore Technologies (Oxford, UK)	LOT: FLP002.10.0043
GeneJET PCR Purification Kit	Thermo Fisher Scientific Inc. (Waltham, USA)	LOT: 00998764, 00998970
Ligation Sequencing Kit SQK-LSK109	Oxford Nanopore Technologies (Oxford, UK)	LOT: CS9109.10.0001
Native Barcoding Expansion 1-12 EXP-NBD104	Oxford Nanopore Technologies (Oxford, UK)	LOT: EN04.10.0013
NucleoSpin® Food Kit	Macherey-Nagel GmbH + Co. KG (Düren, Germany)	LOT: 2107/002
Phusion High-Fidelity PCR Kit	Thermo Fisher Scientific Inc. (Waltham, USA)	LOT: 00960122
Precellys Lysing Kit Soft tissue homogenizing CK14 1.4 mm	Bertin Technologies (Montigny-le-Bretonneux, France)	LOT: 210622-859
Qubit™ dsDNA HS Assay Kit	Thermo Fisher Scientific Inc. (Waltham, USA)	LOT: 2069590
Zirkonoide beads, yttrium stabilized Type ZY-P 2.60-3.30 mm	Sigmund Lindner GmbH (Warmensteinach, Germany)	LOT: 710619

3.4 Biomolecular Reagents

Table 4 Biomolecular reagents

Reagent	Manufacturing Company	Identification Number
Blunt/TA Ligase Master Mix	New England BioLabs Inc. (Ipswich, USA)	LOT: 10115871
NEBNext® Quick Ligation Reaction Buffer 5X	New England BioLabs Inc. (Ipswich, USA)	LOT: 0221803
Quick-Load® purple 1 kb DNA Ladder 50 µg/ml	New England BioLabs Inc. (Ipswich, USA)	LOT: 0071708
Quick-Load® purple 1 kb Plus DNA Ladder 100 µg/ml	New England BioLabs Inc. (Ipswich, USA)	LOT: 10015415
Quick T4 DNA Ligase	New England BioLabs Inc. (Ipswich, USA)	LOT: 10068797
Ultra™ II End End Prep Enzyme Mix	New England BioLabs Inc. (Ipswich, USA)	LOT: 10120040, 10133469
Ultra™ II End End Prep Reaction Buffer	New England BioLabs Inc. (Ipswich, USA)	LOT: 10120041

Table 5 Primer sequences (provided by biomers.net GmbH; Ulm, Germany)

Barcode Target	Primer Name	Primer Sequence	References
ITS	ITS-1	5‘-TCCGTAGGTGAACCTGCGG-3‘	White et al. (1990), LAG (2011)
	ITS-4	5‘-TCCTCCGCTTATTGATATGC-3’	White et al. (1990), LAG (2011)
rbcL	rbcLa_f	5‘-ATGTCACCACAAACAGAGACTAAAGC-3’	Kress and Erickson (2007); Bell et al. (2019)
	rbcL2_r	5‘-TTGATCTCCTTCCATACTTCACAAGC-3’	designed with Primer-BLAST

3.5 Devices

Table 6 Devices

Device	Manufacturing Company
Centrifuge 5430	Eppendorf SE (Hamburg, Germany)
Dark Hood DH-30/32	Biostep GmbH (Burkhardtsdorf, Germany)
Digital Dry Block Heater Grant QBD1	Grant Instruments Ltd. (Cambridge, UK)
Flongle Adapter ADP-FLG001	Oxford Nanopore Technologies (Oxford, UK)
Gelelectrophoresis Mini-Sub® Cell GT	Bio-Rad Laboratories Inc. (Hercules, USA)
Gelelectrophoresis Power-Supply Unit PS 304 MiniPac	Apelex (Lisses, France)
HulaMixer™ Sample Mixer	Thermo Fisher Scientific Inc. (Waltham, USA)
Laminar Flow Chamber Class II HERAsafe™	Heraeus (Hanau, Germany)
Mastercycler® Nexus Thermal Cycler GSX1	Eppendorf SE (Hamburg, Germany)
Microcentrifuge, MiniStar silverline	VWR International (Radnor, USA)
Microscope Axiostar Plus	Carl Zeiss AG (Jena, Germany)
Minicentrifuge MiniSpin® plus	Eppendorf SE (Hamburg, Germany)
MinION Mk1B	Oxford Nanopore Technologies (Oxford, UK)
NanoVue™ Plus Spectrometer	GE Healthcare GmbH (Solingen, Germany)
Qubit™ 3 Fluorometer	Thermo Fisher Scientific Inc. (Waltham, USA)
UV-Transilluminator Vio View UV light UST-20M-8E, 312 nm	Biostep GmbH (Burkhardtsdorf, Germany)
Vortex-Genie 2	Scientific Industries Inc. (New York, USA)

3.6 Software

Table 7 Software

Software	Version	Provider
argusX1	7.14.22	Biostep GmbH (Burkhardtsdorf, Germany)
BLAST	2.13.0	Altschul et al. (1990)
Clustal Omega	1.2.4	Sievers et al. (2011)
Guppy	6.0.1	Oxford Nanopore Technologies (Oxford, UK)
Inkscape	1.1	Inkscape Community
MinKNOW	22.03.6	Oxford Nanopore Technologies (Oxford, UK)
NanoPlot	1.40.2	de Coster et al. (2018)
NEBioCalculator®	1.15.0	New England BioLabs Inc. (Ipswich, USA)
Porechop	0.2.4	Wick (2017)
Primer-BLAST	4.1.0	Ye et al. (2012)
PuTTY	0.76	Tatham (2021)
qfilter	1.0	Wünschiers (2022)
R	4.1.1	Ihaka and Gentleman (1996)
RStudio	1.4.1717	RStudio PBC
R - DHARMa	0.4.6	Hartig (2022)
R - dyplr	1.0.10	Wickham et al. (2022)
R - forcats	0.5.2	Wickham (2022a)
R - ggplot2	3.3.5	Wickham (2016)
R - ggrepel	0.9.1	Slowikowski (2021)
R - glmmTMB	1.1.4	Brooks et al. (2017)
R - scales	1.2.0	Wickham and Seidel (2022)
R - stringr	1.4.1	Wickham (2022b)
R - tidyr	1.2.0	Wickham and Girlich (2022)
R - vegan	2.6.2	Oksanen et al. (2022)
R - xtable	1.8.4	Dahl et al. (2019)

4 Methods

4.1 Pollen Collection

The pollen samples examined in this study were collected from honey bee colonies exclusively. The sampling took place from the 2nd to the 6th of August 2021, during a period from about 9 am to 3 pm as the hives activity was highest during this time of day. A total of 19 different locations around the city of Göttingen in Lower Saxony, Germany served as sampling sites, depicted in the map of Figure 8 (the corresponding coordinates can be found in supplemental Table S1). Each location was equipped with one wooden hive containing four bee colonies. Two samples, each with ten bees, were taken on both sides of the hive (see Figure 7). Since the four different bee colonies were located so close to one another, the assignment of a sample to an individual colony cannot be guaranteed. The bees were caught with an insect net and, regarding the aim of this study, only bees with pollen baskets were kept to ensure enough material for pollen analysis. The samples were conserved in 50 ml Falcon tubes containing 20-30 ml of 70% ethanol (v/v, p.a., undenaturated) and stored at room temperature until they were used for further analysis.

Figure 7: Sampling method at a bee hive (two samples were taken at each side)

Figure 8: Sampling locations: Goe = Göttingen, Gos = Goslar, Nor = Northeim, Wm = Werra-Meißner-Kreis (created with *Google Maps*)

4.2 Sample Preparation

In order to separate bees and pollen, the bees were removed with forceps and checked for residual pollen loads which did not dissolve in the ethanol solution. If present, those pollen remains were scraped off with a fine spatula and added to the sample while the bee body was washed again with 1 ml of 70% ethanol (v/v, p.a.) before being removed from the sample. Afterwards, the Falcon tubes were centrifuged at 7000 rcf for 20 minutes at 20 °C. The ethanol supernatant was removed without disturbing the pollen pellet at the bottom of the tube until less than 1 ml of ethanol was left. Then, the pellet was resuspended in the remaining ethanol and transferred into a 2 ml Eppendorf tube. Subsequently, the bottom of the Falcon tube was washed another time with 400 µl of ethanol in order to remove any residual pollen grains which were then added to same tube. To ensure enough pollen material for DNA extraction, the samples of the same bee hive side (see Figure 7) were pooled in one tube. Next, the tubes were centrifuged at 13,000 rcf for 10 minutes. Afterwards, the supernatant was removed and the pollen was left to dry under a sterile laminar flow chamber for at least 24 hours until the pellet was fully arid.

4.3 DNA Extraction

In order to extract sufficient DNA from the pollen mixtures, two different kits were examined: the DNeasy® Plant Mini Kit from Qiagen which has commonly been used for pollen DNA extraction (Galimberti et al., 2014; Hawkins et al., 2015; Bänsch et al., 2020b) and the NucleoSpin® Food Kit from Macherey-Nagel as recommended by Bell et al. (2017a). In order to establish a protocol for sufficient DNA extraction, the method was tested first with residual pollen samples derived from honey bees and wild bees by Prudnikow (2021). An amount of 0.015 g pollen dry mass was used for each extraction based on the findings of previous studies (Bänsch, 2019; Prudnikow, 2021).

Both kit protocols required the disruption of the pollen exine in the first place before DNA could be extracted. For this reason, the samples were treated with two different sizes of ceramic beads: 60 beads of 1.4 mm (Precellys Lysing Kit) and 10 beads of 2.60 - 3.30 mm (zirkonoide, yttrium stabilized). Afterwards, 350 µl of wash buffer from the respective DNA extraction kit (see Sections 4.3.1 and 4.3.2) were added to the tube. The samples were vortexed with the beads for 1 minute, followed by 1 minute on ice. This step was repeated two more times. Subsequently, the mixtures were separated from the beads and transferred into a new 2 ml tube. Afterwards, the ceramic beads were washed a second time with defined volumes of wash buffer (see section 4.3.1 and 4.3.2) and the remaining volume was added to the new tube as well. Before proceeding with the protocol, 10 µl of each sample were used for microscopic analysis (magnification 10x40) in order to evaluate the success of the pollen disruption. For both kits, manufacturer's instructions were followed with slight adjustments described in detail in the subsequent chapters. Afterwards, DNA concentrations were measured with the Qubit™ 3 Fluorometer and the NanoVue™ Plus Spectrometer. Additionally, NanoVue™ analysis allowed the determination of the DNA purity.

4.3.1 DNeasy® Plant Mini Kit modifications

For the DNA extraction with the DNeasy® Plant Mini Kit the protocol was adapted from Prudnikow (2021). The pollen sample and the ceramic beads were diluted in 350 µl Buffer AP1 and left for dissolution for 10 minutes at 30 °C. After vortexing and transferring the mixture, the beads were washed a second time with 150 µl AP1. Manufacturer's instructions were followed (Qiagen, 2016) with few modifications. The centrifugation time at step 4 was changed to 10 minutes in order to gain a consistent pollen pellet. Furthermore centrifugation parameters for step 7, 8, and 11 were adjusted to 10,000 rpm. Finally, the elution volume of Buffer AE at step 11 was reduced to 50 µl per turn (100 µl in total) to increase DNA concentration.

4.3.2 NucleoSpin® Food Kit modifications

For the DNA extraction with the NucleoSpin® Food Kit modifications, the pollen sample and the ceramic beads were mixed in 350 µl Buffer CF and incubated for 5 minutes at 65 °C. After vortexing and transferring the mixture, the beads were washed again with 200 µl Buffer CF. The proximate steps followed manufacturer's instructions (Macherey-Nagel, 2020) with following modifications: the centrifugation parameters at step 2 were adjusted to 13,000 rcf to ensure a consistent pollen pellet. Furthermore, an additional step was performed, where 4 µl RNase A of the DNeasy® Plant Mini Kit were incubated for 10 minutes at 65 °C with each sample prior to DNA extraction in order to degrade RNA and gain pure DNA samples.

4.4 Primer Design

Since only a subsection of *rbcL* had been used for barcoding attempts so far, a new reverse primer needed to be selected in order to amplify the entire gene region. To ensure a universal applicability across different plant species a multiple sequence alignment was conducted first in order to detect conserved gene segments. For that purpose, the *rbcL* sequences of ten different species representing different plant orders and clades (see Table 8) were aligned with the help of *Clustal Omega* (`https://www.ebi.ac.uk/Tools/msa/clustalo/`).

Table 8 Species for the *rbcL* alignment

Clade	Order	Species	Gene ID
Monocotyledons	Asparagales	*Cypripendium calceolus*	42906461
	Liliales	*Lilium martagon*	37626811
	Poales	*Zea mays*	845212
Eudicotyledons	Asterales	*Helianthus annuus*	4055709
	Brassicales	*Brassica napus*	11542073
	Brassicales	*Arabidopsis thaliana*	844754
	Caryophyllales	*Dianthus caryophyllus*	38290557
	Ranunculales	*Ranunculus repens*	35989791
	Rosales	*Fragaria vesca*	10251527
	Sapindales	*Acer campestre*	65325797

```
Zea             GGGCGCGATCTTGCTCGTGAAGGTAATGAAATTATCAAAGCAGCTTGCAAATGGAGTGCT   1359
Cypripedium     GGGCGTGATCTTGCTCGTGAAGGTAATGATATTATTCGTGAAGCTAGCAAATGGAGCCCT   1380
Lilium          GGGCGTGATCTTGCTCGTGAGGGTAATGAAATTATCCGTGAAGCTTGCAAATGGAGTCCT   1359
Brassica        GGACGTGATCTTGCAGTCGAGGGTAATGAAATTATCCGTGAGGCTTGCAAATGGAGTCCT   1359
Arabidopsis     GGACGTGATCTTGCAGTCGAGGGTAATGAAATTATCCGTGAAGCTTGCAAATGGAGTCCT   1359
Ranunculus      GGACGTGATCTTGCTCGTGAAGGTAATGAAATTATCCGTGAGGCTTGCAAATGGAGCCTT   1359
Helianthus      GGACGCGATCTTGCTACTGAGGGTAATGAAATTATCCGTGAGGCTACCAAATGGAGTCCT   1359
Dianthus        GGACGTGATCTTGCTCGCGAGGGTAATACTATTATTCGCGAGGCTTGCAAATGGAGTCCT   1359
Fragaria        GGACGTGATCTCGCTCGTGAGGGTAATGACATTATTCGTGAGGCTTGTAAATGGAGTCCT   1359
Acer            GGACGCGATCTTGCTCGCGAGGGTAATGAAATTATCCGTGAGGCTAGCAAATGGAGTGCT   1359
                ** ** ***** **    ** ******   *****    *  ***   ********   *

Zea             GAACTAGCCGCAGCTTGTGAAATATGGAAGGAGATCAAATTTGATGGTTTCAAAGCGATG   1419
Cypripedium     GAACTAGCCGCTGCTTGTGAAGTATGGAAAGAGATCAAATTTGATTTCGACCCAGTGGAT   1440
Lilium          GAACTAGCTGCTGCTTGTGAAGTATGGAAGGAGATCAAATTCGAGTTCGAACCAGTAGAT   1419
Brassica        GAACTAGCTGCTGCTTGTGAAGTATGGAAGGAGATCACATTTAACTTCCCAACCATCGAT   1419
Arabidopsis     GAACTAGCTGCTGCTTGTGAAGTATGGAAAGAGATCACATTTAACTTCCCAACCATCGAT   1419
Ranunculus      GAACTAGCCGCTGCTTGTGAGGTATGGAAGGAGATCAAATTCGAATTTGAAGCAATGGAT   1419
Helianthus      GAACTAGCTGCTGCTTGTGAAGTATGGAAGGAGATCAAATTTGAGTTCCAGGCAATGGAT   1419
Dianthus        GAACTAGCTGCTGCTTGTGAAGTATGGAAGGAAATCAAATTTGAATTCGAAGCAATGGAT   1419
Fragaria        GAACTAGCTGCTGCTTGTGAAGTATGGAAAGAGATCAAATTTGAATTCGAAGCAATGGAT   1419
Acer            GAATTGGCTGCTGCTTGTGAAGTATGGAAGGAGATCAAATTTGAATTTGAAGCAATGGAT   1419
                *** * ** ** ******** ******* ** **** ***  *

Zea             GATACCATATAA--------------------------        1431
Cypripedium     AAGCTAGATAAATAG-----------------------        1455
Lilium          AAGCTAGATACAGAGAAGAAATAA--------------        1443
Brassica        AAATTAGATGGCCAAGACTAG-----------------        1440
Arabidopsis     AAATTAGATGGCCAAGAGTAG-----------------        1440
Ranunculus      ACTTTGTAA-----------------------------        1428
Helianthus      ACTTTGGATACGGATAAGGATAAAGATAAGAAGAGATAA       1458
Dianthus        ACAATCTAA-----------------------------        1428
Fragaria        ACTTTGTAA-----------------------------        1428
Acer            ACTTTGTAA-----------------------------        1428
```

Figure 9: Alignment of different plant *rbcL* sequences by Clustal Omega to identify conserved bases (*), Representation of the 3' end; Indicated in red is the region used for primer design

The region located as close as possible to the 3' end of the leading strand with the longest sequence of conserved bases was chosen as a starting point for the further primer design (depicted in Figure 9).

The evaluation of the primers was carried out with the *NCBI Primer-BLAST* tool (`https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi`). The *rbcL* sequence of the sunflower *Helianthus annuus* thereby served as template since it represented of the consensus site of all ten species best. The *rbcLa_f* primer was used as forward primer for the query. A maximum difference of ± 1 K was tolerated regarding the melting

temperatures. A gradient PCR was conducted in order to identify the optimal annealing temperature for the newly designed *rbcL* primer. PCR settings were the same as described in section 4.5 except that the same sample was amplified in twelve different columns over temperature intervals of 53-64 °C (1 K steps) for approximation and of 58-63.5 °C (0.5 K steps) for refinement. The maximum concentration yield was identified with the Qubit™ 3 Fluorometer and resulted in an optimal annealing temperature of 62 °C.

4.5 Polymerase Chain Reaction (PCR)

Table 9 Reagents for PCR mastermix (MM)

Reagent	Concentration	Volume [µl]
Phusion HF buffer (5x)	1x	4
Primer fwd (10 µM)	0.4 µM	0.8
Primer rev (10 µM)	0.4 µM	0.8
dNTPs (10 mM)	0.2 mM	0.4
Nucleasefree water	—	12.8
Phusion DNA polymerase	0.02 U/µl	0.2
Total MM volume	—	19
DNA template (20 ng)	1 ng/µl	1

Table 10 PCR program settings for ITS amplification

Step	Temperature	Time	Cycles
Initial denaturation	98 °C	30 s	1
Denaturation	98 °C	20 s	
Annealing	52 °C	30 s	35
Extension	72 °C	90 s	
Final extension	72 °C	10 min	1
Hold	4 °C		

Table 11 PCR program settings for *rbcL* amplification

Step	Temperature	Time	Cycles
Initial denaturation	98 °C	30 s	1
Denaturation	98 °C	20 s	
Annealing	62 °C	30 s	35
Extension	72 °C	90 s	
Final extension	72 °C	10 min	1
Hold	4 °C		

The PCR protocol described in this chapter was adapted from Prudnikow (2021). The barcodes were amplified using the primers depicted in Section 3.4 Table 5. Reagents used for the mastermix (MM) preparation of each PCR approach are displayed in Table 9. DNA extraction samples were adjusted accordingly to a template concentration of 20 ng/µl. The respective PCR program settings are represented in Table 10 for ITS and in Table 11 for *rbcL*. Each PCR approach additionally included a *no template control* (NTC) to eliminate contamination of the master mix. The preparation process was carried out on ice. After amplification with a thermal cycler, the samples were purified using the GeneJET PCR Purification Kit with a final elution volume of 40 µl. The concentration of each sample was measured with the Qubit™ 3 Fluorometer and DNA

purity was determined by the NanoVue™ Plus Spectrometer. Furthermore, DNA length and master mix purity were verified with a gel electrophoresis (see Section 4.6).

4.6 Gel Electrophoresis

For gel electrophoresis, a 1% (w/v) gel was prepared by dissolving 300 mg agarose in 30 ml TAE-buffer together with 3 µl of SYBR® Safe DNA gel stain. Each DNA sample was mixed with 1 µl of Orange G loading dye before being applied onto the gel. DNA extraction samples were applied with 10 µl, while only 5 µl were used for PCR amplified samples. The Quick-Load® purple 1 kb DNA ladder served as reference. The run-time amounted 35 minutes at 100 V.

4.7 Nanopore Sequencing

Sequencing preparation followed the Nanopore protocol for Native barcoding amplicons (Oxford Nanopore Technologies, 2021), thereby combining the Ligation Sequencing Kit (SQK-LSK109), the Native Barcoding Expansion 1-12 (EXP-NBD104), and the Flow Cell Priming Kit (EXP-FLP002). Different reagents from *New England BioLabs Inc.* complemented the protocol (see Section 3.4 Table 4). Molarity calculations were carried out with the *NEBioCalculator®* (`https://nebiocalculator.neb.com/#!/dsdnaamt`), based on a mean value of 750 bp for ITS and 1500 bp for *rbcL*. Since the sequencing was carried out with Flongle flow cells (Type FLO-FLG001) instead of MinION flow cells the protocol had to be adjusted accordingly as described in the following paragraphs.

For the first preparation step, the modified *end-prep* reaction, the samples were primarily adjusted to a volume of 24 µl containing 100 fmol of amplicon DNA. Subsequently, 1.75 µl of the Ultra II End-prep reaction buffer and 1.5 µl of the Ultra II End-prep enzyme

mix were added. After incubation, the samples were mixed with 30 µl of AMPure XP beads. Finally, the DNA was dissolved and retrieved in 12.5 µl nuclease-free water.

In the second step, each sample was tagged by adding 1.25 µl of an individual native barcode combined with 12.5 µl of the Blunt/TA Ligase Master Mix. Afterwards, 25 µl of AMPure XP beads were added to the reaction. The barcoded samples were retained in 13 µl nuclease-free water. After concentration measurement, the samples were pooled equimolar to a total of 100 fmol and adjusted to a volume 33.5 µl with nuclease-free water.

The third working step included the *adapter ligation and clean-up*. Therefore, each sample was mixed with 2.5 µl Adapter Mix II (AMII), 10 µl NEBNext Quick Ligation Reaction Buffer (5X), and 5 µl Quick T4 DNA Ligase. The reaction was then purified with 25 µl of AMPure XP beads and 125 µl of Short Fragment Buffer (SFB). In the end, the DNA was eluted in 7 µl Elution Buffer (EB).

For the final step, 20 fmol of the DNA library were adjusted to 5 µl with nuclease-free water. A Flongle adapter (ADP-FLG001) was applied in addition to the MinION Mk1B. Before loading the flow cell, a hardware and flow cell check were performed to ensure adequate preconditions for each sequencing run. Flongle flow cells with less than 50 active nanopores were not used for sequencing. Each sequencing run was set for 24 hours and only stopped if no active pores were available anymore. The reads were processed by the MinKNOW software and saved in FAST5 format.

4.8 Data Processing

Full commands are listed in supplemental Section *Commands*. Processing steps described from Section 4.8.1 to Section 4.8.3 were carried out on a Linux server. Details about software tool versions and providers can be seen in Section 3.6 Table 7. An overview of the different read processing steps is displayed in Figure 10.

4.8.1 Generating the BLAST Databases

To generate the two BLAST databases for ITS and *rbcL*, all DNA sequences belonging to the respective barcodes were downloaded from *NCBI Nucleotide* (`https://www.ncbi.nlm.nih.gov/nuccore`) as FASTA-files using the following search parameters:

```
ITS OR internal transcribed spacer [All Fields] AND plants [filter]
```

```
rbcL OR rbc-L OR rubisco large subunit [All Fields] AND plants [filter]
```

To refine the databases, the sequences should be filtered with a species list for regional plants. This list was created according to the Red Lists of fern and vascular plants of Lower Saxony, Hesse, and Thuringia which are the surrounding federal states of Göttingen (Garve, 2004; Korsch et al., 2010; HLNUG, 2019). The final list was compared with the NCBI sequence titles by `grep --no-group-separator -F -w -A1 -f PLANTLIST.txt INPUT.fasta` and matches were saved within a new FASTA-file. Using this file, a database was created with `makeblastdb` by specifying the type as nucleotide `-dbtype nucl`. These steps were carried out for ITS and *rbcL* sequences separately.

4.8.2 Basecalling and Trimming

High-accuracy basecalling `-c dna_r9.4.1_450bps_hac.cfg` was performed with the *Guppy* basecaller software (Oxford Nanopore Technologies). Thereby, read splitting

was activated `--do_read_splitting` as well as trimming of adapters `--trim_adapters` and Nanopore barcode tagging sequences `--trim_barcodes`. Quality score filtering was deactivated by `--disable_qscore_filtering` to be performed afterwards, since the Nanopore quality score might differ from the classical Phred quality score (Delahaye and Nicolas, 2021). In the end, the basecalled reads were saved in compressed FASTQ-files with `--compress_fastq`. The tool *Porechop* (Wick, 2017) was used to trim the primer sequences from the reads in order to avoid interference with the alignment later on, especially since the ITS primers are situated outside of the barcode sequence (see Figure 5). The primer nucleotide sequences were therefore added to the `adapters.py` document in 5'→3' and 3'→5' orientation before running the program.

4.8.3 BLAST and Filtering

Before applying BLAST, the FASTQ-files were filtered for specific Q-score parameters. For this purpose, the tool *qfilter* (Wünschiers, 2022) was applied with the following settings: 70% of the read (`-p 70`) should have a Q-score of at least 15 (`-s 15`) while the each read was supposed to have a minimum Q-score of 10 (`-m 10`).

The alignment of the reads with the respective database was carried out with Nucleotide BLAST `blastn` and saved in output format 6 `-outfmt6`. Since BLAST generates multiple hits for one read, the output file was always filtered subsequently for a minimum alignment length (min. 400 bp for ITS and min. 1000 bp for *rbcL*) and for an alignment identity of at least 95%. With these parameters the best hit of each read was extracted for the final analysis. The series of read processing steps is summarized in Figure 10.

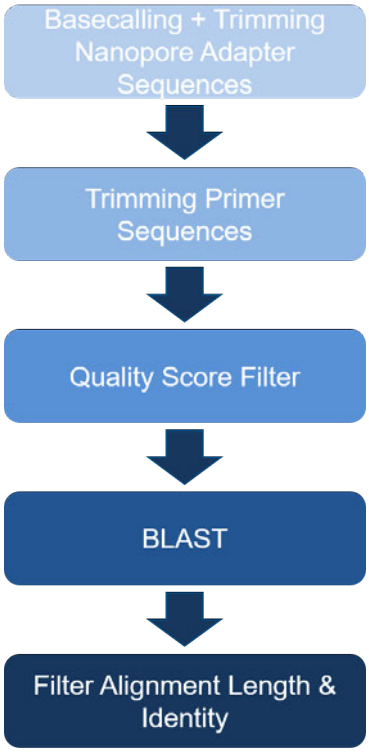


Figure 10: Schematic representation of the sequencing read processing steps

4.8.4 Data Visualisation and Statistical Analysis

Data visualisation and statistical analysis were carried out with RStudio (R version 4.1.1) if not stated otherwise. Three different landscape gradients in a 1 km radius around the bee hives were examined, represented in Figure 11: the proportion of organic crop area, the percentage of semi-natural habitats, and the annual flowerfield coverage (the corresponding data was provided by the research group of Functional Agrobiodiversity; Georg-August-University Göttingen). Generalized linear mixed modelling was conducted with the `glmmTMB` package (Brooks et al., 2017). Diversity analysis was carried out with the `vegan` package (Oksanen et al., 2022). Other R packages utilized during this study are listed in detail in Section 3.6 Table 7.

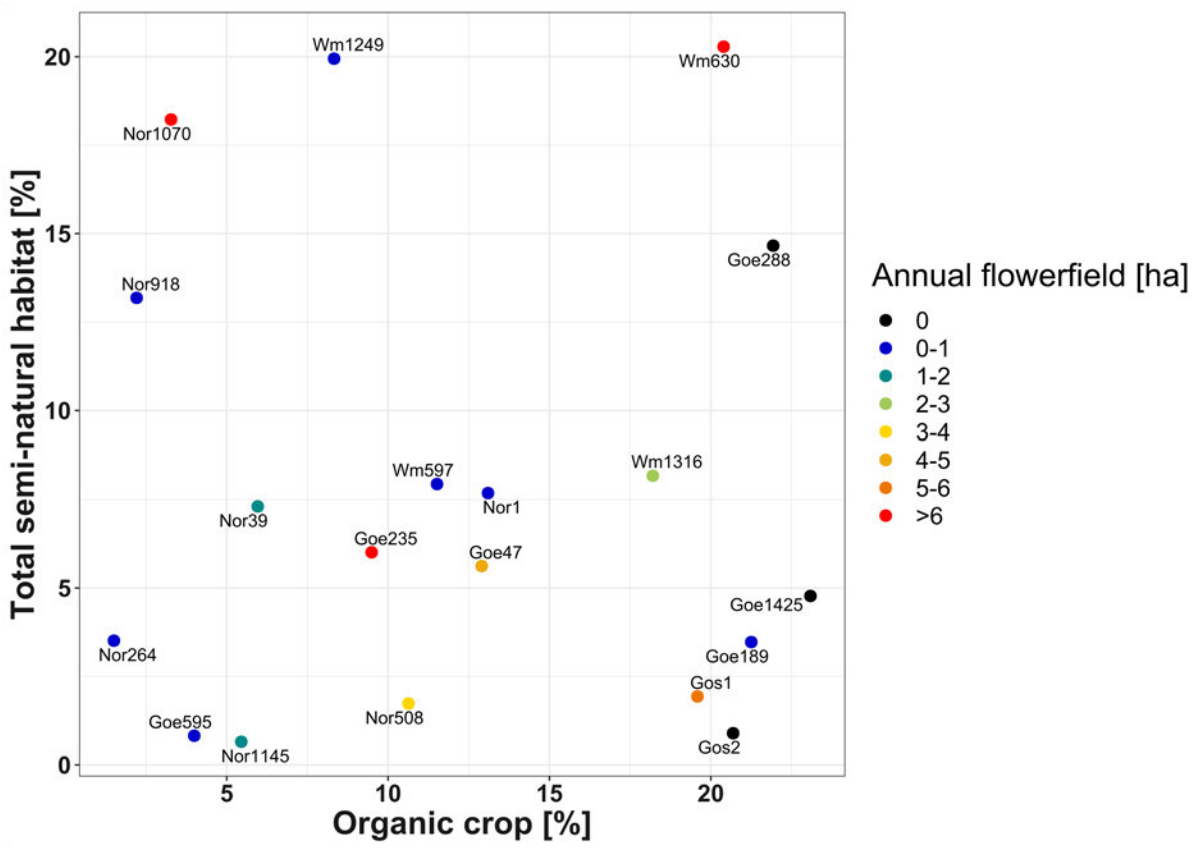


Figure 11: Classification of the 19 *ComBee* investigation sites regarding semi-natural habitat, organic crop percentage, and annual flowerfield area; Colors represent the annual flowerfield area; Goe = Göttingen, Gos = Goslar, Nor = Northeim, Wm = Werra-Meißner-Kreis (provided by the research group of Functional Agro-biodiversity, Georg-August-University Göttingen)

5 Results and Discussion

5.1 Pollen Collection Yields

In the course of this study, honey bee pollen samples were taken from 19 different investigation sites of the *ComBee*-project (depicted in the map of Figure 8). In general, four samples were collected at each location, except for Wm1249. At this location, only two full samples could be collected at one side of the hive, as the colonies on the other bee hive side were inactive.

In previous studies, pollen dry mass quantities of 0.015 g were determined to be appropriate for DNA extraction attempts (Bänsch, 2019; Prudnikow, 2021). However, some of the samples did not yield the required amount, presumably due to smaller pollen basket sizes of the honey bees at the corresponding location. In order to obtain sufficient quantities of starting material for the DNA extraction, the pollen samples of the same bee hive sampling side were pooled (see Section 4.1 Figure 7). The total weight of the pooled samples can be seen in Figure 12. The minimum amount of pollen was collected in sample Goe288_2 with 0.0185 g, the maximum amount in sample Gos1_1 with 0.0937 g. No significant effect of the landscape variables on the total weight of the pollen samples could be detected by generalized linear mixed model analysis. Overall, 37 samples remained for the further investigation process, since location Wm1249 could only be represented by one pooled sample (Wm1240_1).

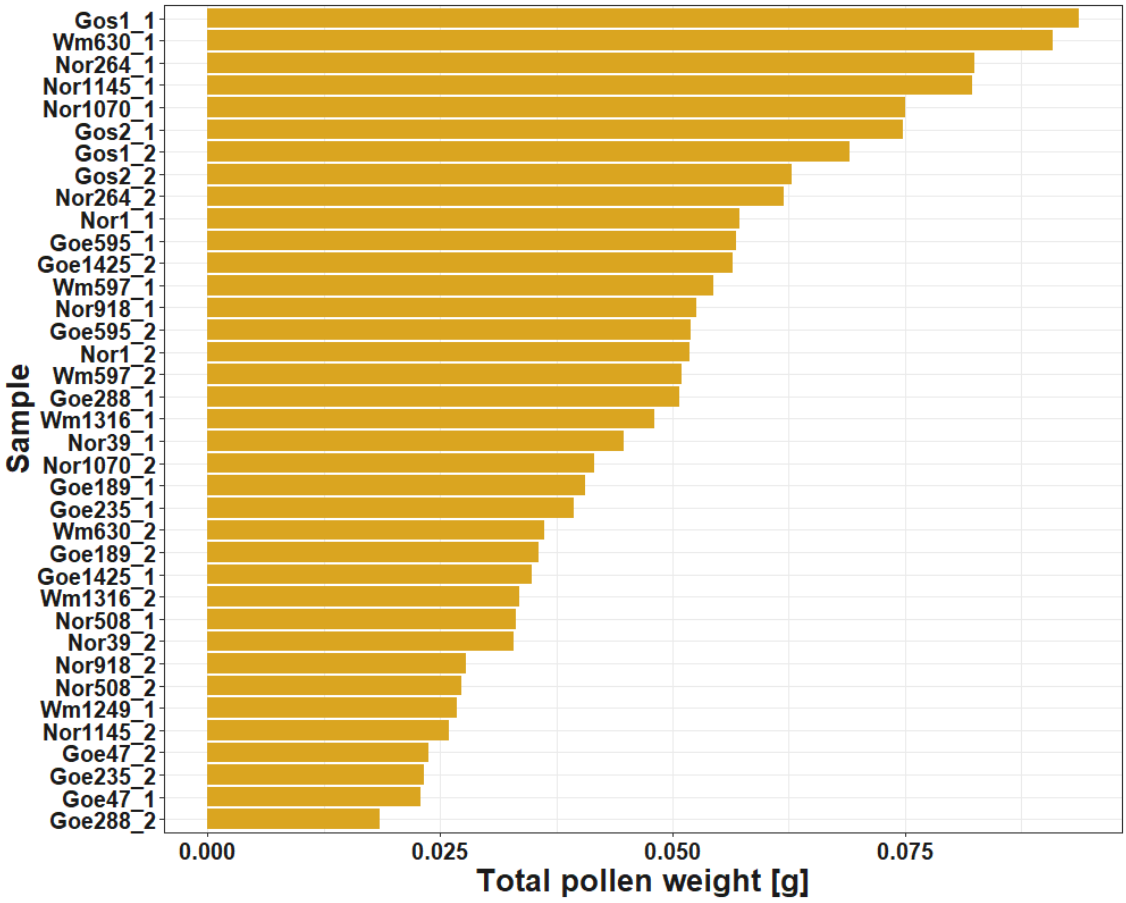


Figure 12: Total pollen weight of the pooled samples; Sampling locations: Goe = Göttingen, Gos = Goslar, Nor = Northeim, Wm = Werra-Meißner-Kreis

5.2 Assessment of the Pollen Disruption

The diversity of pollen structures is remarkable because of the different forms of the outer wall, also referred to as the pollen exine (see Figure 13). The term describes a robust layer of sporopollenin, a highly inert biopolymer consisting of fatty acids and phenolics (Ariizumi and Toriyama, 2011; Li et al., 2019).

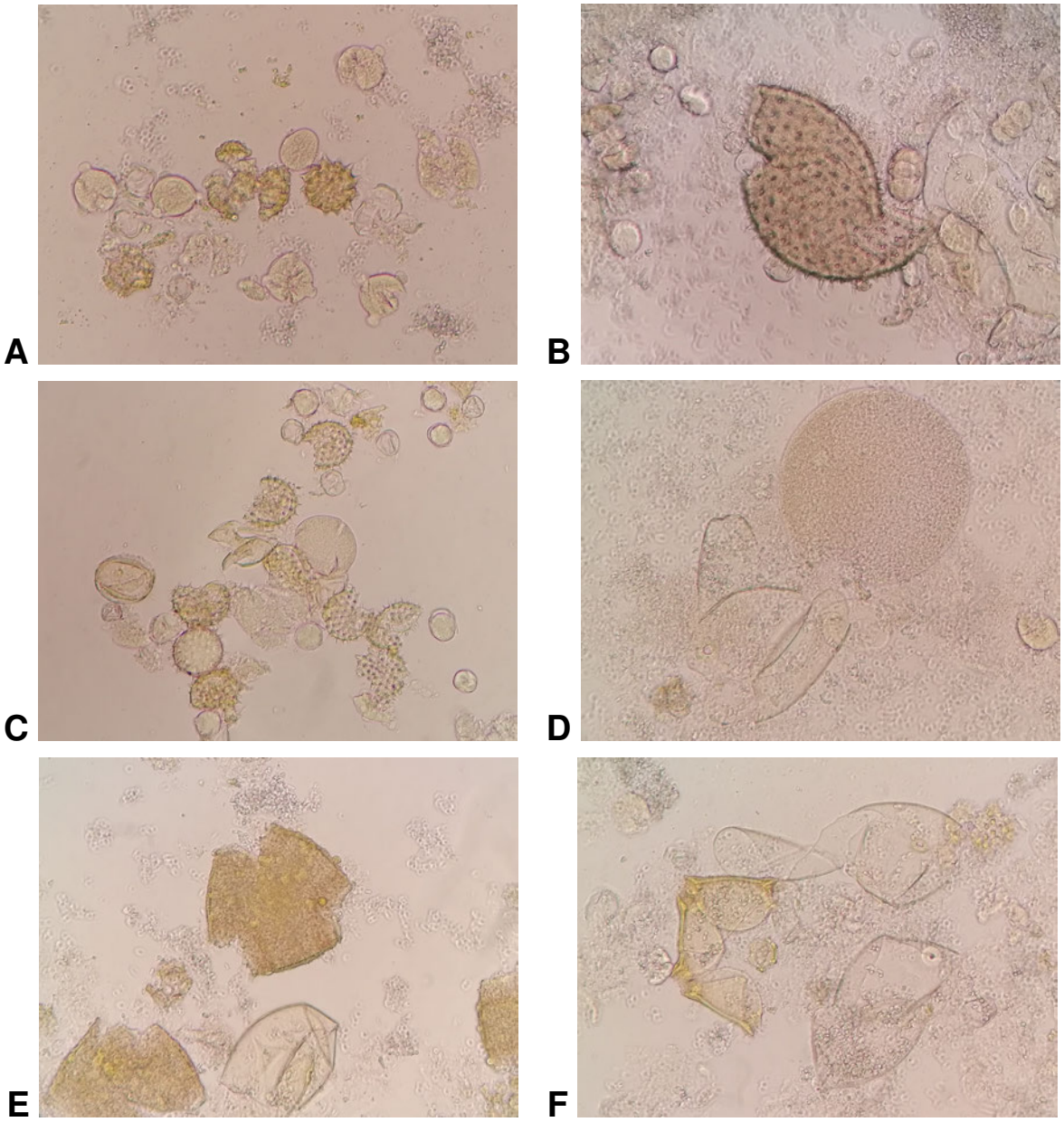


Figure 13: Microscopic assessment after the pollen disruption with ceramic beads; Magnification 10x40; A = Goe288_1; B = Goe595_1; C = Goe1425_2; D = Nor264_1; E = Nor264_2; F = Wm630_1

Especially for entomophilous plants, the exine not only protects the pollen interior from mechanical and environmental impacts (e.g. drastic temperature shifts, UV radiation or microbial pathogens) but also ensures the attachment and dispersion by pollinating insects (Ariizumi and Toriyama, 2011; Borg and Twell, 2011). In order to access the DNA from the inner content, the first challenge is the disruption of the pollen exine (Kraaijeveld et al., 2015). The utilization of two different sizes of ceramic beads of 1.4 mm and 2.60-3.30 mm was proven suitable for that purpose in previous studies (Bänsch, 2019; Prudnikow, 2021; Swenson and Gemeinholzer, 2021). As seen in the microscopic images of Figure 13 the exine of different types and sizes of pollen grains was successfully cracked. Even if the outer layer was only slightly damaged (e.g. Figure 13D) it still poses a target site for DNA extraction. However, especially smaller pollen grains or those with a more robust exine evaded the procedure occasionally (e.g. Figures 13A-C), which can bias subsequent results. In order to solve this problem, the use of more varying bead sizes or an increase of the vortexing time would be possible (Simel et al., 1996). Nevertheless, it was observed that the vast majority of the samples, including small pollen grains, was damaged sufficiently to extract DNA from the inner substance. Statistically, multiple grains of one plant are included in a sample and therefore the species' presence is reflected in later results (e.g. Figure 13A and C).

5.3 Evaluation of the two DNA Extraction Kits

To extract DNA from the pollen samples, the protocol of the DNeasy® Plant Mini Kit was performed, since it is widely applied among pollen metabarcoding studies (Galimberti et al., 2014; Hawkins et al., 2015; Bänsch et al., 2020b). The success was determined with a fluorometric and a spectrometric assay. Test samples from honey bee and wild bee pollen (see Figure 14) were investigated to test the protocol before proceeding with the actual samples.

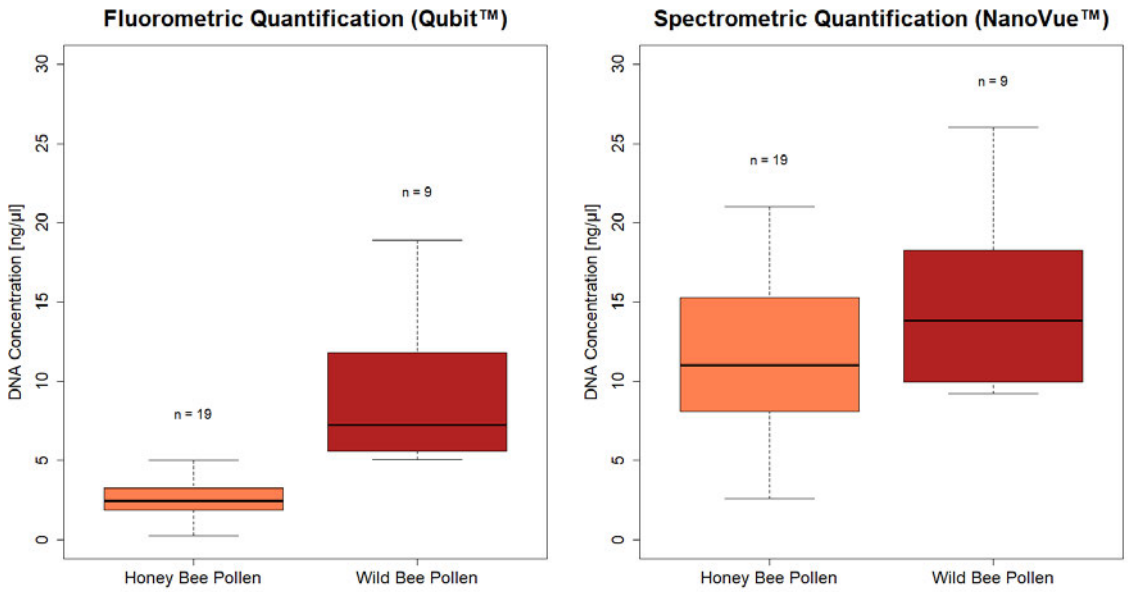


Figure 14: DNA concentration of honey bee and wild bee pollen samples after DNA extraction with the DNeasy® Plant Mini Kit measured with fluorometric (Qubit™) and spectrometric (NanoVue™) assay.

Figure 14 demonstrates that the DNA concentration of the honey bee pollen samples was generally lower than of the wild bee samples. The discrepancy is especially seen within the first graph of the Qubit™ assay whereas only a slight difference is visible within the NanoVue™ measurements. This is owed to the circumstance that the NanoVue™ Plus Spectrometer is based on the detection of absorbance at different wavelengths (230 nm, 260 nm, 280 nm and 320 nm). The method is therefore suitable to determine the purity of the nucleic acid solution but it is also prone to disturbances

caused by pollution when it comes to the estimation of DNA concentration (Matlock, 2015; Koetsier and Cantor, 2019). This fact explains the extensive differences in DNA concentration amounts between the spectrometric and the fluorometric assay. In contrast to NanoVue™, the use of a Qubit™ fluorometer requires a preparation of the samples with a fluorescent stain (Thermo Fisher Scientific Inc., 2016). This specific dye directly intercalates with the dsDNA which makes this assay a lot more sensitive and accurate (Thermo Fisher Scientific Inc., 2021). For that reason, the fluorometric quantification will be used for further DNA concentration assessment. The DNA concentration detected with the Qubit™ assay exhibited only 2 – 3 ng/µl on average for honey bee and between 5 – 19 ng/µl for wild bee pollen samples when diluted in 100 µl elution buffer. These values are a lot lower than indicated by the manufacturer, with an expected concentration of 38 ng/µl (Qiagen, 2020). However, the DNeasy® Plant procedures are optimized for the analysis of leaf samples which are better accessible for DNA extraction – other plant tissues may therefore vary from the benchmark values (Qiagen, 2020). An error during the work process could be ruled out since honey bee and wild bee pollen were treated during the same extraction runs and still wild bee samples exhibited higher concentration values. It is possible that the wild bee pollen samples were more diverse (Rollin et al., 2013; Bänsch, 2019; Urbanowicz et al., 2020) resulting in higher DNA amounts since pollen grain shape, size, and wall composition also affect DNA extraction success (Simel et al., 1996; Borg and Twell, 2011; Hawkins et al., 2015) Nevertheless, the DNA concentration of the honey bee samples was considered too low to ensure a correct representation of the sample which is why the NucleoSpin® Food Kit, recommended by Bell et al. (2017a), was investigated additionally during this study (see Figure 15).

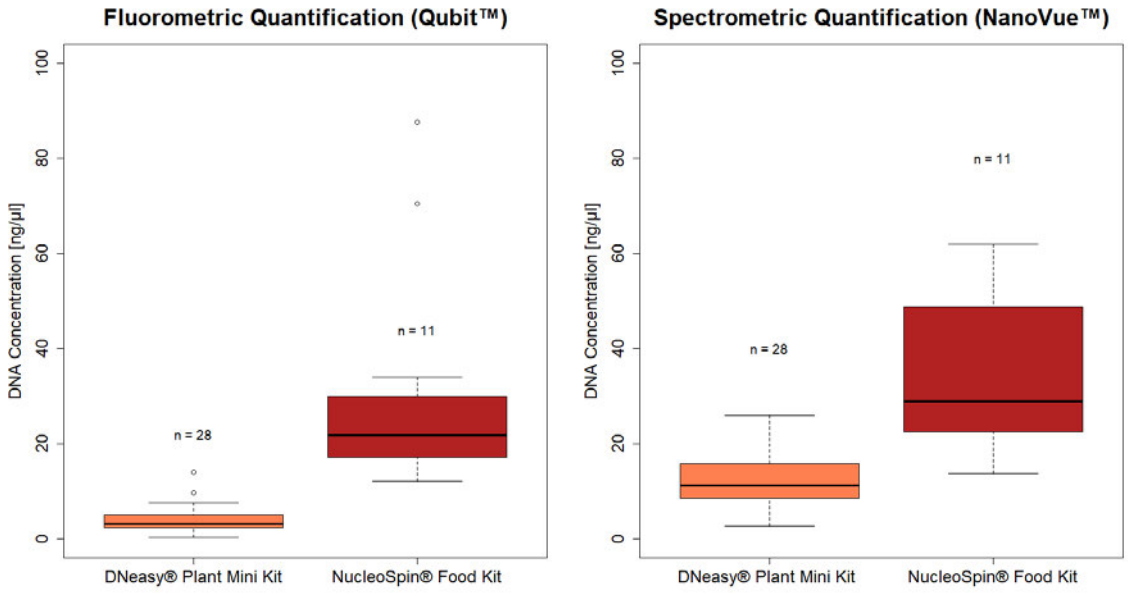


Figure 15: DNA concentration measured with fluorometric (Qubit™) and spectrometric (NanoVue™) assay after DNA extraction of wild bee and honey bee pollen samples with the DNeasy® Plant Mini Kit and the NucleoSpin® Food Kit. The results of the different bee species are graphically represented as one.

As visible in Figure 15, higher concentration values could be achieved by DNA extraction with the NucleoSpin® Food Kit than with the DNeasy® Plant Mini Kit. Again, the value range varied extensively between the fluorometric and spectrometric approach which is why only Qubit™ values were considered for DNA concentration assessment due to the higher accuracy of the method (Thermo Fisher Scientific Inc., 2016). With the NucleoSpin® Food Kit DNA concentration values between 13 – 88 ng/µl were obtained in the same elution volume (50 µl) as for the DNeasy® Plant Mini Kit. Furthermore, honey bee and wild bee pollen samples exhibited equally good concentration values. However, only a small amount of samples was used for this preliminary investigation, which is why the two subsets of honey bee and wild bee pollen are graphically represented as one. Nevertheless, the variation between wild and honey bee pollen DNA extraction should be investigated in subsequent studies.

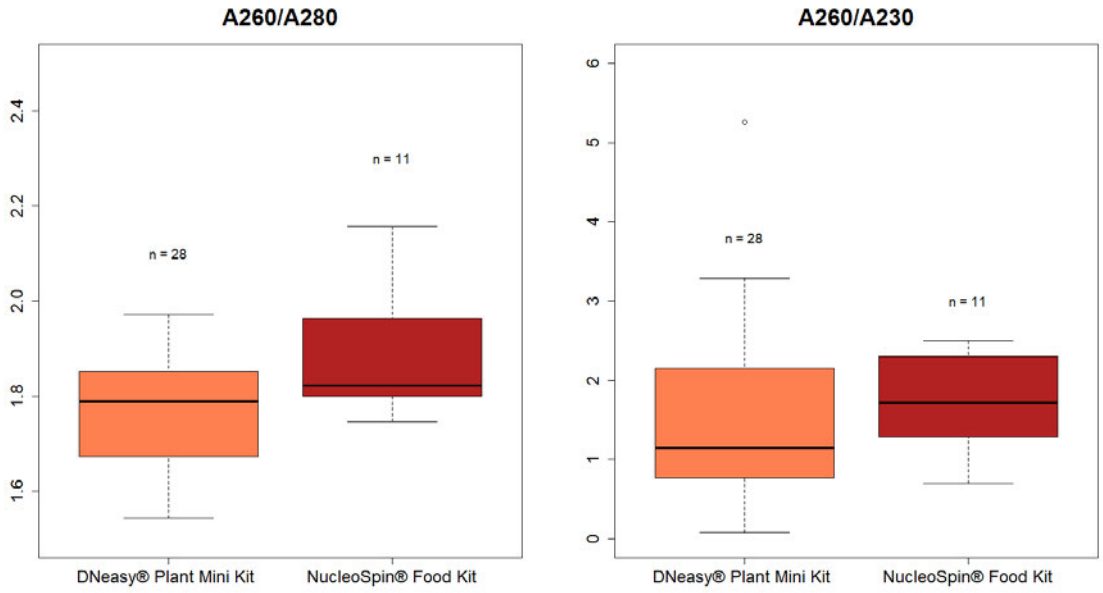


Figure 16: Determination of the DNA purity by ratios A260/A280 and A260/A230; Measurements were carried out with the NanoVue™ Plus Spectrometer after DNA extraction of wild bee and honey bee pollen samples with the DNeasy® Plant Mini Kit and the NucleoSpin® Food Kit. The results of the different bee species are graphically represented as one.

With the help of the NanoVue™ Plus Spectrometer the two kits were further examined regarding the generated DNA purity (Figure 16). Spectrometry is based on the specific light absorbance of different macromolecules. Thereby, the threshold ranges for pure nucleic acids are considered 1.8 – 2 for quotient A260/A280 and 2 – 2.2 for A260/A230 (Matlock, 2015; Koetsier and Cantor, 2019). In the case of this study, the A260/A280 ratio was generally adequate although the mean value of the DNeasy® Plant Mini Kit was a bit lower with 1.721 than the one of the NucleoSpin® Food Kit with 1.886. Apparently, the second kit also exhibits less variation in the A260/A280 ratio – the only exception was the first extraction with the NucleoSpin® Food Kit which exhibited values over 2. This could be explained by the fact, that the standard protocol is typically performed without an RNase and the samples therefore still contained a great amount of RNA (Macherey-Nagel, 2020) which was also shown by gel electrophoresis (see Figure 17). The problem was solved by adopting the RNase A treatment from the DNeasy® Plant

Mini Kit to the protocol of the NucleoSpin® Food Kit. In contrast to A260/280, the values for the A260/230 quotient varied extensively throughout the different samples. The threshold value between 2 – 2.2 could only be reached infrequently. This fluctuation can be caused by residuals from the extraction kit such as phenol and guanidine (Matlock, 2015; Koetsier and Cantor, 2019). Again, the NucleoSpin® Food Kit performed slightly better than the DNeasy® Plant Mini Kit and generally exhibited less deviation.

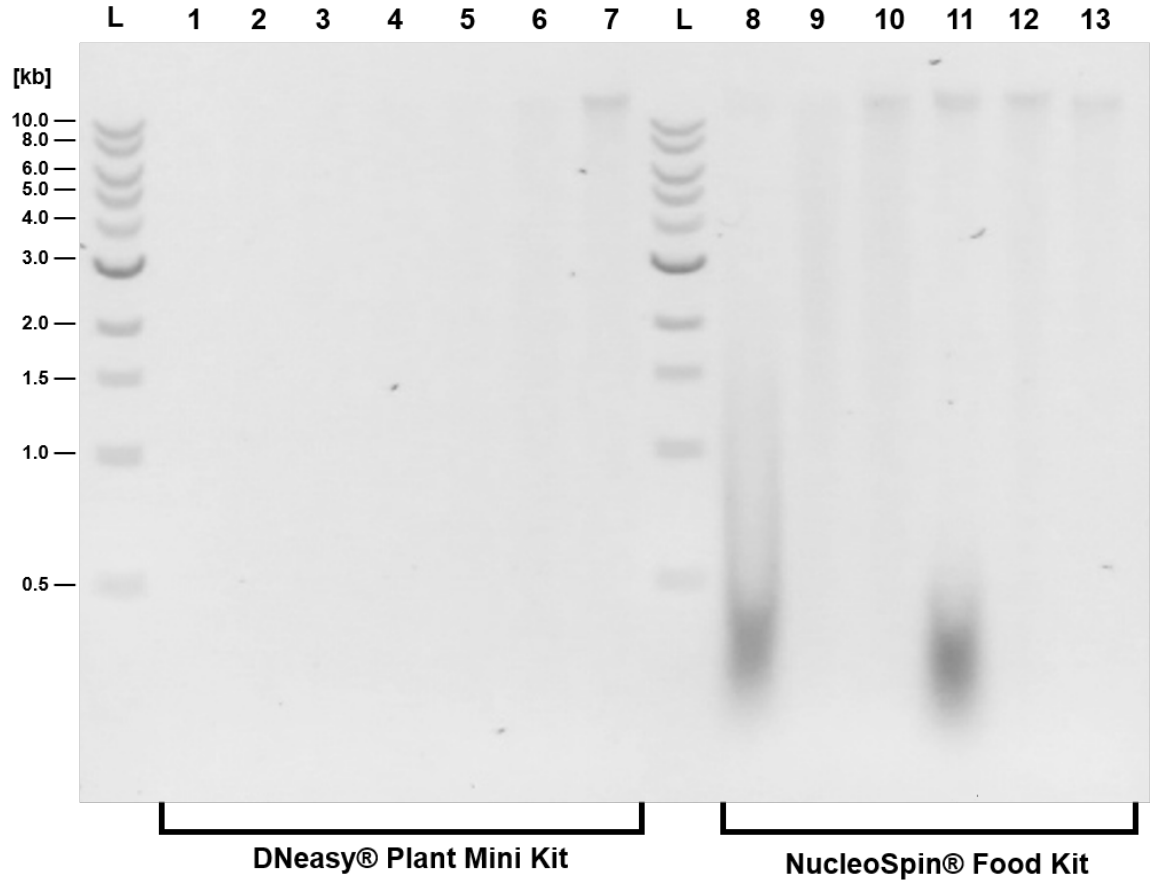


Figure 17: Gel electrophoresis after DNA extraction with the DNeasy® Plant Mini Kit (No. 1-7) and the NucleoSpin® Food Kit (No. 8-13). No. 8 and 11 without RNase A treatment. L = DNA Ladder. Agarose concentration 1%.

In addition to fluorometric and spectrometric measurements, agarose gel electrophoresis was conducted in order to assess the quality of the extracted DNA. According to Qiagen (2016) sizes ranging between 20 – 40 kb are expected. However, as seen in Figure 17 the DNA concentration gained by the DNeasy® Plant Mini Kit was gener-

ally too low to be depicted by electrophoresis (NO. 1-6). Only sample 7 (18.9 ng/µl) contained enough DNA to exhibit a clearly visible band over 10 kb in the gel. On the contrary, all extractions with the NucleoSpin® Food Kit (No. 8-13) exhibited traces of nucleic acids during electrophoresis. The two samples 8 and 11 that were extracted without the addition of RNase A can easily be distinguished due to the blurry bands below 0.5 kb which are characteristic for RNA (Jaakola et al., 2004). The other samples exhibit no such bands, instead they show only one stronger band over 10.0 kb, a sign for highly genomic DNA, and a blurry trace along the column which indicates different sizes of DNA fragments (Jaakola et al., 2004). In sample 9 no clear band over 10.0 kb is visible indicating that the DNA might be more fragmented in this sample which is a possible outcome according to the manufacturer (Macherey-Nagel, 2020). Since the barcode regions are still comparably small in size, this should not be a problem for the further analysis. The extensive differences between the two kits is difficult to explain since precise contents are not revealed by the manufacturing companies. One possible reason could be the utilization of an additional column within the DNeasy® Plant Mini Kit protocol. The QIAshredder column should serve as a first filtration step in order to remove cell debris and salt precipitates, however, it may have shearing effects on the DNA and could also be responsible for the low concentration yields (Qiagen, 2020). Nevertheless, since the NucleoSpin® Food Kit exhibited better results concerning DNA concentration and purity, this extraction kit was used exclusively for the samples from Göttingen during this study (see Figure 18).

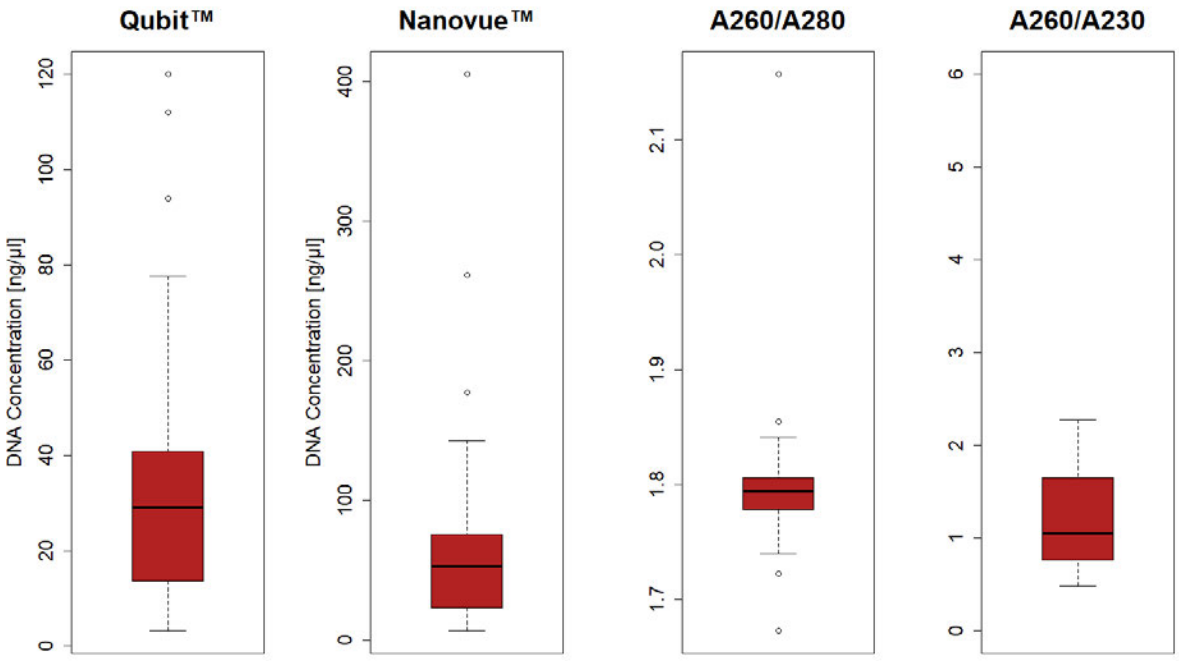


Figure 18: DNA extraction results for the honey bee pollen samples of the different agro-environments (n = 37). DNA extraction was carried out with the NucleoSpin® Food Kit. Quantity assessment by fluorometric Qubit™ and spectrometric Nanovue™assay. Purity evaluation by wavelength ratios A260/A280 and A260/A230.

Figure 18 displays the DNA extraction results for the honey bee pollen samples of the different agro-environments achieved with the NucleoSpin® Food Kit. The mean DNA concentration value measured with the Qubit™ fluorometer amounts to 33.15 ng/µl, whereas 70.65 ng/µl were detected on average by the NanoVue™ spectrometer. DNA purity measurements yielded 1.797 for the A260/A280 ratio which even excels the pureness levels of the test samples. The A260/A230 quotient still fluctuated, however, neglecting extreme outliers the mean value was 1.15. To conclude, the NucleoSpin® performed very well in this study and can be recommended for other pollen DNA extraction attempts.

5.4 PCR Amplification of the Barcodes

5.4.1 Design of *rbcL* reverse primer

In order to amplify the respective barcode regions across a wide range of plant taxa, the use of universal primers is essential. In the case of ITS, the primers ITS-1 and ITS-4 (White et al., 1990) were already examined for the purpose of plant DNA metabarcoding by Prudnikow (2021). The primer binding sites are situated within the conserved sequences coding for the 18S- and 26S-rRNA (see Figure 5) and account for an amplicon size of approximately 700 bp (Baldwin et al., 1995; Álvarez and Wendel, 2003; Wang et al., 2015a). For *rbcL* however, only the subsection *rbcLa* (see Figure 4) has been utilized for barcoding approaches (Cabelin and Alejandro, 2016; Laha et al., 2017; Richardson et al., 2015). Since nanopore sequencing was conducted in further processes of this study, an alternative reverse primer was designed with the NCBI Primer-BLAST tool in order to obtain an almost full length amplicon. The best suited results are depicted in Table 5.

Table 12 *rbcL* primer details (f = forward, r = reverse)

Name	Sequence (5'→3')	Length [nt]	Tm [℃]	GC [%]
rbcLa_f	ATGTCACCACAAACAGAGACTAAAGC	26	61.68	42.31
rbcL1_r	CTCCTTCCATACTTCACAAGCAGC	24	61.69	50.00
rbcL2_r	TTGATCTCCTTCCATACTTCACAAGC	26	61.08	42.31

The rbcLa_f primer exhibits a self complementarity of 3.00 and both reverse primers of 4.00. The self 3' complementarity accounts 2.00 for rbcLa_f, 3.00 for rbcL1_r, and 2.00 for rbcL2_r. The amplicon size of the first reverse primer amounts 1392 bp and 1397 bp for the second one. The primer sequences were designed with GC-clamps at the 3'-ends, which is recommended due to the more stable hydrogen bonds which allow correct and stable hybridization and thereby a reliable starting point for the polymerase enzyme (Nybo, 2013). When it comes to the evaluation of primers, different

criteria need to be assessed: primer length commonly ranges from 18-25 nucleotides, the self complementarity of the primer itself and the respective pair should be as low as possible. A GC content between 40-60% is regarded appropriate and the melting temperature should not exceed a range of 55-65 ℃ or vary too much within one primer pair (Hansen, 2004). The rbcLa_f primer with 26 nucleotides is comparatively long, nonetheless, its applicability across a wide range of plant species has been demonstrated in various studies (Kress and Erickson, 2007; Galimberti et al., 2014; Laha et al., 2017). The design of the reverse primers aimed to set the melting temperature to a maximum of 1 K variation compared to rbcLa_f. Of the two candidates identified, primer rbcL1_r exhibited the least temperature deviation with only 0.01 K compared to rbcL2_r with 0.6 K. GC contents of both primers were adequate and self complementarity values were acceptable. The final decision was made with regard to the position of the primers within the conserved regions identified by Clustal Omega. As seen in Figure 19 primer rbcL2_r exhibited more consecutive conserved bases at its 3' end in contrast to primer rbcL1_r. Since the correct matching of the 3' end is essential for polymerase hybridization, rbcL2_r was used for *rbcL* reverse strand amplification in this study.

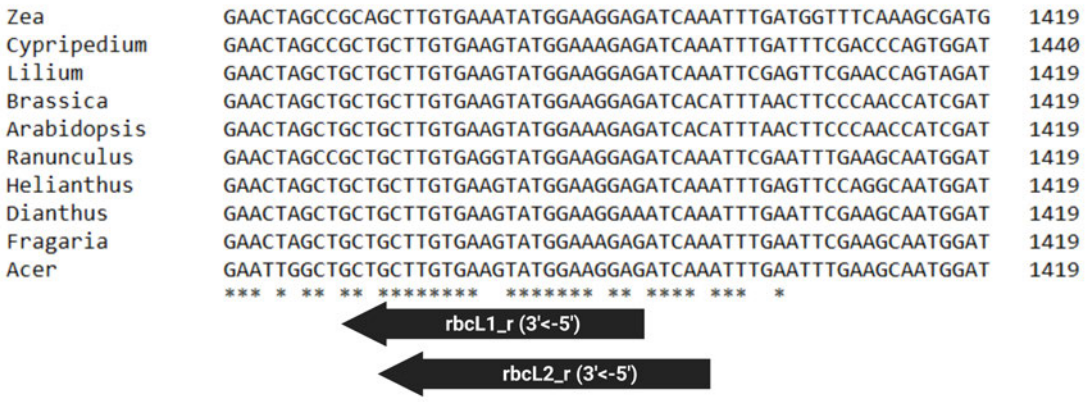


Figure 19: Schematic representation of the designed reverse primers (rbcL1_r and rbcL2_r) integrated into the Clustal Omega alignment.

5.4.2 Gel Electrophoresis Results

Each barcode was amplified using a specific protocol, described in detail in section 4.5. After the PCR step, the samples were purified and applied onto a gel (1%), together with the unpurified *no template control* (NTC). None of the electrophoresis gels exhibited bands within the NTC-section, which proves the purity the master mix.

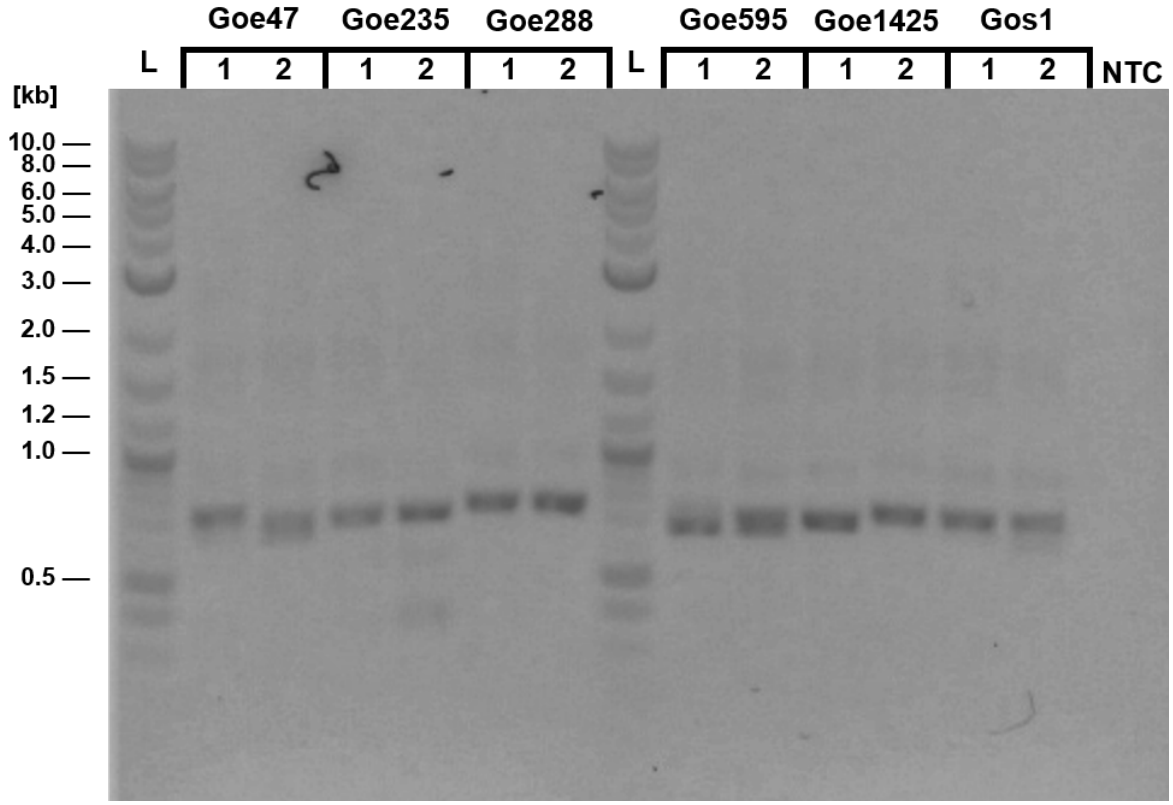


Figure 20: Electrophoresis gel of the purified ITS amplicons. L = DNA Ladder. NTC = no template control. Agarose concentration 1%.

Figure 20 and supplemental Figure S1 display the electrophoresis results for the purified ITS amplicons. The gels exhibit bands between 0.7 kb and 0.8 kb which is characteristic for the average length of ITS. However, all samples display blurry traces above 0.8 kb with vague bands at about 0.9 kb and between 1.5 – 2.0 kb. In addition, some samples show bands at a lower level around 0.6 kb (e.g. Gos2_2, Nor39_1 in Figure S1) and also at under 0.5 kb (Goe235_2). As a spacer region, ITS exhibits length variations to a greater extent than a gene sequence, since it is less conserved and more

susceptible to mutation events such as insertions and deletions (Bolson et al., 2015; Wang et al., 2015a).

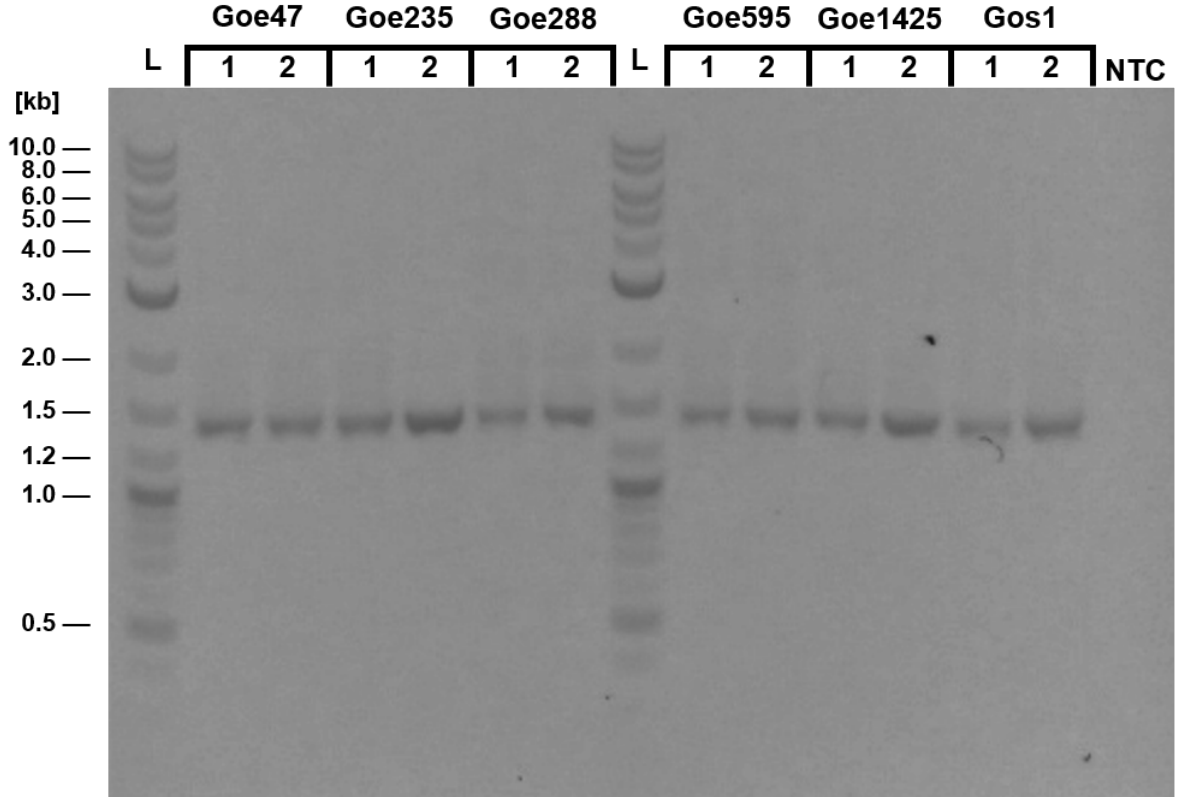


Figure 21: Electrophoresis gel of the purified *rbcL* amplicons. L = DNA Ladder. NTC = no template control. Agarose concentration 1%.

Figure 21 and supplemental Figure S2 display the purified *rbcL* amplicons. A clear band is visible closely under 1.5 kb length, which resembles the expected amplicon size between 1400-1500 bp (Bell et al., 2017b; Newmaster et al., 2006) and proves the correct hybridization of the newly designed reverse primer. In contrast to ITS, no other band sizes are visible in the gel, only a short blurry trace above 1.5 kb, which derives from longer sequence amplicons that are generated the beginning of a PCR.

5.5 Sequencing with the MinION Mk1B

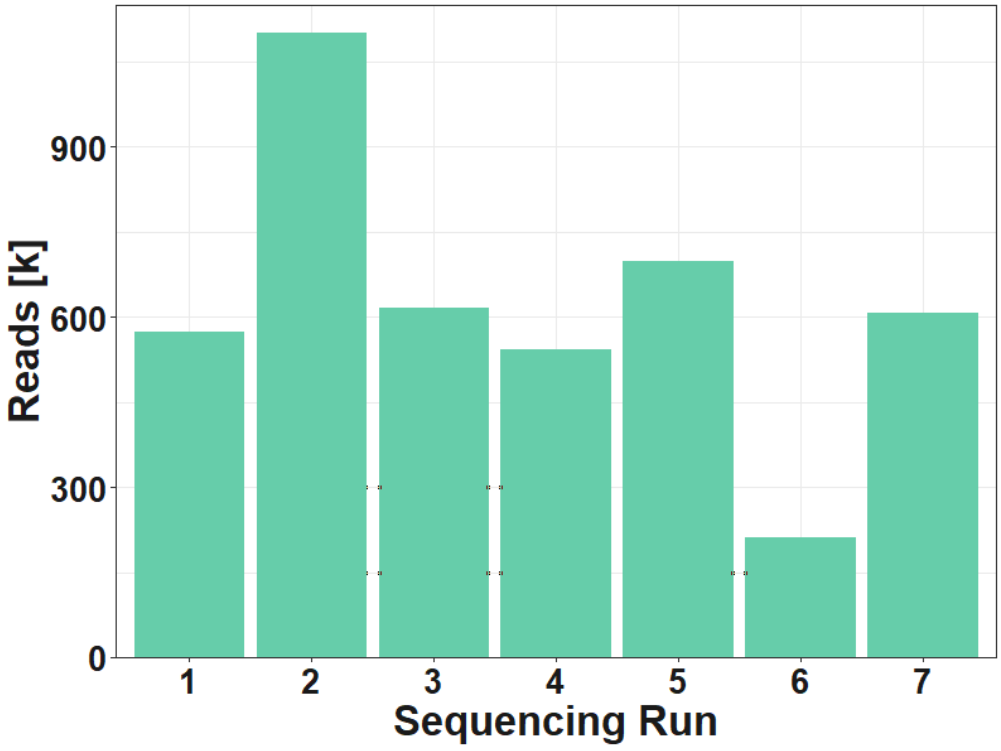


Figure 22: Amount of generated reads for each sequencing run. No. 1-3: ITS samples. No. 4: ITS and *rbcL* samples. No. 5-7: *rbcL* samples.

Sequencing was carried out with Flongle flow cells and seven multiplexed runs were performed in total to sequence all samples. A summary of the different runs is displayed in supplemental Table S2. The amount of generated reads displayed in Figure 22 varied from 212.28 k to 1.1 M with a mean value of 621.62 k. This variation can be explained by the quality of the flow cells depending on the number of available nanopores. The sequencing run with the lowest read amount (No. 6) for example was carried out with 56 pores available at the start, in contrast to the one with the highest read number (No. 2), which started with 91 active nanopores. An additional factor for the low read quantity in sequencing run No. 6 might be disturbances while loading the flow cell. Even a small amount of air can be detrimental to the protein pores and have a negative impact on the success of the sequencing run (Oxford Nanopore Technologies, 2021).

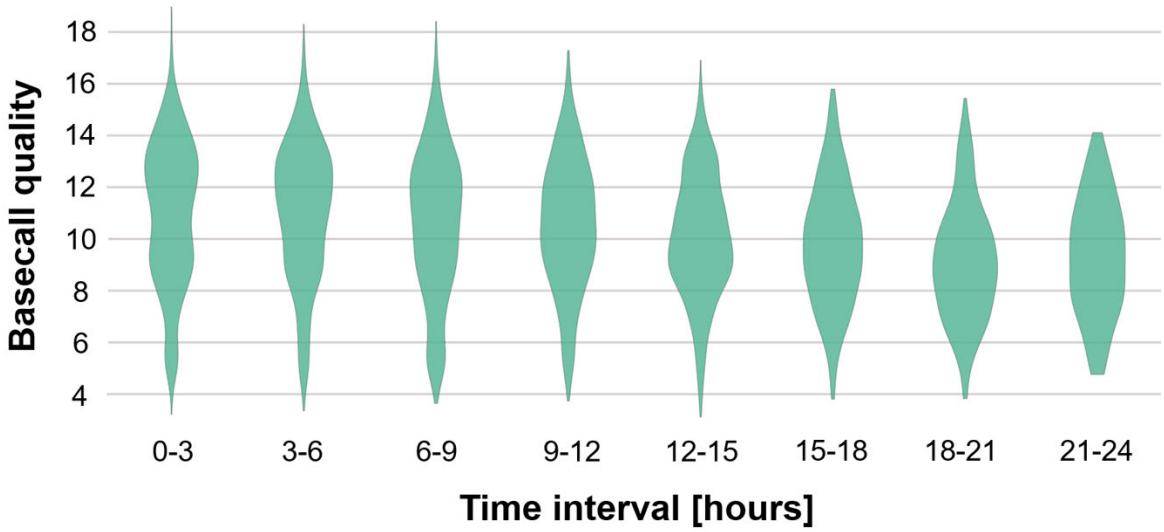


Figure 23: Representation of the basecalled quality over time of sequencing run No. 4 (created with *NanoPlot*).

Figure 23 displays the shift of the basecalled quality over time. In the beginning, the violin plot exhibits density curves at about 13 and one at 9. After 3-6 hours, the curve below 10 flattens, while the other one is still present. However, in the interval of 6-9 hours the upper curve is reduced as well and it can be observed that the Q-score decreases continuously until the main curve falls below a Q-score of 10. To conclude, the read quality declines over time during a nanopore sequencing run, which poses a problem when working with nanopore sequencing techniques. In this study, the risk of false positive classification results was reduced by using high-accuracy basecalling and by filtering the outcome (Delahaye and Nicolas, 2021). The filter was set for sequences with a Q-score of at least 15 (3.16% error rate) for 70% of the bases while the rest of the sequence should not fall below a Q-score of 10 (10.0% error rate). Nevertheless, the mean Q-score of the different sequencing runs varied between 9.1 and 10.6 which resembles literature descriptions (Delahaye and Nicolas, 2021; Wang et al., 2021). The reason for this relatively high error probability might be assigned to the proportion of GC content on the one hand and the occurrence of short repeated regions on the other (Delahaye and Nicolas, 2021). One approach for solving this problem

is including the sequencing depth in the downstream analysis (Delahaye and Nicolas, 2021). However, this is only suitable for genome assemblies and not for metabarcoding attempts where every read stands for an individual taxonomic result. Most promise lies in the improvement of the corresponding technology and the development of new software tools since nanopore technologies have and will experience further investigations and reforms in the future (Delahaye and Nicolas, 2021; Lee et al., 2021; Rang et al., 2018). Read accuracy already improved extensively from under 60%, at the time when nanopore sequencing was first introduced (Rang et al., 2018), to currently on average 90% (Delahaye and Nicolas, 2021; Wang et al., 2021). Further error rate reduction is approached by software revision (Delahaye and Nicolas, 2021) and by the investigation of new materials such as graphene, to create nanopores with a very thin transit channel which enables a more precise nucleotide detection (Deamer et al., 2016). However, the chance to sequence full-length barcoding regions in one working step is a great advantage for metabarcoding assays (Leidenfrost et al., 2020). This long-read aspect in combination with the lower cost and size of ONT techniques hold great potential not only for barcoding, but also for transcriptomic and epigenetic analyses (Leidenfrost et al., 2020; Peel et al., 2019; van Dijk et al., 2018).

5.6 Evaluation of the DNA Barcodes

To compare the outcomes of the DNA barcodes, two different aspects were analyzed: the species richness and the relative read abundance. The term species richness is defined as the number of different species, irrespective of their frequency within a habitat or sample (Gotelli and Colwell, 2001). Therefore, every species is equally weighted since each taxonomic name is only counted once. Species richness analysis was used as a first step, to estimate the percentage of species that both barcodes had detected compared to the proportion of species that one of the barcodes identified exclusively (represented in Figure 24). The species frequency was then included in the second step by analyzing the species read abundance. Though the classical concept of ecological abundance describes the total number of organisms within a certain habitat (Gotelli and Colwell, 2001), this study will use the term to signify the quantity of reads within a sample (displayed in Figure 25). Species richness and read abundance analysis were both divided into three subgroups: (I) species detected only by *rbcL*, (II) species detected only by ITS, and (III) species detected by both DNA bacodes. Furthermore, the two analyzing steps were carried out on the genus level.

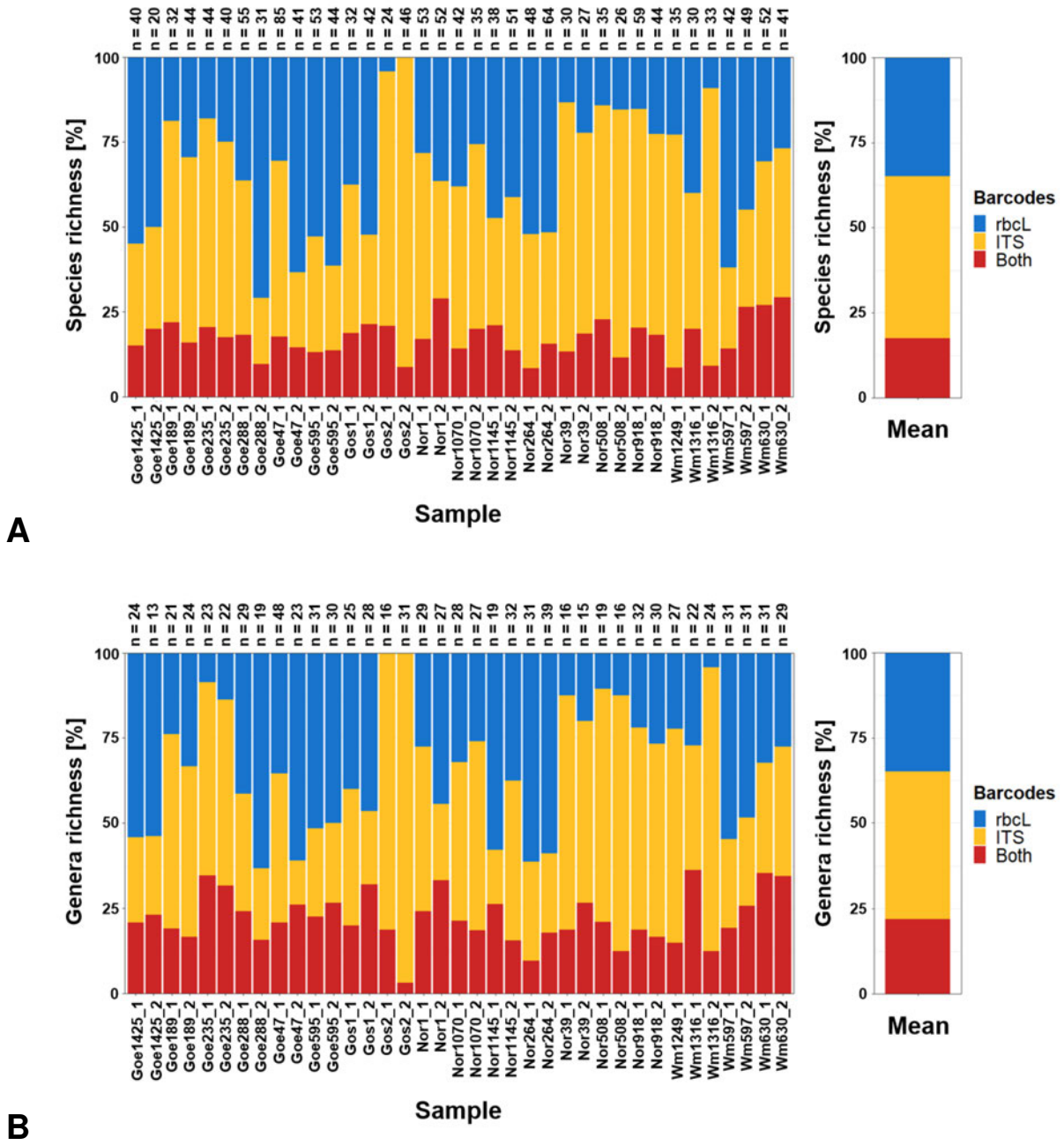
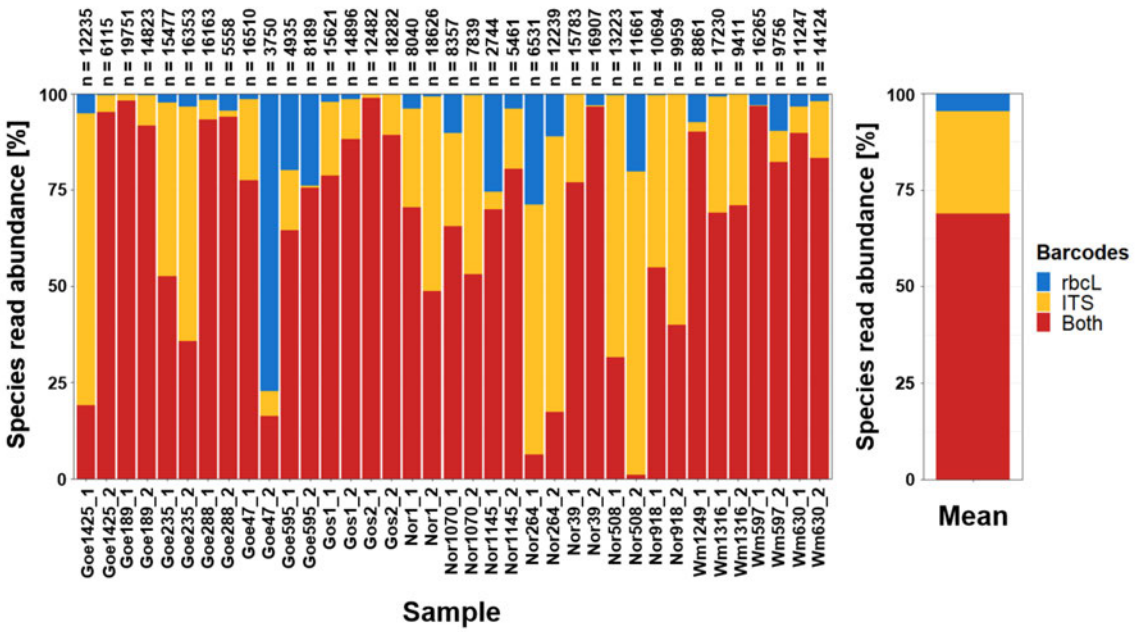


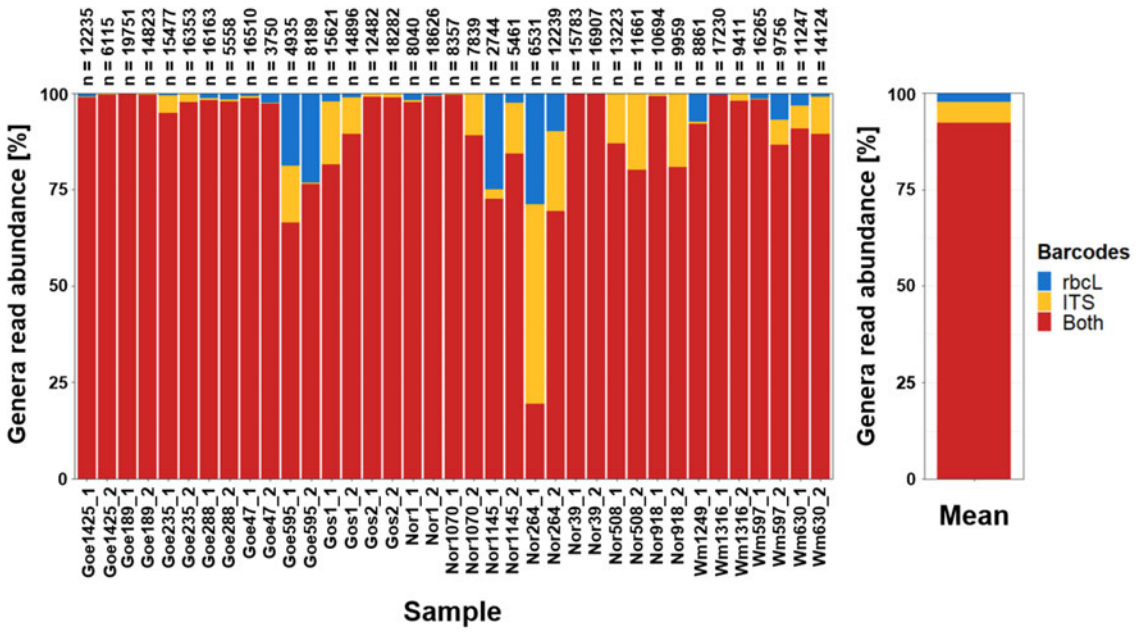
Figure 24: Comparison of the two barcodes ITS and *rbcL* regarding species (A) and genera richness (B) in percent. Representation of the individual samples (left) and the mean proportion (right). The subcategories display the percentage of species detected solely by *rbcL* (blue), by ITS (orange), and the intersection proportion of both barcodes (red).

Figure 24A represents the relative species richness divided into the three subcategories (*rbcL*, ITS, and Both). The mean amount of species names occuring in both barcode results comprises 17.6%, while 47.5% were present exclusively in the ITS and 34.9% in the *rbcL* subset. The number of different species ranged from 20 (Goe1425_2) to 85 (Goe47_1). The proportions of the three subcategories were not changing significantly when looking at the genera richness (see Figure 24B), where on average 21.9% of genera names were found in both barcode subsets whereas 43.3% were present in ITS and 34.8% in *rbcL* exclusively. The number of different genera ranged from 13 (Goe1425_2) to 48 (Goe47_1).

Comparing the analysis of species and genera richness, the percentage of *rbcL* genera barely changes in contrast to ITS, which shows that ITS was able to detect a greater variety of different species within a genus group. This observation can be attributed to the high variability of ITS as a spacer region, which is less conserved and therefore more susceptible to mutation events than the gene sequence of *rbcL* (Bolson et al., 2015; Wang et al., 2015a). Another influencing factor for the discrepancy between *rbcL* and ITS is the degradation process of plastid DNA during pollen development (Matsushima et al., 2011; Sakamoto and Takami, 2018). In fact, the relative number of *rbcL* detected species was on average lower than of ITS and from a total of 269 different plant species identified in this study, 83 were detected by *rbcL* and 112 by ITS exclusively. However, when looking at the genus level, a total of 141 different genera were detected, thereof 49 only by *rbcL* and 40 solely by ITS. To conclude, *rbcL* detected less different species but more plant genera than ITS. The result resembles literature descriptions, that ascribe less discriminatory power to *rbcL* concerning species level classification but great potential regarding plant genera identification (Cabelin and Alejandro, 2016; Kress and Erickson, 2007; Richardson et al., 2021). To further investigate the extent of difference between the two barcoding subsets, the read abundance was taken into account, displayed in Figure 25.



A



B

Figure 25: Comparison of the two barcodes ITS and *rbcL* regarding the read abundance of species (A) and genera (B) in percent. Representation of the individual samples (left) and the mean proportion (right). The subcategories display the relative abundance of species detected solely by *rbcL* (blue), by ITS (orange), and the intersection proportion of both barcodes (red).

Figure 25A represents the species read abundance with the three subgroups *rbcL*, ITS, and the intersection of both barcodes. The sample read amount varies between 2744 (Nor1145_1) and 19751 (Goe189_1). This discrepancy is caused by differences in read amount and quality during the different sequencing runs (discussed in Section 5.5), which is why only the relative read abundance will be examined (Gotelli and Colwell, 2001). In contrast to the results for the species richness, the mean barcode abundance consensus is markedly higher with 68.9%. ITS species account for 26.5% and *rbcL* for 4.6% of the remaining amount. When looking at the relative abundance of the genera in Figure 25B, the percentage that both barcodes have in common increases to 92.3%. Only 5.3% of the genera still occur exclusively in ITS and 2.4% in *rbcL*. To conclude, the two barcodes seem to differ significantly when only looking at the species richness. These findings could result from different levels of taxonomic resolution, sequencing success, or data availability of the respective barcode (Bell et al., 2017a). Nevertheless, when taking the abundance into account as well, the majority of classifications reach a higher consensus, especially on a genus level. This shows that species differences between the two barcodes only accounted for a small percentage of the whole sample. To conclude, the results show that the most abundant species in the outcome of the DNA barcodes coincide. However, the two-locus approach also enabled the representation of less abundant species that were mainly detected by a single barcode but are important to portray a more diverse picture regarding species richness (Bell et al., 2017a; Richardson et al., 2021).

5.7 Statistical Analysis and Diversity Assessment

During this study three different types of landscape variables were examined in a 1 km radius around the bee hives: the organic crop proportion, the percentage of semi-natural habitats, and the annual flowerfield area. To test the effect of these variables against the results of this study, generalized linear mixed modelling was conducted. The results for the two DNA barcodes ITS and *rbcL* were thereby combined.

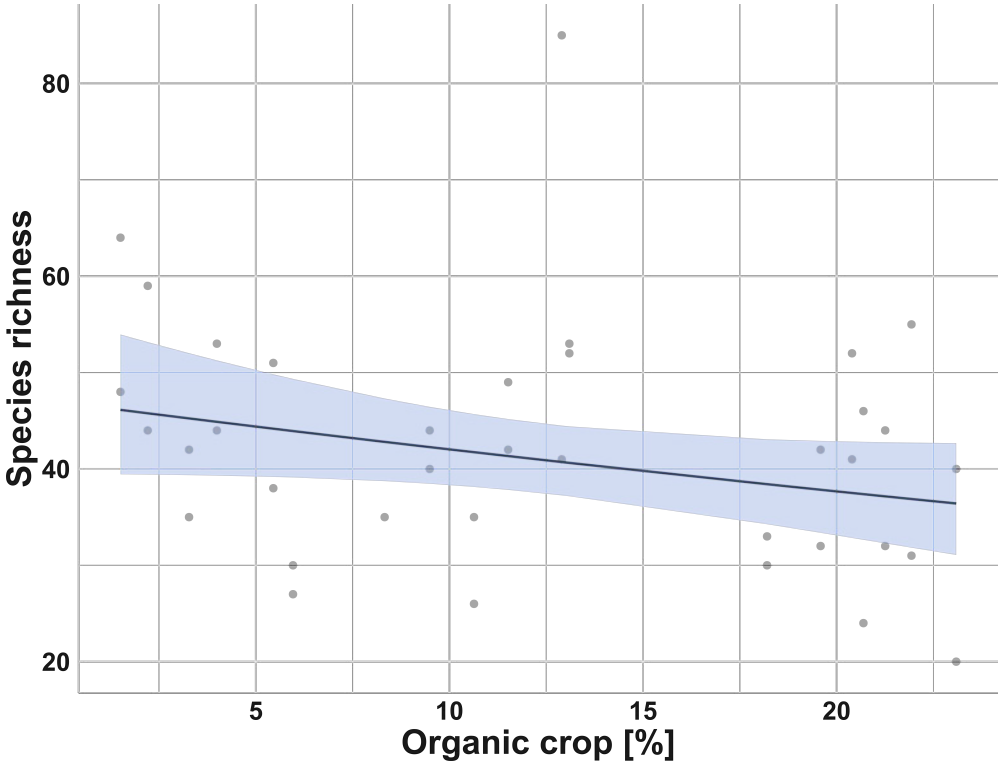


Figure 26: Effect of organic crop proportion on species richness related by a generalized linear mixed model. Species richness represents the barcoding results of *rbcL* and ITS combined.

Regarding species richness, a marginally significant outcome could be detected concerning the organic crop proportion (see supplemental table S3). As visible in Figure 26, the graph exhibits a negative correlation, indicating that species richness decreased with growing proportion of organic farming area. This circumstance might be explained by the late sampling time in August, when the main flowering period of many plant species was already over (Melgar et al., 2012; Werchan et al., 2018). This is

underlined by the fact that the proportion of annual flowerfield had no significant effect on species richness. Furthermore, tree species, which are an important pollen source earlier in the year and might have contributed to species richness enhancement, could not be recorded this late in the season (Melgar et al., 2012; Saunders, 2018; Werchan et al., 2018).

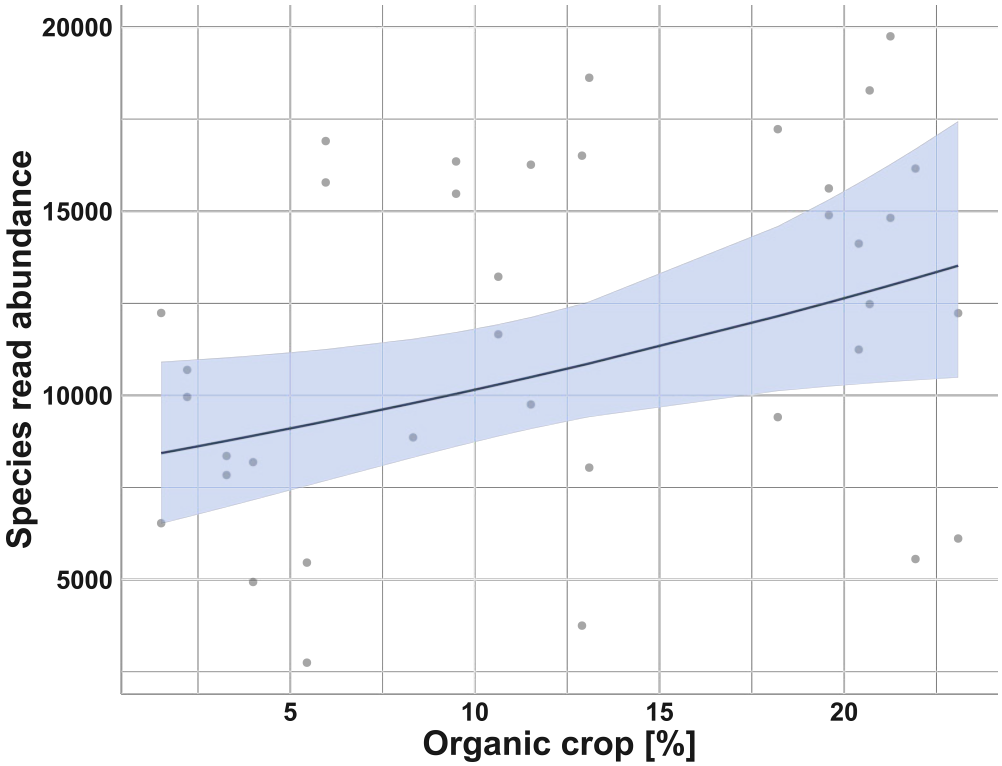


Figure 27: Organic crop correlation effect on species read abundance generated by a generalized linear mixed model. Species read abundance was thereby calculated based on the combined barcoding outcome of *rbcL* and ITS.

The examination of the species read abundance generated a significant result for organic crop proportion (see supplemental Table S4). A positive correlation could be observed in Figure 27, representing an increase of species read abundance with growing organic farming area. However, this result needs to be considered with caution, since the success of DNA extraction, PCR amplification, and sequencing method can influence the read outcome and complicate conclusions about quantitative species abundance (Bänsch et al., 2020b; Lamb et al., 2019). Furthermore, polyploidy and barcode copy number is another factor to consider when it comes to quantitative predictions,

since multiplications of a genome enlarge the number of reads but only represent one species (Bell et al., 2019; Peel et al., 2019). There are different attempts to eliminate these factors. So far, the combination with microscopic optical pollen analysis is one possible approach (Bänsch et al., 2020b; Leontidou et al., 2017; Richardson et al., 2021). In addition, upcoming automated classification processes by flow cytometry in combination with deep learning could add new perspectives to that field as they diminish the expenditure of time and expertise needed for pollen classification (Dunker et al., 2021; Olsson et al., 2021; Sevillano et al., 2020). In order to improve metabarcoding itself, the application of multiple DNA barcodes was suggested by Richardson et al. (2019), though this also multiplies the workload. Peel et al. (2019) presented an approach for reverse metagenomics (RevMet) but the pipeline needs further adaption prior to wide-range application on mixed samples. Shotgun-metagenomics, which signify the sequencing of random genome sections of a mixed sample are not considered fruitful, since the classification requires extensive databases and computing effort in the downstream analysis (Peel et al., 2019). A more promising approach would be a non-PCR based protocol for barcode extraction, which might be in reach using the new CRISPR/Cas systems – however, this attempt has only been conducted for single-species detection so far and poses challenges regarding the identification of species-wide applicable barcode cutting sites (Williams et al., 2019). It can thereby be concluded that the optimal method for reliable quantitative pollen barcoding is still subject of discussion.

To estimate the diversity within a specific environment, the evaluation of the Shannon index (H') is a common practice, thereby combining species richness and abundance (Danner et al., 2017; Richardson et al., 2021). The calculation is based on the proportion of individuals (P_i) as displayed in Formula 1.

$$(1) \qquad Shannon\ index: \quad H' = -\sum(P_i) \times ln(P_i)$$

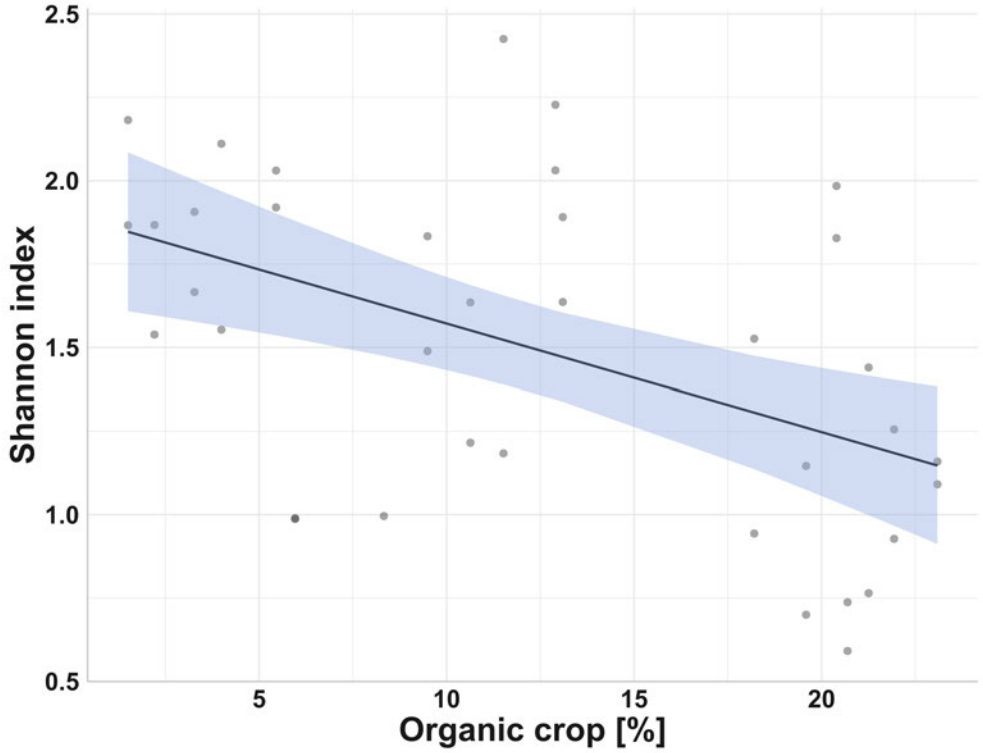


Figure 28: Impact of organic crop proportion on the Shannon diversity index estimated by generalized linear mixed model analysis. The Shannon index thereby includes the the combined barcoding results of *rbcL* and ITS.

The model shows that organic crop proportion had a significant effect on the Shannon index (see Figure 28 and supplemental Table S5). The graph displays a negative correlation, thus, the Shannon diversity index decreases with higher proportion of organic farming area. Again, this might be traced back to the late sampling time at the beginning of August, after the main annual flowering period (Melgar et al., 2012; Saunders, 2018; Werchan et al., 2018). However, it could also indicate the need for more diversity enhancing measures for this season. Flower strips are an important method to enhance species diversity during mass-flowering periods and avoid nutritional mismatches when flowering sources are rare (Haaland et al., 2011; Scheper et al., 2015). However, the main focus of attention in the past was drawn on pollen quantity instead of quality (Haaland et al., 2011; Pamminger et al., 2019b). First suggestions for improvement on that matter describe complementing approaches. For example, the sunflower (*Helianthus annuus*), which also belonged to the most dominant species in this study

(see Figures 30 and 31), provides pollinators with high-quality nectar as an energy source but is otherwise poor in nutrients (Frias et al., 2016; Pamminger et al., 2019a). Maize (*Zea mays*) as a wind-pollinated species on the contrary is low in nectar as well as protein content (Pamminger et al., 2019b). Both plants are important crops with wide-ranging field areas – in order to supply nearby pollinators with the nutritional amount they need for a healthy development, these fields could be surrounded with flower strips of complementing plant species. More precisely, pollinators close to sunflower fields would profit from plants with higher protein content, while a maize field can be compensated with plant species supplying both nutrients and nectar (Pamminger et al., 2019b). However, this approach presupposes the knowledge about the nutritional quality of many different plant species for pollinators. More studies are needed on that matter, since pollen quality can differ significantly within plant families, genera, and even habitats (di Pasquale et al., 2013; Frias et al., 2016; Pamminger et al., 2019a). Nevertheless, complementary flower strips can pose a promising approach in order to gain diversity enrichment and provide pollinators with the nutrients they need for a healthy development (Huang, 2012; Pamminger et al., 2019b).

In regard to biodiversity, the Shannon index itself is an estimation, however, it does not comprise information about the species distribution within a habitat. For this reason, species evenness was calculated as presented in Formula 2 based on the Shannon index (H') divided by its maximum possible value (H'_{max}), which equals the natural logarithm of the total number of species (S).

$$(2) \qquad Evenness: \; E = \frac{H'}{H'_{max}} \qquad H'_{max} = ln(S)$$

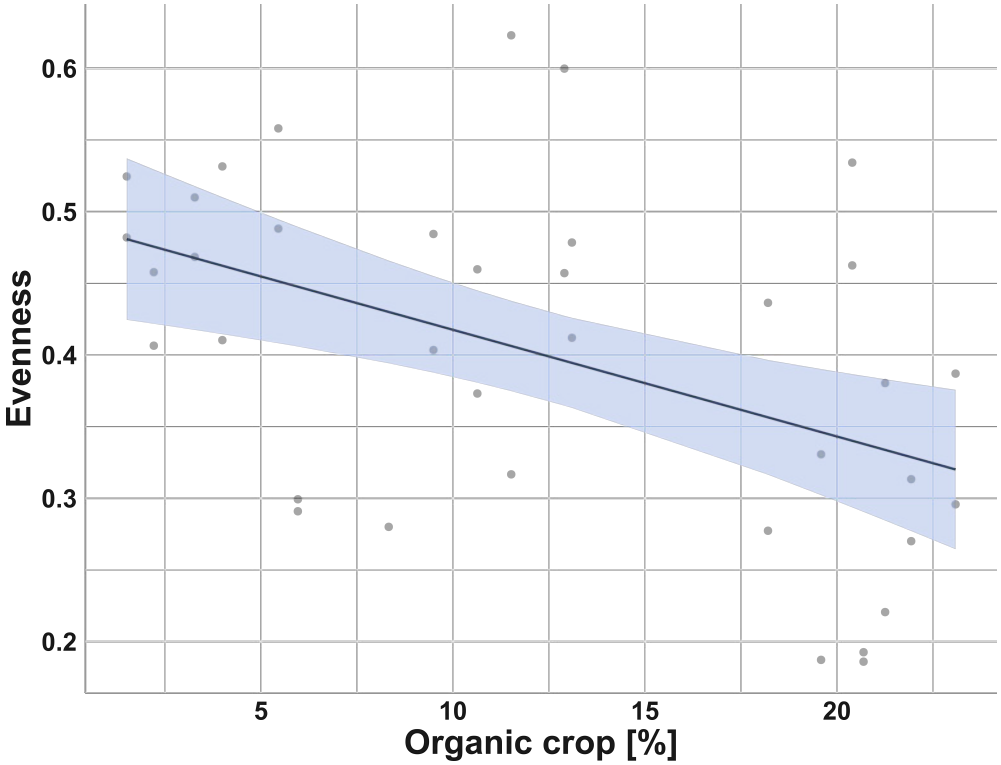


Figure 29: Effect of organic crop percentage on species evenness related by generalized linear mixed modelling. Species evenness was calculated based on the combined barcoding outcomes of *rbcL* and ITS.

The generalized linear mixed model resulted in a significant effect regarding organic crop and species evenness (see Figure 29 and supplemental Table S6). The graph shows that species evenness declines with rising organic crop proportion indicating more dominant species in organic farming areas. The following graphs display the 30 most prevalent species, divided into their presence in the different samples (Figure 30) as well as regarding the total read outcome (Figure 31). The graphs represent the combined results of the two barcode subsets *rbcL* and ITS.

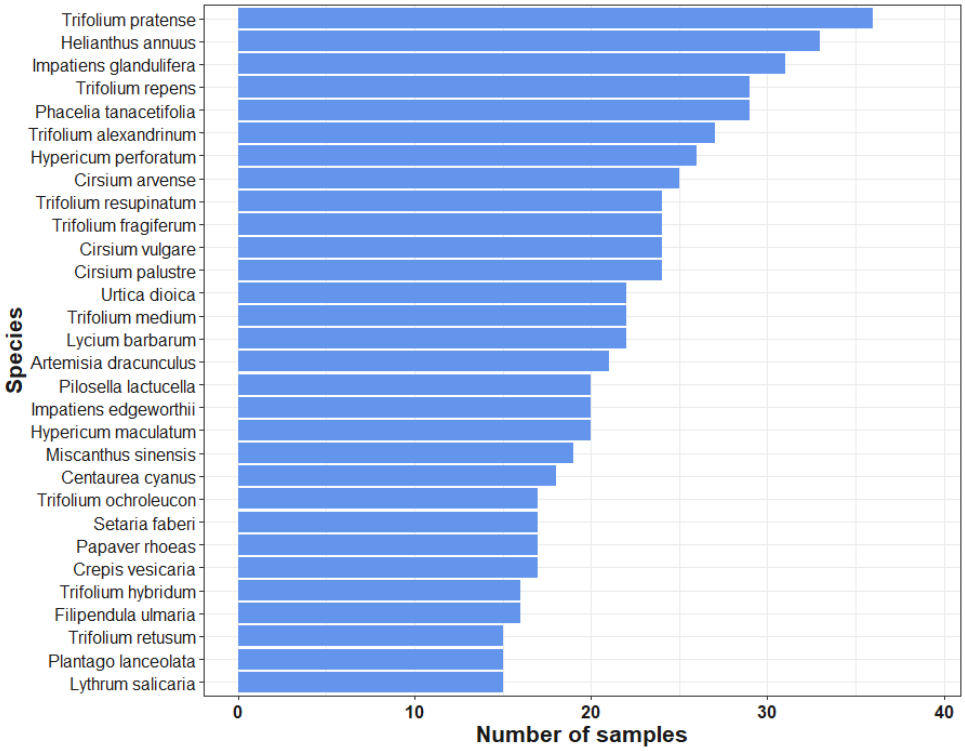


Figure 30: Presentation of the 30 plant species with highest sample presence. The graph is based on the combined results of the barcodes of *rbcL* and ITS.

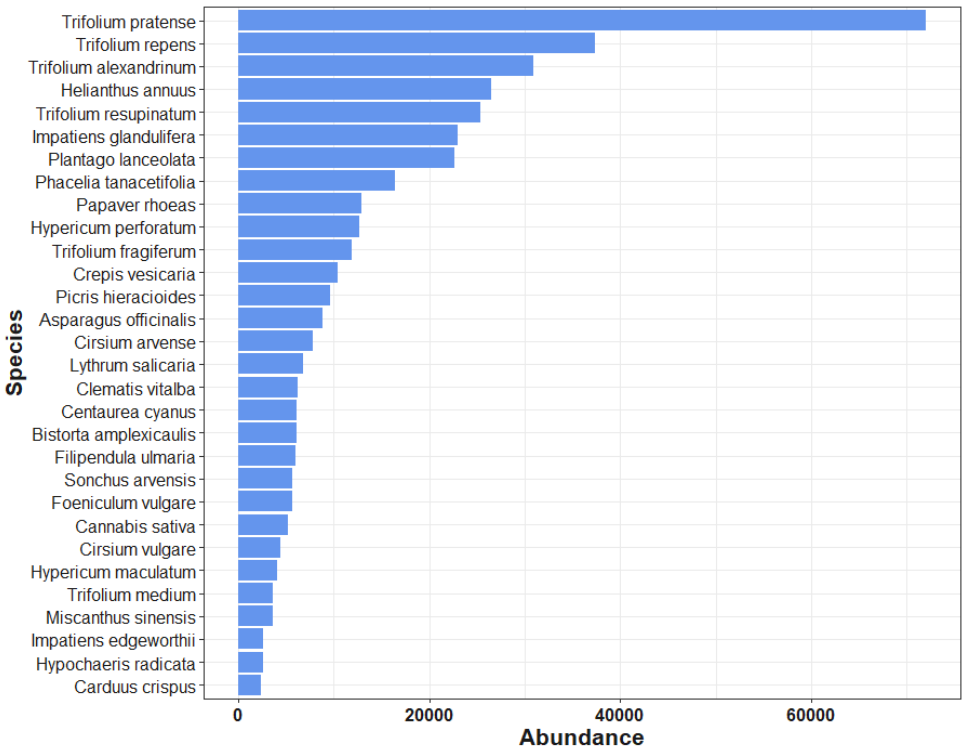


Figure 31: Representation of the 30 plant species with the highest read abundance. The subsets of the barcodes *rbcL* and ITS were thereby combined.

The most dominant species regarding sample and total abundance in this study belong to the genus *Trifolium spp.*, with *Trifolium pratense* and *Trifolium repens* as the most prevalent. Further species with high sample presence in addition to their total abundance are *Helianthus annuus*, *Impatiens glandulifera*, and *Phacelia tanacetifolia*. Interestingly, two grass species were detected in multiple samples, namely *Miscanthus sinensis* and *Setaria faberi*. *Miscanthus sinensis* is even counted among the thirty species with the highest read amount, as well as *Cannabis sativa*. Since all these species are anemophilous plants and typically not pollinated by insects, this outcome is rather surprising. One could argue, that the pollen was collected by coincidence, since anemophilous plants produce a large amount of pollen to gain higher chances on pollination by random wind distribution (O'Brien and Arathi, 2019; Timerman and Barrett, 2021). Then again, the two grass species were detected in multiple samples, which were also treated separately during PCR and sequencing steps which reduces the chance of cross-contaminations. Furthermore, *Miscanthus sinensis* and *Cannabis sativa* exhibited high read abundance. The findings support the hypothesis of Saunders (2018) which states that there are indeed interactions between wind-pollinated plant species and pollinating insects, though often neglected or declared as contaminations in scientific analyses. Other studies even describe the visitation and use of *Cannabis sativa* by honey bees in times of floral scarcity (Dalio, 2012; O'Brien and Arathi, 2019). However, collected by chance or not, it is an interesting question for investigation to which extent bees feed on pollen from non-entomophilous plants.

When interpreting the outcome regarding the landscape variables, it needs to be considered that all samples of this study only represent one point of time in the year. Further sampling throughout different months would be needed in order to portray a more complete picture of the whole season (Danner et al., 2017; Richardson et al., 2021; de Vere et al., 2017). A study conducted by Danner et al. (2017) indicates that the diversity of honey bee collected pollen is affected solely by seasonal changes and not by landscape variables. On the one hand, this resembles the results of this study

as Danner et al. (2017) showed that the species richness and Shannon index of the pollen samples were declining in August. On the other hand, a significant outcome regarding organic crop proportion was detected in this work which is an affecting landscape variable. The findings could indicate the need for further ecological measures during this time of year, especially since multiple anemophilous plants with poorer nutritional value were detected in the pollen samples, which might have been collected in dearth of other food resources (Dalio, 2012; O'Brien and Arathi, 2019; Saunders, 2018). Furthermore, it needs to be noted that even if the percentage of semi-natural habitats and the annual flowerfield area did not have a significant influence on honey bees in this case, it should not be disregarded that these factors are of vital importance as food source and living environment for wild, non-managed pollinating species (Bänsch et al., 2020a; Bertrand et al., 2019; Kremen et al., 2004).

5.8 Conclusion

All in all, it can be concluded that the use of the DNA barcodes ITS and *rbcL* proved suitable for the investigation of honey bee foraging behavior (Bell et al., 2017a; Richardson et al., 2021). A dual-barcoding approach is highly recommended, as well as the combination of plastid and nuclear encoded DNA loci, to gain higher species richness resolution (Bell et al., 2019; Kress et al., 2005; Richardson et al., 2021). The application of Nanopore techniques furthermore allowed the one-step, full-length sequencing of the two barcodes, which provided more genetic sequence information for taxonomic classification in the subsequent data analysis (Leidenfrost et al., 2020; Maestri et al., 2019).

The study has shown that the formulation of optimal ecological guidelines and methods for pollinator conservation is a complex task as it demands an evaluation of many different influencing factors. However, this circumstance should not be discouraging – on the contrary, it should be a motivation to continue efforts on decoding the complex interactions of plant-pollinator networks in order to identify suitable measures to ensure their protection and preservation in the future.

6 Outlook

Nanopore sequencing and DNA metabarcoding were the main methods applied in this study to analyze honey bee pollen samples. Nevertheless, these methods still hold room for improvement. On the one hand, nanopore sequencing generates reads with a relatively high error rate. High-accuracy basecalling and filtering can be an option, however, technology and software revision pose the greatest chance in order to overcome this drawback. On the other hand, DNA metabarcoding approaches are facing difficulties when it comes to the quantitative interpretation of the results, due to bias during DNA extraction, PCR, and sequencing success as well as barcode copy number. Different ways and methods are suggested to gain more certainty when it comes to quantifying the outcome of pollen metabarcoding approaches. One possibility would be the combination of genetic and optical approaches, including automated processes based on deep learning classification. Moreover, barcode-specific extraction using CRISPR/Cas systems instead of PCR amplification could be another potential method in the future. In addition to the improvement of the work procedures, a further data processing refinement of pollen metabarcoding databases could be considered, for example by integrating the flowering periods of the different plant species. In addition, analysing sequences of closely related plant species in order to identify their sequence uniqueness is another possible data improvement approach. For consecutive pollinator investigations it is furthermore recommended to analyze samples from different times in the year in order to gain more information about foraging shifts throughout the seasons. Thereby, it would be an interesting question of research, to which extend bees collect pollen from anemophilous plants and what are the main drivers for this behavior.

7 Summary

The loss of pollinating insects due to intensive land use, climate change, and pathogen distribution imposes far-reaching ecological and economical consequences. For this reason, fast and large-scale monitoring techniques have become more important than ever. This study portrays honey bee foraging behavior in different agro-environmental settings around the city of Göttingen (Germany), which were classified regarding organic crop proportion, percentage of semi-natural habitat, and the area of annual flowerfields. Pollen samples of these different landscapes were analyzed with DNA metabarcoding methods combined with the long-read nanopore sequencing technology. Within this report, a procedure for the adequate extraction of pollen DNA is described. In addition, two different types of DNA barcodes were examined: the spacer region ITS localized in the nucleus, and the plastid gene *rbcL*. On that account, a new reverse primer was designed during this study for the full-length amplification of (*rbcL*). It could be shown, that the results of the two DNA barcodes matched to a degree of nearly 70% regarding species read abundance and over 90% concerning genera read abundance. Furthermore, the dual-locus approach allowed to portray a more diverse picture regarding species richness, since over 70% of the species taxa in this study were detected by one of the barcodes exclusively. The analysis of the three different landscape variables exhibited a significant negative correlation of the proportion of organic crop area regarding species richness, Shannon diversity index, and species evenness. Furthermore, non-entomophilous plant species were detected in multiple samples, which could indicate alternative honey bee foraging during floral scarcity. However, the samples of this study represent just one point of time in the year and further examinations would be needed to display a more complete picture of honey bee foraging throughout the seasons.

8 Zusammenfassung

Der Verlust von Bestäuberinsekten durch intensive Landnutzung, Klimawandel und die Verbreitung von Pathogenen hat weitreichende ökologische und ökonomische Konsequenzen. Aus diesem Grund sind schnelle und umfangreiche Monitoring-Methoden bedeutsamer als je zuvor. Diese Studie untersucht das Sammelverhalten von Honigbienenvölkern in unterschiedlichen Agrarlandschaften rund um die Stadt Göttingen, welche hinsichtlich ihres Anteils an ökologischer Landwirtschaft, dem Prozentsatz an semi-naturalen Habitaten sowie der jährlichen Blühfläche klassifiziert wurden. Pollenproben aus den verschiedenen Landschaften wurden mithilfe von DNA-Metabarcoding und Nanopore-Sequenzierung genetisch analysiert. Dabei wurde ein geeignetes Verfahren für die DNA-Extraktion von Pollen erarbeitet. Weiterhin wurden zwei verschiedene DNA-Barcodes untersucht: der im Zellkern lokalisierte Spacer ITS und das plastidäre Gen *rbcL*. Für diesen Zweck wurde ein neuer *reverse* Primer konzipiert, um die Amplifikation der gesamten *rbcL* Genregion zu ermöglichen. Es konnte gezeigt werden, dass die Read-Abundanz der beiden Barcodes zu 70% auf Spezies- und bis zu 90% auf Gattungsebene übereinstimmte. Zudem erlaubte der duale Ansatz eine bessere Erfassung der Artenvielfalt, da über 70% der klassifizierten Spezies nur durch je einen der beiden Barcodes detektiert werden konnten. Die Analyse der drei Landschaftsvariablen ergab eine signifikant negative Korrelation hinsichtlich des Anteils ökologischer Landwirtschaft in Bezug auf Artenvielfalt, Shannon-Diversitätsindex und die Gleichverteilung der Arten. Des Weiteren wurden anemophile Pflanzenspezies in den Proben detektiert, was auf ein alternatives Sammelverhalten der Honigbienen aufgrund von Nahrungsknappheit hindeuten könnte. Nichtsdestotrotz muss bei der Interpretation dieses Ergebnisses beachtet werden, dass die Probennahme dieser Studie nur einen Zeitpunkt abbildet und mehr Untersuchungen notwendig wären, um das Sammelverhalten von Honigbienen über das Jahr besser darstellen zu können.

Acknowledgements

I hereby would like to thank everyone, who accompanied, guided, and motivated me during the development of my master's thesis.

First of all, I want to thank my supervisors Röbbe Wünschiers and Lisa Prudnikow for their advice, inspiring ideas, and support throughout the work process. Thank you, Röbbe Wünschiers, for inviting me to the *WunschAG*, for your time, your spirit of research, and for the Linux first aid. Thank you, Lisa Prudnikow, for all instructions, for your scientific enthusiasm, for travel organization, and your help whenever needed.

Furthermore, special thanks go to Catrin Westphal, for a spontaneous meeting that settled the topic of this thesis by introducing me to *ComBee*, and for the great time in Göttingen. At this point, I also need to thank the whole *ComBee*-team for their constructive feedback and inspiration, particularly Kathrin Czechofsky and Annika Haß, who helped me through the jungle of statistics.

I also would like to thank all members of the research group *WunschAG* and the faculty group *Biotechnology and Chemistry* at the *University of Applied Sciences Mittweida* with special regards to Nils Schön for bioinformatic advice and to Sandra Feik and René Kretschmer for laboratory organisation. Moreover, thank you all for such a great working atmosphere.

Thank you, also to my fellow students Lisa and Mareike for their help in words and deeds, in science and in life. Not to forget, I want to express my gratitude to my family who always encouraged me on my way. And, last but not least, I thank my partner Arvid who always kept my back, who supported me through all ups and downs – even if this meant that I was spending more time on this project than with him.

Thank you all!

Bibliography

Alaux, C., Dantec, C., Parrinello, H. and Le Conte, Y. (2011). Nutrigenomics in honey bees: Digital gene expression analysis of pollen's nutritive effects on healthy and varroa-parasitized bees. *BMC Genomics* 12: 1–13, doi:10.1186/1471-2164-12-496.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410, doi:10.32388/RHQ6VJ.

Álvarez, I. and Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–434, doi:10.1016/S1055-7903(03)00208-2.

Ariizumi, T. and Toriyama, K. (2011). Genetic regulation of sporopollenin synthesis and pollen exine development. *Annual Review of Plant Biology* 62: 437–460, doi:10.1146/annurev-arplant-042809-112312.

Arnot, D. E., Roper, C. and Bayoumi, R. A. (1993). Digital codes from hypervariable tandemly repeated DNA sequences in the *Plasmodium falciparum* circumsporozoite gene can genetically barcode isolates. *Molecular and Biochemical Parasitology* 61: 15–24.

Aslan, C. E., Zavaleta, E. S., Tershy, B. and Croll, D. (2013). Mutualism disruption threatens global plant biodiversity: A systematic review. *PloS ONE* 8: 1–11, doi:10.1371/journal.pone.0066993.

Baldwin, B. G. (1992). Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the compositae. *Molecular Phylogenetics and Evolution* 1: 3–16.

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S. and Donoghue, M. J. (1995). The ITS region of nuclear ribosomal DNA: A valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247–277, doi:10.2307/2399880.

Bänsch, S. (2019). Managing strawberry pollination with wild bees and honey bees: Facilitation or competition by mass-flowering resources? Dissertation, *Georg-August-Universität Göttingen*, Göttingen.

Bänsch, S., Tscharntke, T., Gabriel, D. and Westphal, C. (2020a). Crop pollination services: Complementary resource use by social vs solitary bees facing crops with contrasting flower supply. *Journal of Applied Ecology* 58: 476–485, doi:10.1111/1365-2664.13777.

Bänsch, S., Tscharntke, T., Wünschiers, R., Netter, L., Brenig, B., Gabriel, D. and Westphal, C. (2020b). Using ITS2 metabarcoding and microscopy to analyse shifts in pollen diets of honey bees and bumble bees along a mass-flowering crop gradient. *Molecular Ecology* 29: 5003–5018, doi:10.1111/mec.15675.

Bell, K. L., Burgess, K. S., Botsch, J. C., Dobbs, E. K., Read, T. D. and Brosi, B. J. (2019). Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. *Molecular Ecology* 28: 431–455, doi:10.1111/mec.14840.

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

Bell, K. L., Fowler, J., Burgess, K. S., Dobbs, E. K., Gruenewald, D., Lawley, B., Morozumi, C. and Brosi, B. J. (2017a). Applying pollen DNA metabarcoding to the study of plant-pollinator interactions. *Applications in Plant Sciences* 5: 1–10, doi:10.3732/apps.1600124.

Bell, K. L., Loeffler, V. M. and Brosi, B. J. (2017b). An *rbcL* reference library to aid in the identification of plant species mixtures by DNA metabarcoding. *Applications in Plant Sciences* 5: 1–7, doi:10.3732/apps.1600110.

Bertrand, C., Eckerter, P. W., Ammann, L., Entling, M. H., Gobet, E., Herzog, F., Mestre, L., Tinner, W. and Albrecht, M. (2019). Seasonal shifts and complementary use of pollen sources by two bees, a lacewing and a ladybeetle species in european agricultural landscapes. *Journal of Applied Ecology* 56: 2431–2442, doi:10.1111/1365-2664.13483.

BMEL (2019). Zukunftsstrategie ökologischer Landbau. *Bundesministerium für Ernährung und Landwirtschaft* https://www.bmel.de/SharedDocs/ Downloads/DE/Broschueren/ZukunftsstrategieOekologischerLandbau2019.html [accessed 01.11.2022].

Bolson, M., Smidt, E. d. C., Brotto, M. L. and Silva-Pereira, V. (2015). ITS and *trnH-psbA* as efficient DNA barcodes to identify threatened commercial woody angiosperms from southern Brazilian Atlantic rainforests. *PloS ONE* 10: 1–18, doi:10.1371/journal.pone.0143049.

Borg, M. and Twell, D. (2011). Pollen: Structure and development. *eLS, John Wiley & Sons, Ltd* 1: 1–11, doi:10.1002/9780470015902.a0002039.pub2.

Breeze, T. D., Gallai, N., Garibaldi, L. A. and Li, X. S. (2016). Economic measures of pollination services: Shortcomings and future directions. *Trends in Ecology & Evolution* 31: 927–939, doi:10.1016/j.tree.2016.09.002.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M. and Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* 9: 378–400, doi:10.32614/RJ-2017-066.

Cabelin, V. L. D. and Alejandro, G. J. D. (2016). Efficiency of *matK, rbcL, trnH-psbA, and trnL-F* (cpDNA) to molecularly authenticate philippine ethnomedicinal apocynaceae through DNA barcoding. *Pharmacognosy Magazine* 12: 384–388, doi:10.4103/0973-1296.185780.

CBOL Plant Working Group (2009). A DNA barcode for land plants. *PNAS* 106: 12794–12797, doi:10.1073/pnas.0905845106.

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M. and Palmer, T. M. (2015). Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances* 1: 1–5, doi:10.1126/sciadv.1400253.

Chase, M. W., Salamin, N., Wilkinson, M., Dunwell, J. M., Kesanakurthi, R. P., Haider, N., Haidar, N. and Savolainen, V. (2005). Land plants and DNA barcodes: Short-term and long-term goals. *Phil. Trans. R. Soc. B* 360: 1889–1895, doi:10.1098/rstb.2005.1720.

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. and Leon, C. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PloS ONE* 5: 1–8, doi:10.1371/journal.pone.0008613.

Coster, W. de, D'Hert, S., Schultz, D. T., Cruts, M. and van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics (Oxford, England)* 34: 2666–2669, doi:10.1093/bioinformatics/bty149.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A. and Swinton, J. (2019). xtable: Export tables to LaTeX or HTML. R package version 1.8-4, http://xtable.r-forge.r-project.org/ [accessed 01.11.2022].

Dalio, J. S. (2012). *Cannabis sativa*–An important subsistence pollen source for *Apis mellifera*. *IOSR Journal of Pharmacy and Biological Sciences* 1: 1–3, doi:10.9790/3008-0140103.

Danner, N., Keller, A., Härtel, S. and Steffan-Dewenter, I. (2017). Honey bee foraging ecology: Season but not landscape diversity shapes the amount and diversity of collected pollen. *PloS ONE* 12: 1–14, doi:10.1371/journal.pone.0183716.

de Vere, N., Jones, L. E., Gilmore, T., Moscrop, J., Lowe, A., Smith, D., Hegarty, M. J., Creer, S. and Ford, C. R. (2017). Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. *Scientific Reports* 7: 1–10, doi:10.1038/srep42838.

Deamer, D., Akeson, M. and Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology* 34: 518–524, doi:10.1038/nbt.3423.

Delahaye, C. and Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases. *PloS ONE* 16: 1–29, doi:10.1371/journal.pone.0257521.

di Pasquale, G., Salignon, M., Le Conte, Y., Belzunces, L. P., Decourtye, A., Kretzschmar, A., Suchail, S., Brunet, J.-L. and Alaux, C. (2013). Influence of pollen nutrition on honey bee health: Do pollen quality and diversity matter? *PloS ONE* 8: 1–13, doi:10.1371/journal.pone.0072016.

Dunker, S., Motivans, E., Rakosy, D., Boho, D., Mäder, P., Hornick, T. and Knight, T. M. (2021). Pollen analysis using multispectral imaging flow cytometry and deep learning. *New Phytologist* 229: 593–606, doi:10.1111/nph.16882.

Fišer Pečnikar, Ž. and Buzan, E. V. (2014). 20 years since the introduction of DNA barcoding: From theory to application. *Journal of Applied Genetics* 55: 43–52, doi:10.1007/s13353-013-0180-y.

Frias, B. E. D., Barbosa, C. D. and Lourenço, A. P. (2016). Pollen nutrition in honey bees *(Apis mellifera)*: Impact on adult health. *Apidologie* 47: 15–25, doi:10.1007/s13592-015-0373-y.

Galimberti, A., Mattia, F. de, Bruni, I., Scaccabarozzi, D., Sandionigi, A., Barbuto, M., Casiraghi, M. and Labra, M. (2014). A DNA barcoding approach to characterize pollen collected by honeybees. *PloS ONE* 9: 1–13, doi:10.1371/journal.pone.0109363.

Garve, E. (2004). Rote Liste und Florenliste der Farn- und Blütenpflanzen in Niedersachsen und Bremen. *Niedersächsisches Landesamt für Ökologie* 5th ed: 1–76.

Gotelli, N. J. and Colwell, R. K. (2001). Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4: 379–391, doi:10.1046/j.1461-0248.2001.00230.x.

Haaland, C., Naisbit, R. E. and Bersier, L.-F. (2011). Sown wildflower strips for insect conservation: A review. *Insect Conservation and Diversity* 4: 60–80, doi:10.1111/j.1752-4598.2010.00098.x.

Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörren, T., Goulson, D. and Kroon, H. de (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PloS ONE* 12: 1–21, doi:10.1371/journal.pone.0185809.

Hansen, A. (2004). Bioinformatik: Ein Leitfaden für Naturwissenschaftler. *Springer Basel AG* 2nd ed: 113–116, doi:10.1007/978-3-0348-7855-5.

Hartig, F. (2022). DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models. R package version 0.4.6, https://cran.r-project.org/web/packages/DHARMa/vignettes/DHARMa.html [accessed 01.11.2022].

Hawkins, J., de Vere, N., Griffith, A., Ford, C. R., Allainguillaume, J., Hegarty, M. J., Baillie, L. and Adams-Groom, B. (2015). Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. *PloS ONE* 10: 1–20, doi:10.1371/journal.pone.0134735.

Hebert, P. D. N., Cywinska, A., Ball, S. L. and deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* 270: 313–321, doi:10.1098/rspb.2002.2218.

Hebert, P. D. N. and Gregory, T. R. (2005). The promise of DNA barcoding for taxonomy. *Systematic Biology* 54: 852–859, doi:10.1080/10635150500354886.

Hilu, K. W. and Liang, H. (1997). The *matK* gene: Sequence variation and application in plant systematics. *American Journal of Botany* 84: 830–839.

HLNUG (2019). Rote Liste der Farn- und Samenpflanzen Hessens. *Hessisches Landesamt für Naturschutz, Umwelt und Geologie* 5th ed: 1–274.

Huang, Z. (2012). Pollen nutrition affects honey bee stress resistance. *Terrestrial Arthropod Reviews* 5: 175–189, doi:10.1163/187498312X639568.

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5: 299–314.

Jaakola, L., Pirttilä, A. M., Vuosku, J. and Hohtola, A. (2004). Method based on electrophoresis and gel extraction for obtaining genomic DNA-free cDNA without DNase treatment. *Biotechniques* 37: 744–748, doi:10.2144/04375BM06.

Jain, M., Fiddes, I. T., Miga, K. H., Olsen, H. E., Paten, B. and Akeson, M. (2015). Improved data analysis for the MinION nanopore sequencer. *Nature Methods* 12: 351–356, doi:10.1038/nmeth.3290.

Jiao, W.-B. and Schneeberger, K. (2017). The impact of third generation genomic technologies on plant genome assembly. *Current Opinion in Plant Biology* 36: 64–70, doi:10.1016/j.pbi.2017.02.002.

Kasianowicz, J. J., Brandin, E., Branton, D. and Deamer, D. W. (1996). Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. USA* 93: 13770–13773, doi:10.1073/pnas.93.24.13770.

Khalifa, S. A. M., Elshafiey, E. H., Shetaia, A. A., El-Wahed, A. A. A., Algethami, A. F., Musharraf, S. G., AlAjmi, M. F., Zhao, C., Masry, S. H. D., Abdel-Daim, M. M., Halabi, M. F., Kai, G., Al Naggar, Y., Bishr, M., Diab, M. A. M. and El-Seedi, H. R. (2021). Overview of bee pollination and its economic value for crop production. *Insects* 12: 1–23, doi:10.3390/insects12080688.

Klein, A.-M., Vaissière, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C. and Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. *Proc. R. Soc. B* 274: 303–313, doi:10.1098/rspb.2006.3721.

Kluser, S. and Peduzzi, P. (2007). Global pollinator decline: A literature review. *UNEP/GRID Europe* 1: 1–12, doi:10.1371/journal.pbio.0050168.

Koetsier, G. and Cantor, E. (2019). A practical guide to analyzing nucleic acid concentration and purity with microvolume spectrophotometers. https://www.science.smith.edu/cmbs/wp-content/uploads/sites/36/2020/01/A-Practical-Guide-to-Analyzing-Nucleic-Acid-Concentration-and-Purity-with-Microvolume-Spectrophotometers.pdf [accessed 01.11.2022].

Kolter, A. and Gemeinholzer, B. (2021). Plant DNA barcoding necessitates marker-specific efforts to establish more comprehensive reference databases. *Genome* 64: 265–298, doi:10.1139/gen-2019-0198.

Korsch, H., Westhus, W., Horn, K. and Jansen, W. (2010). Rote Liste der Farn- und Blütenpflanzen (Pteridophyta et Spermatophyta) Thüringens. *Thüringer Landesamt für Umwelt, Bergbau und Naturschutz* 5th ed: 1–26.

Kraaijeveld, K., Weger, L. A. de, Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P. S. and Dunnen, J. T. den (2015). Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Molecular Ecology Resources* 15: 8–16, doi:10.1111/1755-0998.12288.

Krehenwinkel, H., Pomerantz, A. and Prost, S. (2019). Genetic biomonitoring and biodiversity assessment using portable sequencing technologies: Current uses and future directions. *Genes* 10: 1–16, doi:10.3390/genes10110858.

Kremen, C., Williams, N. M., Bugg, R. L., Fay, J. P. and Thorp, R. W. (2004). The area requirements of an ecosystem service: Crop pollination by native bee communities in California. *Ecology Letters* 7: 1109–1119, doi:10.1111/j.1461-0248.2004.00662.x.

Kress, W. J. and Erickson, D. L. (2007). A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PloS ONE* 2: 1–10, doi:10.1371/journal.pone.0000508.

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. and Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *PNAS* 102: 8369–8374.

Laha, R. C., Mandal, S. de, Ralte, L., Ralte, L., Kumar, N. S., Gurusubramanian, G., Satishkumar, R., Mugasimangalam, R. and Kuravadi, N. A. (2017). Meta-barcoding in combination with palynological inference is a potent diagnostic marker for honey floral composition. *AMB Express* 7: 1–8, doi:10.1186/s13568-017-0429-7.

Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G. and Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology* 28: 420–430, doi:10.1111/mec.14920.

Lee, J. Y., Kong, M., Oh, J., Lim, J., Chung, S. H., Kim, J.-M., Kim, J.-S., Kim, K.-H., Yoo, J.-C. and Kwak, W. (2021). Comparative evaluation of nanopore polishing tools for microbial genome assembly and polishing strategies for downstream analysis. *Scientific Reports* 11: 1–11, doi:10.1038/s41598-021-00178-w.

Leidenfrost, R. M., Bänsch, S., Prudnikow, L., Brenig, B., Westphal, C. and Wünschiers, R. (2020). Analyzing the dietary diary of bumble bee. *Frontiers in Plant Science* 11: 1–9, doi:10.3389/fpls.2020.00287.

Leontidou, K., Vernesi, C., Groeve, J. de, Cristofolini, F., Vokou, D. and Cristofori, A. (2017). Taxonomic identification of airborne pollen from complex environmental samples by DNA metabarcoding: A methodological study for optimizing protocols. *Aerobiologia* 1: 1–23, doi:10.1101/099481.

Li, C., Chng, K. R., Boey, E. J. H., Ng, A. H. Q., Wilm, A. and Nagarajan, N. (2016). INC-Seq: Accurate single molecule reads using nanopore sequencing. *GigaScience* 5: 1–11, doi:10.1186/s13742-016-0140-7.

Li, F.-S., Phyo, P., Jacobowitz, J., Hong, M. and Weng, J.-K. (2019). The molecular structure of plant sporopollenin. *Nature Plants* 5: 41–46, doi:10.1038/s41477-018-0330-7.

Loera-Sánchez, M., Studer, B. and Kölliker, R. (2020). DNA barcode *trnH-psbA* is a promising candidate for efficient identification of forage legumes and grasses. *BMC Research Notes* 13: 1–6, doi:10.1186/s13104-020-4897-5.

Lozier, J. D. and Zayed, A. (2017). Bee conservation in the age of genomics. *Conservation Genetics* 18: 713–729, doi:10.1007/s10592-016-0893-7.

Macherey-Nagel (2020). User manual: Genomic DNA from food. https://www.mn-net.com/media/pdf/55/34/a2/Instruction-NucleoSpin-Food.pdf [accessed 01.11.2022].

Maestri, S., Cosentino, E., Paterno, M., Freitag, H., Garces, J. M., Marcolungo, L., Alfano, M., Njunjić, I., Schilthuizen, M., Slik, F., Menegon, M., Rossato, M. and Delledonne, M. (2019). A rapid and accurate MinION-based workflow for tracking species biodiversity in the field. *Genes* 10: 1–11, doi:10.3390/genes10060468.

Maloukh, L., Kumarappan, A., Jarrar, M., Salehi, J., El-Wakil, H. and Rajya Lakshmi, T. V. (2017). Discriminatory power of *rbcL* barcode locus for authentication of some of United Arab Emirates (UAE) native plants. *3 Biotech* 7: 1–7, doi:10.1007/s13205-017-0746-1.

Martinez-Seidel, F., Beine-Golovchuk, O., Hsieh, Y.-C. and Kopka, J. (2020). Systematic review of plant ribosome heterogeneity and specialization. *Frontiers in Plant Science* 11: 1–23, doi:10.3389/fpls.2020.00948.

Matlock, B. (2015). Assessment of nucleic acid purity. https://assets.thermofisher.com/TFS-Assets/CAD/Product-Bulletins/TN52646-E-0215M-NucleicAcid.pdf [accessed 01.11.2022].

Matsushima, R., Tang, L. Y., Zhang, L., Yamada, H., Twell, D. and Sakamoto, W. (2011). A conserved, Mg^{2+}-dependent exonuclease degrades organelle DNA during Arabidopsis pollen development. *The Plant Cell* 23: 1608–1624, doi:10.1105/tpc.111.084012.

Melgar, M., Trigo, M. M., Recio, M., Docampo, S., García-Sánchez, J. and Cabezudo, B. (2012). Atmospheric pollen dynamics in Münster, north-western Germany: A three-year study (2004–2006). *Aerobiologia* 28: 423–434, doi:10.1007/s10453-012-9246-2.

Merget, B., Koetschan, C., Hackl, T., Förster, F., Dandekar, T., Müller, T., Schultz, J. and Wolf, M. (2012). The ITS2 database. *JoVE* 61: 1–5, doi:10.3791/3806.

Milla, L., Sniderman, K., Lines, R., Mousavi-Derazmahalleh, M. and Encinas-Viso, F. (2021). Pollen DNA metabarcoding identifies regional provenance and high plant diversity in Australian honey. *Ecology and Evolution* 11: 8683–8698, doi:10.1002/ece3.7679.

Nazarevich, V. (2015). The sixth species extinction event by humans. *Earth Common Journal* 5: 61–72, doi:10.31542/j.ecj.261.

Newmaster, S. G., Fazekas, A. J. and Ragupathy, S. (2006). DNA barcoding in land plants: Evaluation of rbcL in a multigene tiered approach. *Canadian Journal of Botany* 84: 335–341, doi:10.1139/b06-047.

Nybo, K. (2013). Primer design. *Biotechniques* 54: 249–250, doi:10.2144/000114025.

O'Brien, C. and Arathi, H. S. (2019). Bee diversity and abundance on flowers of industrial hemp *(Cannabis sativa L.)*. *Biomass and Bioenergy* 122: 331–335, doi:10.1016/j.biombioe.2019.01.015.

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H. B. A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M. O., Lahti, L., McGlinn, D., Ouellette, M.-H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C. J. and Weedon, J. (2022). vegan: Community Ecology Package. R package version 2.6-2, https://CRAN.R-project.org/package=vegan [accessed 01.11.2022].

Ollerton, J., Winfree, R. and Tarrant, S. (2011). How many flowering plants are pollinated by animals? *Oikos* 120: 321–326, doi:10.1111/j.1600-0706.2010.18644.x.

Olsson, O., Karlsson, M., Persson, A. S., Smith, H. G., Varadarajan, V., Yourstone, J. and Stjernman, M. (2021). Efficient, automated and robust pollen analysis using deep learning. *Methods in Ecology and Evolution* 12: 850–862, doi:10.1111/2041-210X.13575.

Oxford Nanopore Technologies (2021). Ligation sequencing amplicons – Native barcoding (SQK-LSK109 with EXP-NBD104 and EXP-NBD114). https://community.nanoporetech.com/docs/prepare/library_prep_protocols/native-barcoding-amplicons/v/nba_9093_v109_revn_12nov2019 [accessed 01.11.2022].

Pamminger, T., Becker, R., Himmelreich, S., Schneider, C. W. and Bergtold, M. (2019a). The nectar report: quantitative review of nectar sugar concentrations offered by bee visited flowers in agricultural and non-agricultural landscapes. *PeerJ* : 1–15doi:10.7717/peerj.6329.

Pamminger, T., Becker, R., Himmelreich, S., Schneider, C. W. and Bergtold, M. (2019b). Pollen report: quantitative review of pollen crude protein concentrations offered by bee pollinated flowers in agricultural and non-agricultural landscapes. *PeerJ* : 1–13doi:10.7717/peerj.7394.

Pang, X., Liu, C., Shi, L., Liu, R., Liang, D., Li, H., Cherny, S. S. and Chen, S. (2012). Utility of the *trnH-psbA* intergenic spacer region and its combinations as plant DNA barcodes: A meta-analysis. *PloS ONE* 7: 1–9, doi:10.1371/journal.pone.0048833.

Patel, M. and Berry, J. O. (2008). Rubisco gene expression in C4 plants. *Journal of Experimental Botany* 59: 1625–1634, doi:10.1093/jxb/erm368.

Peel, N., Dicks, L. V., Clark, M. D., Heavens, D., Percival-Alwyn, L., Cooper, C., Davies, R. G., Leggett, R. M. and Yu, D. W. (2019). Semi–quantitative characterisation of mixed pollen samples using MinION sequencing and Reverse Metagenomics (RevMet). *Methods in Ecology and Evolution* 10: 1690–1701, doi:10.1111/2041-210X.13265.

Petersen, G., Johansen, B. and Seberg, O. (1996). PCR and sequencing from a single pollen grain. *Plant Molecular Biology* 31: 189–191, doi:10.1007/BF00020620.

Pfenninger, M., Nowak, C., Kley, C., Steinke, D. and Streit, B. (2007). Utility of DNA taxonomy and barcoding for the inference of larval community structure in morphologically cryptic Chironomus (Diptera) species. *Molecular Ecology* 16: 1957–1968, doi:10.1111/j.1365-294X.2006.03136.x.

Pornon, A., Escaravage, N., Burrus, M., Holota, H., Khimoun, A., Mariette, J., Pellizzari, C., Iribar, A., Etienne, R., Taberlet, P., Vidal, M., Winterton, P., Zinger, L. and Andalo, C. (2016). Using metabarcoding to reveal and quantify plant-pollinator interactions. *Scientific Reports* 6: 1–12, doi:10.1038/srep27282.

Pottier, M., Gilis, D. and Boutry, M. (2018). The hidden face of Rubisco. *Trends in Plant Science* 23: 382–392, doi:10.1016/j.tplants.2018.02.006.

Potts, S. G., Imperatriz-Fonseca, V., Ngo, H. T., Aizen, M. A., Biesmeijer, J. C., Breeze, T. D., Dicks, L. V., Garibaldi, L. A., Hill, R., Settele, J. and Vanbergen, A. J. (2016). Safeguarding pollinators and their values to human well-being. *Nature* 540: 220–229, doi:10.1038/nature20588.

Prudnikow, L. C. (2021). Development of a genetic biomonitoring test for investigating plant-pollinator interactions. Master thesis, *Hochschule Mittweida*, Mittweida.

Qiagen (2016). Quick-start protocol: DNeasy® Plant Mini Kit. https://www.qiagen.com/us/resources/resourcedetail?id=6b9bcd96-d7d4-48a1-9838-58dbfb0e57d0lang=en [accessed 01.11.2022].

Qiagen (2020). DNeasy plant handbook. https://www.qiagen.com/us/resources/resourcedetail?id=f6455f80-dc4f-4ff2-b2de-ae7a3e6c91e0lang=en [accessed 01.11.2022].

Rang, F. J., Kloosterman, W. P. and Ridder, J. de (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biology* 19: 1–11, doi:10.1186/s13059-018-1462-9.

Reuter, J. A., Spacek, D. V. and Snyder, M. P. (2015). High-throughput sequencing technologies. *Molecular Cell* 58: 586–597, doi:10.1016/j.molcel.2015.05.004.

Richardson, R. T., Curtis, H. R., Matcham, E. G., Lin, C.-H., Suresh, S., Sponsler, D. B., Hearon, L. E. and Johnson, R. M. (2019). Quantitative multi-locus metabarcoding and waggle dance interpretation reveal honey bee spring foraging patterns in Midwest agroecosystems. *Molecular Ecology* 28: 686–697, doi:10.1111/mec.14975.

Richardson, R. T., Eaton, T. D., Lin, C.-H., Cherry, G., Johnson, R. M. and Sponsler, D. B. (2021). Application of plant metabarcoding to identify diverse honeybee pollen forage along an urban-agricultural gradient. *Molecular Ecology* 30: 310–323, doi:10.1111/mec.15704.

Richardson, R. T., Lin, C.-H., Quijia, J. O., Riusech, N. S., Goodell, K. and Johnson, R. M. (2015). Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. *Applications in Plant Sciences* 3: 1–9, doi:10.3732/apps.1500043.

Rollin, O., Bretagnolle, V., Decourtye, A., Aptel, J., Michel, N., Vaissière, B. E. and Henry, M. (2013). Differences of floral resource use between honey bees and wild bees in an intensive farming system. *Agriculture, Ecosystems & Environment* 179: 78–86, doi:10.1016/j.agee.2013.07.007.

Ruedenauer, F. A., Biewer, N. W., Nebauer, C. A., Scheiner, M., Spaethe, J. and Leonhardt, S. D. (2021). Honey bees can taste amino and fatty acids in pollen, but not sterols. *Frontiers in Ecology and Evolution* 9: 1–7, doi:10.3389/fevo.2021.684175.

Ruppert, K. M., Kline, R. J. and Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation* 17: 1–29, doi:10.1016/j.gecco.2019.e00547.

Sakamoto, W. and Takami, T. (2018). Chloroplast dna dynamics: Copy number, quality control and degradation. *Plant & Cell Physiology* 59: 1120–1127, doi:10.1093/pcp/pcy084.

Santamaria, M., Fosso, B., Consiglio, A., Caro, G. de, Grillo, G., Licciulli, F., Liuni, S., Marzano, M., Alonso-Alemany, D., Valiente, G. and Pesole, G. (2012). Reference databases for taxonomic assignment in metagenomics. *Briefings in Bioinformatics* 13: 682–695, doi:10.1093/bib/bbs036.

Saunders, M. E. (2018). Insect pollinators collect pollen from wind-pollinated plants: Implications for pollination ecology and sustainable agriculture. *Insect Conservation and Diversity* 11: 13–31, doi:10.1111/icad.12243.

Savolainen, V. and Chase, M. W. (2003). A decade of progress in plant molecular phylogenetics. *Trends in Genetics* 19: 717–724, doi:10.1016/j.tig.2003.10.003.

Scheper, J., Bommarco, R., Holzschuh, A., Potts, S. G., Riedinger, V., Roberts, S. P. M., Rundlöf, M., Smith, H. G., Steffan-Dewenter, I., Wickens, J. B., Wickens, V. J. and Kleijn, D. (2015). Local and landscape-level floral resources explain effects of wildflower strips on wild bees across four European countries. *Journal of Applied Ecology* 52: 1165–1175, doi:10.1111/1365-2664.12479.

Sevillano, V., Holt, K. and Aznarte, J. L. (2020). Precise automatic classification of 46 different pollen types with convolutional neural networks. *PloS ONE* 15: 1–15, doi:10.1371/journal.pone.0229751.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. and Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7: 1–6, doi:10.1038/msb.2011.75.

Simel, E. J., Saidak, L. R. and Tuskan, G. A. (1996). Method of extracting genomic DNA from non-germinated gymnosperm and angiosperm pollen. *Biotechniques* 22: 390–394.

Slowikowski, K. (2021). ggrepel: Automatically position non-overlapping text labels with 'ggplot2'. R package version 0.9.1, https://CRAN.R-project.org/package=ggrepel [accessed 01.11.2022].

Smith, M. A., Woodley, N. E., Janzen, D. H., Hallwachs, W. and Hebert, P. D. (2006). DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *PNAS* 103: 3657–3662, doi:0.1073/pnas.0511318103.

Steckel, J., Westphal, C., Peters, M. K., Bellach, M., Rothenwoehrer, C., Erasmi, S., Scherber, C., Tscharntke, T. and Steffan-Dewenter, I. (2014). Landscape composition and configuration differently affect trap-nesting bees, wasps and their antagonists. *Biological Conservation* 172: 56–64, doi:10.1016/j.biocon.2014.02.015.

Swenson, S. J. and Gemeinholzer, B. (2021). Testing the effect of pollen exine rupture on metabarcoding with Illumina sequencing. *PloS ONE* 16: 1–16, doi:10.1371/journal.pone.0245611.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C. and Willerslev, E. (2007). Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* 35: 1–8, doi:10.1093/nar/gkl938.

Tatham, S. (2021). PuTTY: A free Telnet/SSH client. https://github.com/janinge/putty [accessed 10.10.2022].

Thermo Fisher Scientific Inc. (2016). Qubit dsDNA assay specificity in the presence of single-stranded DNA. Https://assets.thermofisher.com/TFS-Assets/LSG/Application-Notes/qubit-dsdna-assay-specificity-app-note.pdf [accessed 01.11.2022].

Thermo Fisher Scientific Inc. (2021). Qubit fluorometers and assays. https://assets.thermofisher.com/TFS-Assets/BID/brochures/qubit-fluorometers-assays-brochure.pdf [accessed 01.11.2022].

Thomsen, P. F. and Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation* 183: 4–18, doi:10.1016/j.biocon.2014.11.019.

Timerman, D. and Barrett, S. C. H. (2021). The biomechanics of pollen release: New perspectives on the evolution of wind pollination in angiosperms. *Biological Reviews* 96: 2146–2163, doi:10.1111/brv.12745.

Urbanowicz, C., Muñiz, P. A. and McArt, S. H. (2020). Honey bees and wild pollinators differ in their preference for and use of introduced floral resources. *Ecology and Evolution* 10: 6741–6751, doi:10.1002/ece3.6417.

Valentini, A., Pompanon, F. and Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology & Evolution* 24: 110–117, doi:10.1016/j.tree.2008.09.011.

Valière, N., Fumagalli, L., Gielly, L., Miquel, C., Lequette, B., Poulle, M.-L., Weber, J.-M., Arlettaz, R. and Taberlet, P. (2003). Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. *Animal Conservation* 6: 83–92, doi:10.1017/S1367943003003111.

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. and Thermes, C. (2018). The third revolution in sequencing technology. *Trends in Genetics* 34: 666–681, doi:10.1016/j.tig.2018.05.008.

Wang, X.-C., Liu, C., Huang, L., Bengtsson-Palme, J., Chen, H., Zhang, J.-H., Cai, D. and Li, J.-Q. (2015a). ITS1: A DNA barcode better than ITS2 in eukaryotes? *Molecular Ecology Resources* 15: 573–586, doi:10.1111/1755-0998.12325.

Wang, Y., Yang, Q. and Wang, Z. (2015b). The evolution of nanopore sequencing. *Frontiers in Genetics* 5: 1–20, doi:10.3389/fgene.2014.00449.

Wang, Y., Zhao, Y., Bollas, A., Wang, Y. and Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology* 39: 1348–1365, doi:10.1038/s41587-021-01108-x.

Werchan, M., Werchan, B. and Bergmann, K.-C. (2018). German pollen calendar 4.0 – Update based on 2011–2016 pollen data. *Allergo Journal International* 27: 69–71, doi:10.1007/s40629-018-0055-1.

Westphal, C., Hass, A. and Paxton, R. (2021). ComBee: Kombinierte Agrarumweltmaßnahmen, Bienendiversität und Gesundheitszustand von Wild- und Honigbienen. https://www.uni-goettingen.de/de/646422.html [accessed 01.11.2022].

White, T. J., Bruns, T., Lee, S. and Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols* 1: 1–8.

Whitlock, B. A., Hale, A. M. and Groff, P. A. (2010). Intraspecific inversions pose a challenge for the *trnH-psbA* plant DNA barcode. *PloS ONE* 5: 1–7, doi:10.1371/journal.pone.0011533.

Wick, R. R. (2017). Porechop. https://github.com/rrwick/porechop [accessed 10.10.2022].

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. *Springer-Verlag New York* 2nd ed, doi:10.1007/978-3-319-24277-4.

Wickham, H. (2022a). forcats: Tools for working with categorical variables (factors). R package version 0.5.2, https://CRAN.R-project.org/package=forcats [accessed 01.11.2022].

Wickham, H. (2022b). stringr: Simple, consistent wrappers for common string operations. R package version 1.4.1, https://CRAN.R-project.org/package=stringr [accessed 01.11.2022].

Wickham, H., François, R., Henry, L. and Müller, K. (2022). dplyr: A grammar of data manipulation. R package version 1.0.10, https://CRAN.R-project.org/package=dplyr [accessed 01.11.2022].

Wickham, H. and Girlich, M. (2022). tidyr: Tidy messy data. R package version 1.2.0, https://CRAN.R-project.org/package=tidyr [accessed 01.11.2022].

Wickham, H. and Seidel, D. (2022). scales: Scale functions for visualization. R package version 1.2.0, https://CRAN.R-project.org/package=scales [accessed 01.11.2022].

Williams, M.-A., O'Grady, J., Ball, B., Carlsson, J., Eyto, E. de, McGinnity, P., Jennings, E., Regan, F. and Parle-McDermott, A. (2019). The application of CRISPR-Cas for single species identification from environmental DNA. *Molecular Ecology Resources* 19: 1106–1114, doi:10.1111/1755-0998.13045.

Wünschiers, R. (2022). qfilter. https://github.com/awkologist/qfilter [accessed 10.10.2022].

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13: 134, doi:10.1186/1471-2105-13-134.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht. Fotos sowie Abbildungen, die nicht gekennzeichnet wurden, wurden von mir erstellt. Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Mittweida, 03.11.2022

Birgit Pannicke

Supplemental Information

Table S1 Coordinates of the sampling locations

Location ID	Coordinates
Goe1425	51.4570895,10.1550478
Goe189	51.721122799999996,10.2286852
Goe235	51.6926504,10.1930846
Goe288	51.6775506,10.2842807
Goe47	51.7767010,10.2194920
Goe595	51.6071467,9.8639256
Gos1	51.9379130,10.4700470
Gos2	51.977413,10.255802
Nor1	51.825189,10.000652
Nor1070	51.729802,9.808416
Nor1145	51.7987985,10.0952517
Nor264	51.685006,9.965145
Nor39	51.7012912,9.7752581
Nor508	51.835786,9.7443048
Nor918	51.89068,9.88693
Wm1249	51.3209764,9.8079358
Wm1316	51.257383499999996,10.0017318
Wm597	51.1741767,10.1073515
Wm630	51.162446,10.042364

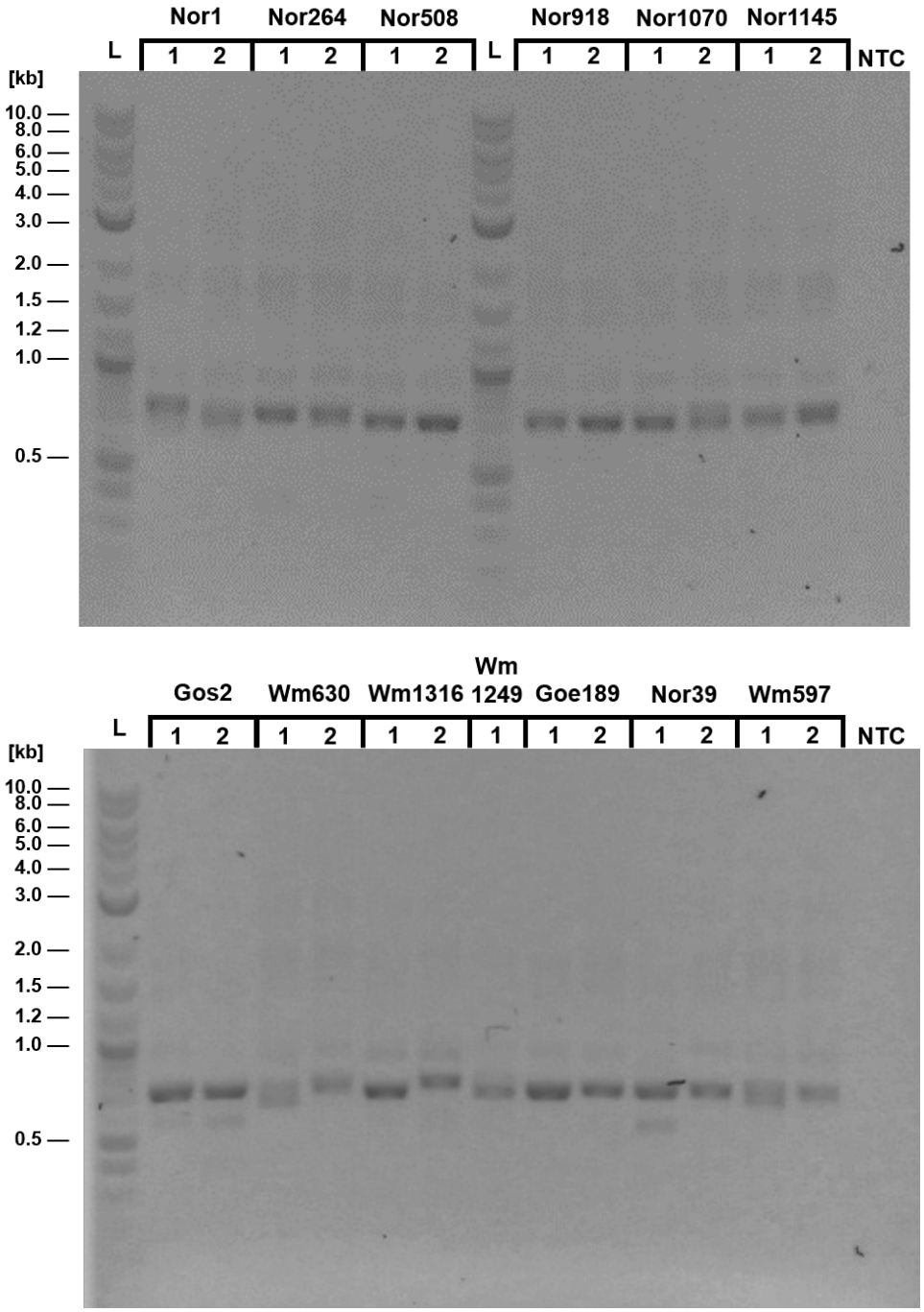


Figure S1: Electrophoresis gels of the purified ITS amplicons; L = DNA Ladder, NTC = no template control; Agarose concentration 1%

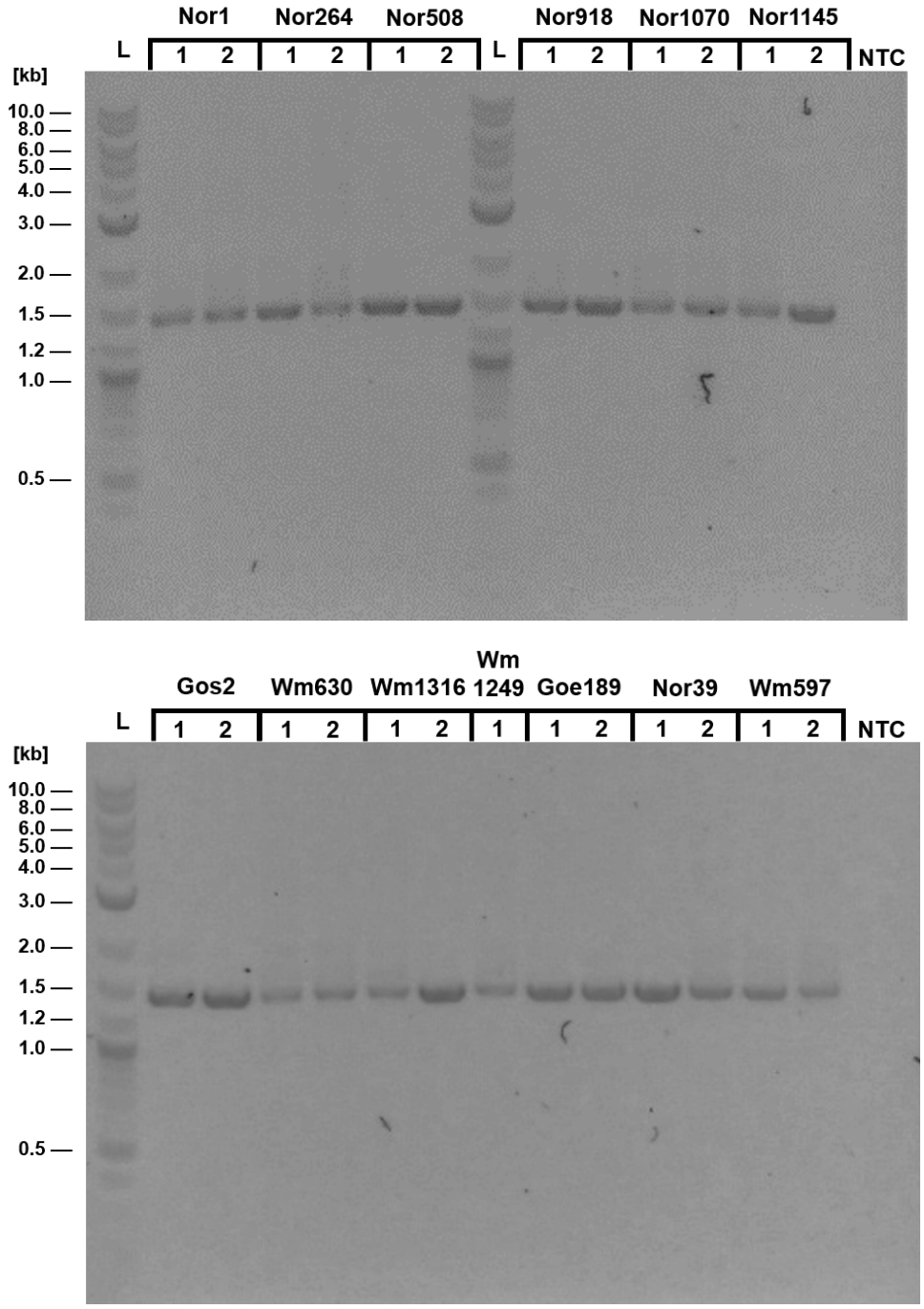


Figure S2: Electrophoresis gels of the purified *rbcL* amplicons; L = DNA Ladder, NTC = no template control; Agarose concentration 1%

Table S2 Sequencing run details

Run No.	Run ID	Reads [k]	Mean Quality Score	Samples (1 & 2)
1	ITS_Goe_10	573.92	10.4	(ITS) Goe1425, Goe235, Goe288, Goe47, Goe595
2	ITS_Nor_12	1100.00	9.7	(ITS) Nor1, Nor1070, Nor1145, Nor264, Nor508, Nor918
3	ITS_GosWm_9	616.97	10.6	(ITS) Gos1, Gos2, Wm1249_1, Wm1316, Wm630
4	ITSrbcL_Goe NorWm_12	542.91	10.6	(ITS) Goe189, Nor39, Wm597 (*rbcL*) Goe189, Nor39, Wm597
5	rbcL_Goe_10	697.48	9.3	(*rbcL*) Goe1425, Goe235, Goe288, Goe47, Goe595
6	rbcL_Nor_12	212.28	10.1	(*rbcL*) Nor1, Nor1070, Nor1145, Nor264, Nor508, Nor918
7	rbcL_GosWm_9	607.79	9.1	(*rbcL*) Gos1, Gos2, Wm1249_1, Wm1316, Wm630

Table S3 Generalized linear mixed model species richness

Predictors	Incidence Rate Ratios	CI	p-Value	
(Intercept)	46.29	37.55 – 57.07	<0.001	***
Organic crop [%]	0.99	0.98 – 1.00	0.076	.
Semi-natural habitat [%]	1.00	0.99 – 1.02	0.577	
Annual flowerfield [ha]	0.99	0.96 – 1.02	0.627	
N_{Land_ID} 37				
Observations 37				

Table S4 Generalized linear mixed model species read abundance

Predictors	Incidence Rate Ratios	CI	p-Value	
(Intercept)	7780.05	5504.81 – 10995.70	<0.001	***
Organic crop [%]	1.02	1.00 – 1.04	0.030	*
Semi-natural habitat [%]	1.00	0.98 – 1.02	0.990	
Annual flowerfield [ha]	1.02	0.97 – 1.07	0.432	
N_{Land_ID} 37				
Observations 37				

Table S5 Generalized linear mixed model Shannon index

Predictors	Estimates	CI	p-Value	
(Intercept)	1.80	1.48 – 2.12	<0.001	***
Organic crop [%]	-0.03	-0.05 – -0.01	<0.001	***
Semi-natural habitat [%]	0.01	-0.02 – 0.03	0.625	
Annual flowerfield [ha]	0.02	-0.02 – 0.07	0.321	
N_{Land_ID} 37				
Observations 37				

Table S6 Generalized linear mixed model species evenness

Predictors	Estimates	CI	p-Value	
(Intercept)	0.47	0.39 – 0.54	<0.001	***
Organic crop [%]	-0.01	-0.01 – -0.00	0.001	***
Semi-natural habitat [%]	0.00	-0.00 – 0.01	0.685	
Annual flowerfield [ha]	0.01	-0.00 – 0.02	0.176	
N_{Land_ID} 37				
Observations 37				

Commands

Plantlist specific filtering:

```
grep --no-group-separator -F -w -A1 -f PLANTLIST.txt INPUT.fasta
    > OUTPUT.fasta
```

Database creation:

```
./ncbi-blast-2.13.0+/bin/makeblastdb -in INPUT.fasta -out OUTPUT -parse_seqids
    -dbtype nucl
```

Basecalling + trimming:

```
guppy_basecaller -i INPUTDIRECTORY -r -s OUTPUTDIRECTORY
    -c dna_r9.4.1_450bps_hac.cfg --disable_qscore_filtering
    --device "cuda:0,1" --do_read_splitting --barcode_kits "EXP-NBD104"
    --trim_barcodes --trim_adapters --compress_fastq
```

Porechop:

```
./porechop-runner.py -i INPUT.fastq.gz -o OUTPUT.fastq.gz
```

Quality score filtering:

```
./qfilter -s 15 -p 70 -m 10 -l n INPUT.fastq > OUTPUT.fastq
```

BLAST:

```
./ncbi-blast-2.13.0+/bin/blastn -db DATABASE -query INPUT.fasta -out OUTPUT.txt
    -outfmt 6 -num_threads 20
```

Filtering the BLAST output:

```
awk '$4 >= MIN_ALIGNMENTLENGTH' INPUT.txt | awk '$3 >= 95.0' | sort -r -k1,1
    -k3,3 | awk '!x[$1]++' > OUTPUT.txt
```

Generating species and genera lists:

```
awk '{print $2}' INPUT.txt | sed 's/_/ /g' | awk '{print $2, $3}' | sort
    | uniq -c | sort -n -r > OUTPUT_species.txt
```

```
awk '{print $2}' INPUT.txt | sed 's/_/ /g' | awk '{print $2, $3}'
    | tr -c '[:alnum:]' '[\n*]' | sort | uniq -c | sort -n -r | grep '[A-Z]'
    > OUTPUT_genera.txt
```

Determining species richness and read abundance:

```
for x in $( for i in  rbcL/rbcL_lists/*species.txt; do echo $i | sed 's/.*rbcL
    //' ; done); do echo $x; awk -v datei=$x 'BEGIN{pfad="ITS/ITS_lists/ITS"
    datei; print "Datei: "datei; while(getline < pfad > 0){test1[$2" "$3]=$1}}
    {test2[$2" "$3]=$1; if(test1[$2" "$3] > 0){nbeide+=test1[$2" "$3]+=test2
    [$2" "$3]; beide++; print $2" "$3" in beiden"}else{nnur2+=test2[$2" "$3];
    nur2++;print $2" "$3" nur rbcL"}}END{for(x in test1){if(test2[x]==0)
    {nnur1+=test1[x]; nur1++; print x" nur in ITS"}}}END{print "#\t"datei"
    \t"nur1"\t"beide"\t"nur2"\t"nnur1"\t"nbeide"\t"nnur2}' rbcL/rbcL_lists/
    rbcL$x; done | grep "#" | less
```

Data availability

All generated data, plots, tables, lists and further commands of this study are in progress for a public repository. At current time, the data is in house.