High-throughput biodiversity assessment – Powers and limitations of meta-barcoding

Hochdurchsatzerfassung von Biodiversität – Stärken und Grenzen von Meta-barcoding

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Integrative Biology

submitted by

Wiebke Sickel

from

Oranienburg

Würzburg, 2016

Submitted on: ……..……..

Office stamp

Members of the *Promotionskomitee*:

Chairperson: Prof Dr Thomas Müller

Primary Supervisor: Dr Alexander Keller

Supervisor (Second): Prof Dr Ingolf Steffan-Dewenter

Supervisor (Third): Prof Dr Jörg Schultz

Date of Public Defence: ……..……

Date of Receipt of Certificates: ……..

Affidavit

I hereby confirm that my thesis entitled 'High-throughput biodiversity assessment - powers and limitations of meta-barcoding' is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg, 07 October 2016
Place, Date Signature

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation 'Hochdurchsatzerfassung von Biodiversität - Stärken und Grenzen von Meta-barcoding' eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, 07. Oktober 2016
Ort, Datum Unterschrift

Acknowledgements

I am highly grateful to my three supervisors, Dr Alexander Keller, Prof Dr Ingolf Steffan-Dewenter and Prof Dr Jörg Schultz. Firstly, for giving me the opportunity to pursue a PhD; secondly for encouraging me to develop and conduct my own research projects; and thirdly for their constant support and advice during these last three years.

I would like to thank the rest of the Molecular Biodiversity Group for valuable discussions, helping me out when needed and generally for creating a great working environment; thank you Gudrun Grimmer, Markus J. Ankenbrand, Anna Voulgari-Kokota, and the Bachelor and Master students Annette Brandel, Mira Becker, Rebecca Balles and Jonas Stelz.

Many thanks go to my collaborators Dr Giulia Zancolli, Dr Ivonne Meuche, Dr Andrea Holzschuh, Dr Frank Förster, Prof Dr T. Ulmar Grafe, Dr Dieter Mahsberg, Dr Stephan Härtel and Jonathan Lanzen. I would also like to thank the staff of the Graduate School of Life Sciences.

Further thanks go to the following people who so far have not been named: Niklas Terhoeven of the Computational Evolutionary Biology Group within the Centre for Computational and Theoretical Biology, Dr Simone Rost and Jens Gräf of the Human Genetics Department, Janina Kay of the Department for Neurobiology and Genetics as well as my family and (other) friends.

Summary

Traditional species identification based on morphological characters is laborious and requires expert knowledge. It is further complicated in the case of species assemblages or degraded and processed material. DNA-barcoding, species identification based on genetic data, has become a suitable alternative, yet species assemblages are still difficult to study. In the past decade meta-barcoding has widely been adopted for the study of species communities, due to technological advances in modern sequencing platforms and because manual separation of individual specimen is not required. Here, meta-barcoding is put into context and applied to the study of bee-collected pollen as well as bacterial communities. These studies provide the basis for a critical evaluation of the powers and limitations of meta-barcoding. Advantages identified include species identification without the need for expert knowledge as well as the high throughput of samples and sequences. In microbiology, meta-barcoding can facilitate directed cultivation of taxa of interest identified with meta-barcoding data. Disadvantages include insufficient species resolution due to short read lengths and incomplete reference databases, as well as limitations in abundance estimation of taxa and functional profiling. Despite these, meta-barcoding is a powerful method for the analysis of species communities and holds high potential especially for automated biomonitoring.

Zusammenfassung

Traditionelle Methoden der Identifizierung von Organismen anhand von morphologischen Merkmalen sind arbeits- und zeitaufwendig und benötigen Expertenkenntnisse der Morphologie. Weitere Probleme liegen in der Analyse von Artgemeinschaften und prozessiertem Material. DNA-barcoding, Artbestimmung anhand von genetischen Merkmalen, hat sich als Alternative herausgebildet, jedoch sind Artgemeinschaften nach wie vor schwierig zu analysieren. Im vergangenen Jahrzehnt wurde meta-barcoding zur Analyse von Artgemeinschaften entwickelt; insbesondere durch die Weiterentwicklung moderner Sequenziergeräte und da eine Auftrennung der Organismen innerhalb einer Gemeinschaft nicht mehr notwendig ist. In der vorliegenden Arbeit wurde zunächst ein Überblick über meta-barcoding erstellt. Die Methode wurde dann für die Analyse von Bienen-gesammeltem Pollen und Bakteriengemeinschaften angewandt. Diese Studien bilden eine gute Basis, um die Vor- und Nachteile von meta-barcoding kritisch zu bewerten. Vorteile beinhalten unter anderem, dass Organismen bestimmt werden können, ohne dass Expertenkenntnisse notwendig sind, sowie der hohe Durchsatz von Proben und Sequenzen. In der Mikrobiologie kann meta-barcoding eine gerichtete Kultivierung von Bakterien erleichtern, die durch meta-barcoding als Zielorganismen indentifiziert wurden. Nachteile finden sich in der manchmal noch unzureichenden Unterscheidung nah verwandter Arten aufgrund von kurzen Sequenzlängen und lückenhaften Referenzdatenbanken, sowie Einschränkungen in der Abschätzung von Abundanzen und Funktionen der Organismen innerhalb der Artgemeinschaft. Trotz dieser Problematiken ist meta-barcoding eine leistungsstarke Methode für die Analyse von Artgemeinschaften und ist besonders vielversprechend für automatisiertes Bio-Monitoring.

Table of contents

Part I.

Introduction

Morphological species identification: A central aspect of biology with limitations

The classification of a specimen to a species remains a central aspect of biology (Wiens and Servedio 2000). Its application ranges from systematic biology and ecology to conservation biology (Wiens and Servedio 2000; Balakrishnan 2005). It can also become important in food safety (Woolfe and Primrose 2004) and law enforcement (Ogden et al. 2009). Traditionally, species identifications are based on morphological characters (Wiens and Servedio 2000; Balakrishnan 2005). Sometimes, other aspects are included, such as behaviour (Balakrishnan 2005). However, there are many situations in which these aspects are not feasible or simply impossible, for example, when the specimen has been processed in some way, which is the case in gut contents (Soininen et al. 2009; Valentini et al. 2009; Pompanon et al. 2012) or in traditional Chinese medicine (Yip et al. 2007; Li et al. 2011; Coghlan et al. 2012).

One other example, where classification based on morphological characters is not sufficient, is pollen analysis, which traditionally utilises light microscopy (Mullins and Emberlin 1997). The pollen grains of closely-related plant species very closely resemble one another, so often the lowest taxonomic level that can be identified is plant family (Williams and Kremen 2007; Galimberti et al. 2014). In addition, pollen grain classification is very laborious (Galimberti et al. 2014) and requires expert knowledge of the respective bioregion the pollen was collected in (Keller et al. 2015).

In the case of bacteria, species identification is further complicated by the need to bring them into culture to study them in detail (Handelsman and Smalla 2003). However, cultivation in standard media only captures a low amount of bacterial diversity to study (Handelsman and Smalla 2003). Additionally, diversity of morphological characters is limited in bacteria (Handelsman and Smalla 2003), which means that numerous tests of bacterial morphology and physiology are needed to describe a bacterial specimen further and to classify it (Gerner-Smidt et al. 1991; Mata et al. 2002; Edberg et al. 1986). This is very tedious and can become very costly. Although some rapid tests for certain groups of bacteria exist (Holmes et al. 1994; Odumeru et al. 1999), these are often tailored for pathogen identification (Holmes et al. 1994; Odumeru et al. 1999) or are restricted to specific groups of bacteria (Nord et al. 1974) and thus fail to detect, yet even classify undescribed species.

DNA-barcoding: Species identification based on genetic material

Thus, in the past decades, species identification based on genetic information, DNA-barcoding, has become an important tool. This is especially the case for organisms difficult to study otherwise, like bacteria, and processed or degraded material. DNA-barcoding relies on a simple comparison of DNA sequences of an unclassified specimen to reference sequences of known identity. It has been quickly adopted in microbiology (Woese and Fox 1977; Fox et al. 1977) and is nowadays required in bacterial species descriptions (Stackebrandt et al. 2002; Blaxter 2016). Since its invention it has also been applied to higher eukaryotes (Hebert et al. 2003). In principle, DNA is isolated from an unknown specimen, a particular part of the genome, the marker gene, is amplified, sequenced and compared to a sequence obtained from a specimen of known identity. If the difference between two sequences is below a certain threshold, e.g. 97% in the case of bacteria (Stackebrandt and Goebel 1994), the unknown specimen is assigned to the same species. If the difference is larger, the unknown specimen belongs to a different species. Usually, the query sequence is compared to a variety of sequences of known identity, saved in a database, such as GenBank (Benson et al. 2013) or BOLD (Ratnasingham and Hebert 2007). If the database contains sequences similar to the query sequence, a species identification can be made. If not, the specimen remains unclassified, but it is still possible to assign it to higher taxonomic levels, such as genus or family, by adjusting the threshold.

The fragment of the genome used for DNA-barcoding is called the marker gene. A good marker gene needs to fulfill various requirements. Firstly, DNA sequence dissimilarities between closely-related species need to be large enough to tell the species apart, so species resolution of the marker gene needs to be sufficient (Hollingsworth et al. 2011). At the same time, the differences within a single species should not be too large to avoid wrongfully assigning specimens of the same species to separate species. In other words, a good marker gene exhibits a barcode gap, which means that the sequence variation within a single species is lower than and does not overlap with the sequence variation between species (Chen et al. 2010; Schoch et al. 2012). For correct classification at higher taxonomic levels, the marker gene needs to represent the genetic disparity at these levels as well. Secondly, the marker gene should also exhibit high amplification success rates across a variety of species (Chen et al. 2010). Ideally, the marker gene can be successfully amplified from a large group of different organisms, such as

different families or even phyla, with a universal set of primers and still distinguish organisms at the species level (Chen et al. 2010; Hollingsworth et al. 2011; Coissac et al. 2012). Thirdly, the length of the fragment to be analysed should be sufficient to cover enough sequence differences for species distinction, but at the same time short enough for successful amplification in degraded material (Chen et al. 2010; Coissac et al. 2012).

For different groups of organisms, different marker genes have proven to be suitable for DNA-barcoding. In bacteria, the 16S ribosomoal RNA (rRNA) gene is most commonly used (Woese and Fox 1977; Fox et al. 1977; Stackebrandt et al. 2002; Blaxter 2016). For animals, the mitochondrial cytochrome oxidase I (COI) gene (Hebert et al. 2003; Benson et al. 2013) and for fungi the internal transcribed spacer (ITS) within the ribosomal cistron (Schoch et al. 2012) have been identified as suitable markers. In the case of plants, the choice of a marker gene is more complicated, and several genes have been proposed with very different success rates (Chen et al. 2010).

Although DNA-barcoding is commonly used for species identification, it is still single specimens that are being dealt with and in the case of assemblages, such as pollen from sediment cores, collected by bees or communities of bacteria, the need to manually separate individuals from one another persists. Again, this is laborious and taxa of low abundance might be missed out (Bent and Forney 2008; Pompanon et al. 2012). Especially in the case of bacteria, the importance of studying complete assemblages rather than focusing on specific taxa has recently been recognised (Junker et al. 2011; Keller et al. 2013; Kueneman et al. 2013). It has become generally accepted that the study of bacterial communities as a whole bears novel and important findings concerning bacterial ecology (Keller et al. 2013; Kueneman et al. 2013). In host-microbe associations, for example, it is probably the bacterial community as a whole rather than singular strains that contributes to the interaction (Junker et al. 2011; Keller et al. 2013; Kueneman et al. 2013).

Thus, an alternative approach that would allow identifying all species within an assemblage simultaneously and without prior separation would highly benefit various research fields, such as agro-ecology (Williams and Kremen 2007; Krupke et al. 2012), palaeo-ecology (Behling et al. 2004; Davies and Tipping 2004; Gugerli et al. 2005), diet analysis (Valentini et al. 2010; Soininen et al. 2009) and community ecology (Peterson et al. 2008; Beil et al. 2008; Loudon et al. 2014), but also applications such as food safety (Galimberti et al. 2014; Bruni et al. 2015), allergen load assessment (Kraaijeveld et al. 2015) and safety issues with medicinal preparations (Coghlan et al. 2012).

Meta-barcoding: Analysing species communities

In the past decade, major advances have been made with high-throughput sequencing (HTS) platforms, improving sequence length and quality whilst dropping costs (Shokralla et al. 2012; Zinger et al. 2012). The ability to read millions of DNA sequences simultaneously (Shokralla et al. 2012), allows species identification of theoretically all species within a sample in parallel, which is termed meta-barcoding. In meta-barcoding, DNA is extracted from a mixture of organisms or specimens, such as soil, gut contents or pollen assemblages. Next, the marker gene is amplified, similarly to DNA-barcoding. However, because the marker gene is amplified from a mixture of different DNA sources, universal primers suitable for amplification from closely and distantly related species at the same time is even more important than in DNA-barcoding. The universal primers need to be sufficient for successful amplification of all organisms in the sample, but also with the same efficiency, so as not to introduce skews in the data (Coissac et al. 2012).

The immense throughput of modern sequencing platforms further allows the analysis of multiple samples at the same time, called multiplexing. In this case, each sample is specifically labeled with a short sequence of known base composition, an index sequence, which is sequenced alongside the actual DNA barcode (Binladen et al. 2007; Kozich et al. 2013). The sample indices allow mapping of obtained DNA sequences to individual samples later on in raw data processing. Multiplexing requires an an additional step in sample processing, normalisation (Harris et al. 2010; Kozich et al. 2013), to account for differential amplification success between samples, which would skew the sequencing output per sample dramatically. Sample processing requires some additional preparation based on the sequence platform chosen, the whole process is called library preparation. Once fully prepared for sequencing, the base composition of each sequence in the sample is read by the sequencer alongside the sample index, if multiplexing is performed.

Most sequencing platforms perform two sequencing runs, one forward and one reverse sequence read. These can be joined during raw data processing (Aronesty 2011), to cover a longer barcode sequence. Further common data processing steps include quality filtering, since the data often contain sequencing errors (Coissac et al. 2012; Kozich et al. 2013), demultiplexing, i.e. mapping sequences to samples, and chimera checking, which removes a common PCR artifact (Caporaso et al. 2010; Edgar et al. 2011). Taxonomic classification in meta-barcoding generally relies on the same princi-

ples as DNA-barcoding (see above). However, the vast amount of sequencing data, with up to 600 Gigabases with the Illumina HiSeq 2000 (Coissac et al. 2012; Shokralla et al. 2012), requires some amendments. Running every single sequence against a database would be computationally very intensive, so sequences are clustered into operational taxonomics units (OTUs) before taxonomic classification (Caporaso et al. 2010; Edgar 2010; Edgar 2013; Blaxter 2016). In bacteria, OTUs are commonly clustered based on a 97% sequence identity threshold (Cuesta-Zuluaga and Escobar 2016), other groups of organisms might require other thresholds. For actual taxonomic identification, one representative sequence per OTU is run against the chosen reference database and taxonomy is assigned based on this representative (Ji et al. 2013).

In microbiology, meta-barcoding has been adopted rapidly, because it bypasses cultivation and subsequently the separation of individual strains before community assembly (Zinger et al. 2012), but the potential for diet analysis (Valentini et al. 2009; Soininen et al. 2009; Pompanon et al. 2012) and biodiversity monitoring (Shokralla et al. 2012) has also been realised. Meta-barcoding has successfully been applied to a variety of higher organisms as well, including fungi (Bálint et al. 2014), plants (Keller et al. 2015) and animals (Yu et al. 2012).

Objectives

This thesis aims to critically analyse the powers and limitations of meta-barcoding. To achieve this, the first step is to create an overview of the role of meta-barcoding in biology (Publication P.1). Then, meta-barcoding was applied to the analysis of bee-collected pollen with the aim to establish a method for pollen meta-barcoding (Publications P.2 and P.3). In the third instance, bacterial communities were analysed with meta-barcoding to infer host-microbe associations (Publications P.4 and P.5). I then critically evaluated meta-barcoding with a focus on its application in biodiversity assessments.

Part II.

Publications

Main findings of the publications

In the following, the main findings of the five publications are summarised. Publication P.1 is a mini-review of meta-barcoding in biological research and thus gives a short overview of this methodology. Publications P.2 and P.3 introduce a method for pollen analysis using meta-barcoding. Afterwards, publications P.4 and P.5 apply meta-barcoding to bacterial communities in association with reptile (P.4) and plant (P.5) hosts.

Mini-review. The overview (Publication P.1) summarises the principle behind DNA-barcoding in general and the workflow of meta-barcoding in particular. It highlights some of the main advantages of meta-barcoding over other approaches but also introduces the challenges associated with that method. These will be discussed in more detail later in the thesis.

Pollen analysis. Applying meta-barcoding to the analysis of pollen (Publication P.2) constitutes in the first instance a proof of principle. It was possible to establish a method for sequencing DNA extracted from pollen and to automatically identify the plant origin of the pollen with a bioinformatical pipeline. However, there are many different approaches for pollen meta-barcoding (Richardson et al. 2015; Kraaijeveld et al. 2015) and even slight differences in protocols limit the comparability of studies. Thus, a detailed protocol for a suggested standard method for pollen meta-barcoding was developed (Publication P.3). It contains step by step descriptions of both the laboratory and the bioinformatic workflow.

The laboratory method was based on a previously published pipeline for bacterial community analysis Kozich et al. 2013 and was adapted to be suitable for pollen analysis. Incorporated into the oligo scaffold provided by Kozich et al. 2013 were primers amplifying the internal transcribed spacer 2 (ITS2), a genetic marker, which has been reported suitable for plant barcoding (Chen et al. 2010). Additionally, this marker was chosen because a comprehensive database of ITS2 sequences was available (Schultz et al. 2006).

The method cannot only be applied to pollen analysis but also to any other question of plant species identification in mixed samples or samples where taxonomic identification is not possible otherwise.

Bacterial community analysis. Analysing bacterial communities constitutes the most important application of meta-barcoding because it is difficult to study bacteria otherwise. Meta-barcoding was applied to the analysis of

bacterial communities in two study systems: pet reptiles (Publication P.4) and Bornean pitcher plants (Publication P.5). It was possible to (i) refute the long-held belief that pet reptiles carry human pathogens in their oral cavities and (ii) describe the bacterial community in two *Nepenthes* species. The main rationale for applying meta-barcoding in these cases was to circumvent the methodological bias of bacterial isolation and cultivation on standard media and instead of this, describe the complete bacterial community.

CHAPTER 1

Metabarcoding put into context

P.1 DNA-Metabarcoding - ein neuer Blick auf organismische Diversität

Authors:

Alexander Keller, Gudrun Grimmer, Wiebke Sickel*, Markus J. Ankenbrand
*Co-author

published: 2016

Journal: *BIOspektrum*

Volume: 22, Pages: 147-150

Genetische Ökologie

DNA-Metabarcoding – ein neuer Blick auf organismische Diversität

ALEXANDER KELLER, GUDRUN GRIMMER, WIEBKE SICKEL, MARKUS J. ANKENBRAND
LEHRSTUHL FÜR TIERÖKOLOGIE UND TROPENBIOLOGIE, BIOZENTRUM, UNIVERSITÄT WÜRZBURG

Taxon identification is one of the fundamental challenges in biological research. Usually, classifications are based on specimen morphology, sometimes supported by their behaviour, ecology or biochemistry. Technological advances now allow using genomic fragments as a taxon barcode. With the latest developments of high-throughput sequencers this can go even further: identifying complete assemblages simultaneously, with various applications in ecology, conservation, forensics and health security.

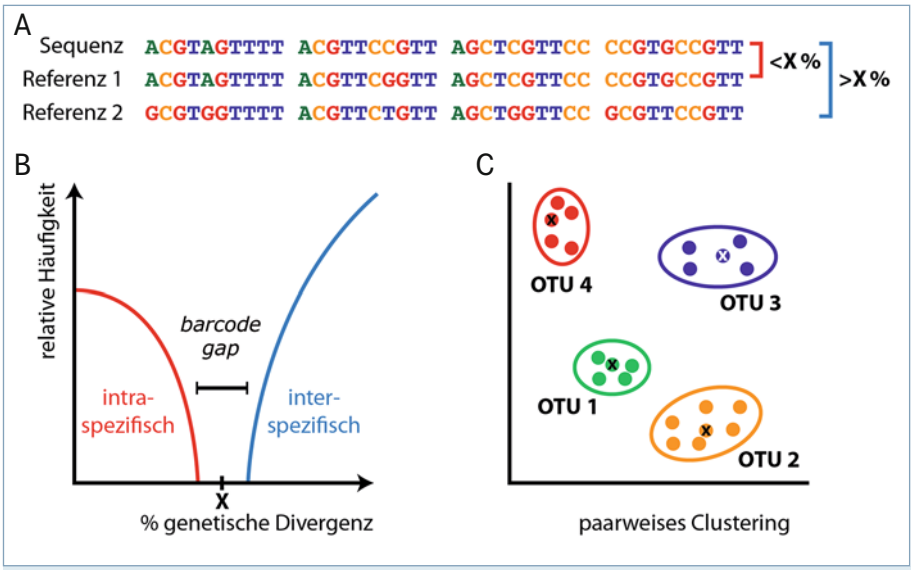
DOI: 10.1007/s12268-016-0669-0

■ Die Identifikation von Organismen stellt eine der grundlegendsten und ältesten Herausforderungen in der biologischen Forschung dar. Traditionell wird diese Erkennung und Abgrenzung von anderen Lebewesen über morphologische Merkmale durchgeführt, ggf. werden je nach taxonomischer Gruppe auch ethologische, biochemische oder ökologische Informationen zurate gezogen. Durch die technologischen Entwicklungen in den vergangenen Jahren stehen uns heute zusätzlich genomische Daten in Form von DNA-Sequenzen zur Verfügung, die auch bei der Klassifizierung und Unterscheidung von Organismen hilfreich sein können.

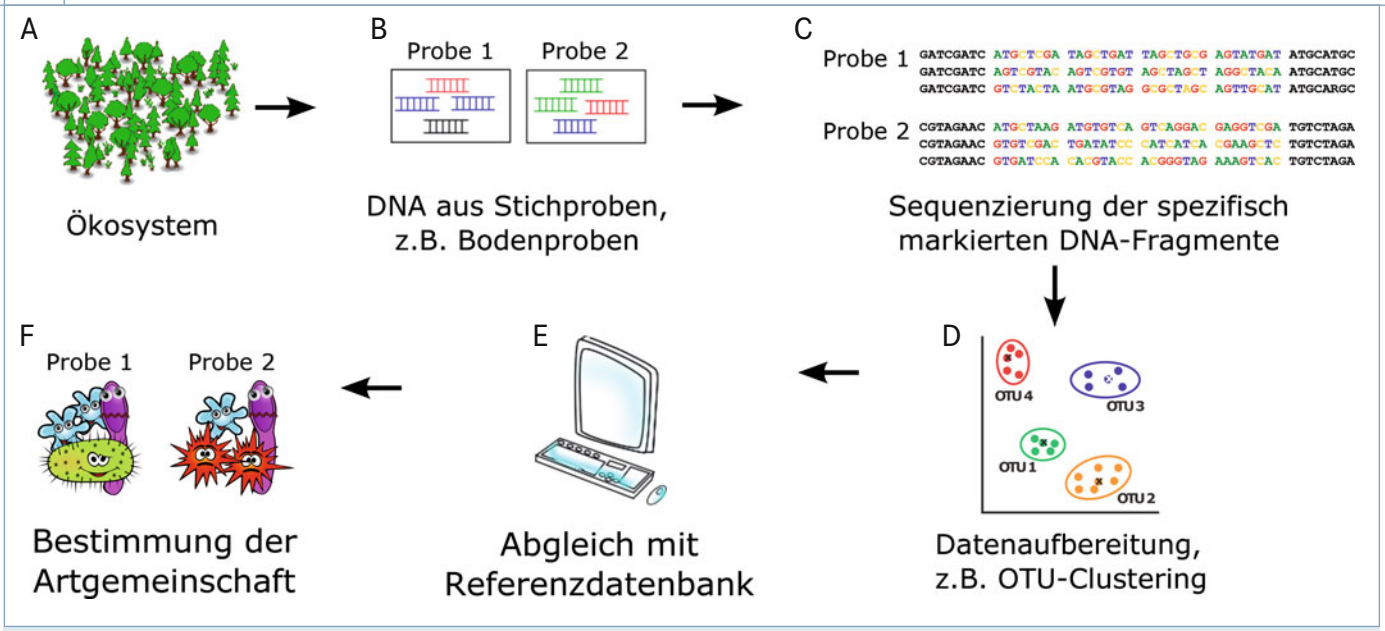
DNA-Barcoding unterstützt die traditionelle Arterkennung

In der Diversitätsforschung wurde die Sequenzierung genomischer DNA-Fragmente schon relativ früh eingesetzt, um die evolutive Geschichte von Organismen zu rekonstruieren [1]. Dabei werden Sequenzen unterschiedlicher Organismen miteinander verglichen, Unterschiede ermittelt und diese zur Erstellung eines phylogenetischen Stammbaums verwendet. Vor allem in der Mikrobiologie etablierte sich diese Methode schnell, da sie nicht mehr auf die wenigen erfassbaren Merkmale der Individuen angewiesen war [1]. Dementsprechend verwundert es nicht, dass erste Schritte zur Katalogisierung von Organismen anhand von Sequenzen auch in mikrobiologischen Werken zu finden sind [2]. Erst im Jahr 2003 wurde diese Methode unter dem Namen DNA-Barcoding auch für höhere Eukaryoten etabliert [3]. Inzwischen ist die Methode weit verbreitet und wird durch zahlreiche Initiativen gestützt. Die grundlegenden Ziele des DNA-Barcodings sind die flächendeckende Katalogisierung der organismischen Diversität und deren Nutzung als Referenz für weiterführende Fragestellungen.

Das Prinzip des DNA-Barcodings besteht darin, ein kurzes Fragment der genomischen DNA zu analysieren, das repräsentativ für eine bestimmte Art ist und eindeutig auf diese zurückgeführt werden kann. Über einen bioinformatischen Vergleich mittels eines Schwellenwertes (*barcoding gap*) kann die Identität einer unbekannten Sequenz anhand einer Referenzdatenbank bestimmt werden (**Abb. 1A**). Dieser Schwellenwert wird so definiert, dass intraspezifische von interspezifischer genomischer Variation unterschieden wird (**Abb. 1B**). Ein großer Vorteil dieser



▲ **Abb. 1:** Bioinformatischer Ablauf einer DNA-Barcoding-Studie. **A**, Sequenzidentitäten mit Referenzen kleiner dem Schwellenwert X gelten als erfolgreiche Artidentifizierung. **B**, X wird bestimmt durch die *barcode gap* zwischen der Variation innerhalb einer Art und zu anderen Arten. **C**, Einordnung ähnlicher Sequenzen in taxonomische Einheiten (OTU, *operational taxonomic unit*) eines Metabarcoding-Datensatzes; nur eine repräsentative Sequenz wird mit der Datenbank abgeglichen.



▲ **Abb. 2:** Überblick über Metabarcoding. Ein Ökosystem (**A**) mit schwer unterscheidbaren Arten wird untersucht und die DNA aus verschiedenen Stichproben isoliert (**B**) und sequenziert (**C**). Nach der Datenaufbereitung (OTU, *operational taxonomic unit*; **D**) und einem Datenbankabgleich (**E**) wird die Artgemeinschaft für jede Stichprobe separat ermittelt (**F**).

Methode ist die Reproduzierbarkeit der Identifikation. Eine erfolgreiche Arterkennung kann somit nicht nur von erfahrenen Taxonomen und Experten bestimmter Artengruppen durchgeführt werden. Für die taxonomischen Großgruppen werden meist unterschiedliche genomische Bereiche verwendet: Für Bakterien ist die ribosomale 16S-RNA etabliert, für Pilze ITS(*internal transcribed spacer*)-Bereiche, für Pflanzen Abschnitte der ITS oder Plastid-Gene, wohingegen bei Tieren dominant mitochondriale Marker eingesetzt werden. Neuere Studien setzen verschiedene Regionen kombiniert ein, um die taxonomische Sicherheit zu erhöhen [4].

Erfassung komplexer Artgemeinschaften mit DNA-Metabarcoding

Neue Hochdurchsatztechnologien erlauben es nun, einen Schritt weiterzugehen. Es wird eine Vielzahl von Sequenzen aus einer Ausgangsprobe generiert; im Kontext der Diversitätsforschung kann dies eingesetzt werden, um nicht nur einzelne Individuen, sondern eine Vielzahl von Organismen simultan zu erfassen (**Abb. 2**, [5]). Moderne Plattformen erlauben hierbei außerdem, verschiedene Proben gleichzeitig zu prozessieren (*multiplexing*), dabei wird jede Probe spezifisch markiert (**Abb. 2C**).

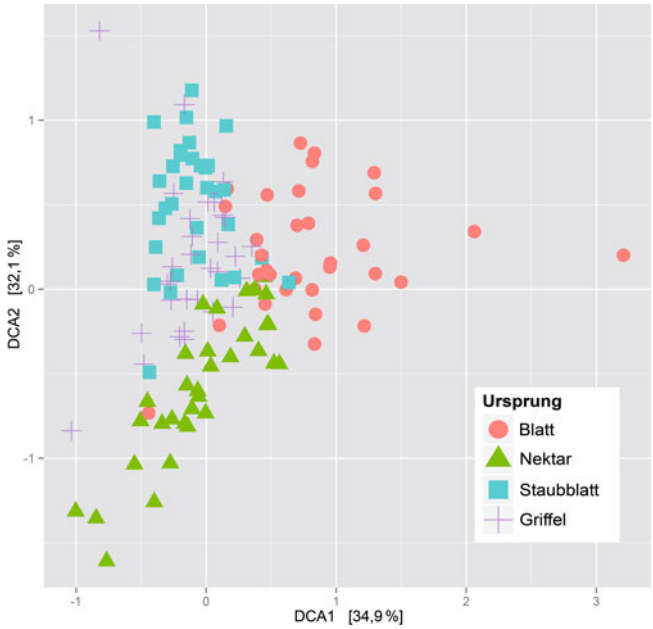
Je nach Technologie ergeben sich mehrere Millionen Sequenzen, sodass der direkte Vergleich mit Referenzdatenbanken unpraktikabel wird. Man verwendet daher oft einen Zwischenschritt: Über ein Clustering-Verfahren werden innerhalb eines Datensatzes Sequenzen nach Ähnlichkeit in taxonomische Einheiten (OTUs, *operational taxonomic units*) zusammengefasst (**Abb. 1D**). Aus diesen Einheiten wird jeweils nur eine repräsentative Sequenz mit der Referenzdatenbank verglichen. Da besonders im mikrobiellen Bereich der Anteil an unbekannten Organismen sehr groß werden kann, werden zudem Algorithmen eingesetzt, die bei fehlenden Referenzsequenzen die unbekannte Sequenz so gut wie möglich in übergeordnete Gruppen klassifizieren (z. B. Gattung, Familie, Ordnung).

Auch das Metabarcoding etablierte sich zuerst in der bakteriellen Ökologie. Komplette Gemeinschaften werden hier auf einmal erfasst, ohne die einzelnen Organismen vorher zu trennen [6]. Es bedarf auch keiner vorherigen Kultivierung der einzelnen Bakterien, welche für einen Großteil nicht praktikabel ist. Obwohl diese Methode noch sehr jung ist, hat sie schon enorm zu einem neuen Verständnis von mikrobieller Diversität und der Strukturierung von Gemeinschaften beigetragen [6]. Die Etablierung des Metabarcodings befindet sich derzeit auch für Eukaryoten im Aufwind und verspricht hier ebenso eine gute Erfassung der Biodiversität. Artgemeinschaften von Pilzen [7], Pflanzen [8] und Tieren [9] konnten über die Hochdurchsatzsequenzierung bereits erfolgreich erfasst werden und ermöglichen einen neuen Blick auf die Mechanismen der Etablierung und Strukturierung von Artgemeinschaften und Ökosystemen.

Jedoch ergeben sich durch das Metabarcoding auch neue Herausforderungen. Die Abundanzwerte stellen nicht unbedingt die tatsächliche Abundanz einer erfassten Art dar. Da die zugrunde liegende Polymerasekettenreaktion (PCR) kein linearer Prozess ist, kann es zu einer Überschätzung oder Unterschätzung kommen [5]. Hinzu kommt, dass die Biomasse zwischen den Arten variieren kann und dass diese auch unterschiedlich gut labortechnisch aufgeschlossen werden können. Beide Faktoren beeinträchtigen die Vergleichbarkeit von Abundanzen zwischen den Arten. Durch qualitativ schlechte Sequenzierergebnisse können Sequenzen fehlklassifiziert werden und damit zu einer artifiziellen Überschätzung der tatsächlichen Biodiversität führen. Von entscheidender Bedeutung für jede taxonomische Klassifizierung eines Metabarcoding-Datensatzes ist die Quantität und Qualität der zugrunde liegenden Referenzdatenbank, in welcher sich auch fehlerhafte Sequenzen befinden können, besonders bei nicht-kurierten Datenbanken [10]. Dem Großteil dieser neuen Schwierigkeiten kann durch eine akkurate bioinformatische Auswertung und diverse Korrekturmechanismen nach der Sequenzierung entgegengewirkt werden. Auch hier verspricht die Umstellung von einzelnen auf mehrere Marker Vorteile, ist derzeit jedoch analytisch schwerer umsetzbar als bei Einzelorganismen [11].

Anwendungsbereiche von DNA-Metabarcoding

Biodiversitätserfassung und Charakterisierung von Artgemeinschaften stellen einen essenziellen Bestandteil der ökologischen Forschung und des Naturschutzes dar. Die Mög-



◀ **Abb. 3:** Feinskalige Analyse von Bakteriengemeinschaften auf Blüten. Die Datenpunkte entsprechen Einzelproben mit jeweils einer gesamten Artgemeinschaft, aufgetragen nach Ihrer Ähnlichkeit zueinander mittels DCA(*detrended correspondence analysis*)-Ordinationsanalyse. Mikrohabitate wie Griffel, Nektar- und Staubblätter einer Blüte sowie die Blätter stellen sehr unterschiedliche Voraussetzungen für Bakterien dar. Sie beherbergen dadurch mehrere verschiedene, diverse und gut unterscheidbare Gemeinschaften, die ohne Metabarcoding bisher unterschätzt wurden (nach [6]).

lichkeit, Proben im Hochdurchsatz und ohne Auftrennung in einzelne Individuen prozessieren zu können, erlaubt generell eine Erhöhung der Stichproben (und damit der statistischen Sicherheit) sowie der Anzahl an durchführbaren Experimenten [11]. Es können auch für taxonomisch schlecht erfasste Gebiete und Artgruppen Studien durchgeführt werden. Zudem kann die Eingliederung in ökologische Nischen sehr feinskalig untersucht werden, da wenig Ausgangsmaterial notwendig ist (**Abb. 3**, [6]). Es ergeben sich neue Möglichkeiten des Naturschutzes durch das Metabarcoding von Umgebungs-DNA. Im aquatischen Bereich kann der Nachweis bedrohter oder invasiver Arten durch abgestoßene Hautzellen, Exkremente oder andere Körperbestandteile direkt über das Wasser erfolgen, ohne dass Individuen gefangen werden müssen. Durch die Beprobung von Erdschichten können Rückschlüsse auf die Biodiversität im Verlauf der Erdgeschichte gezogen werden [11].

Metabarcoding wird zudem sehr erfolgreich bei der Erfassung von zwischenartlichen Interaktionen sowie zur Identifikation von Pathogenen und Symbionten eingesetzt [6]. Es können Netzwerke aus Pflanzen und deren Bestäubern direkt über die Sequenzierung von Pollen erfasst werden [8]. Die Bedeutung von bakteriellen Gemeinschaften im Darmtrakt für die Immunabwehr und die Nährstoffversorgung ist bekannt, doch bietet die neue Forschungsmethode nun die Möglichkeit, diese Gemeinschaften systematisch zu untersuchen und im Kontext diverser Hintergründe (z. B. Ernährung und Krankheiten) auszuwerten.

Für die Sicherung des Lebensqualität der Menschen kann das Metabarcoding in einer Vielzahl von Bereichen eingesetzt werden [11]. Die Nahrungsqualität kann durch die Erfassung der pflanzlichen und tierischen Bestandteile überprüft und gesichert werden. Allergene wie Pollen in der Luft sowie Blüteereignisse bei Algen können frühzeitig erfasst und damit präventive Maßnahmen eingeleitet werden. Das Metabarcoding kann außerdem zur Erfassung von Krankheitserregern verwendet werden und damit zur Hygiene in Städten und Verkehrszentren beitragen. Kliniken sowie wissenschaftliche Labore können durch regelmäßige Prüfung auf Kontaminationen hin untersucht werden. Auch forensische Analysen lassen sich durch die Methode verbessern, indem Algen, Pollen und weitere Pflanzenbestandteile zur Ursprungsermittlung herangezogen werden.

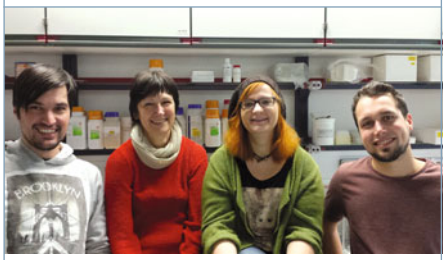
Die Bandbreite an Applikationen ist groß, und durch die anhaltenden technologischen Weiterentwicklungen wird sowohl die Qualität als auch die Quantität der Daten durch Metabarcoding ständig verbessert und kosteneffizienter gestaltet. Mit dieser Entwicklung zeigt sich auch ein Trend in der Ausbildung der Wissenschaftler, von taxonomischen Experten hin zu bioinformatischen Analytikern. Diese verschiedenen Blickwinkel, von Metabarcoding und traditionellen Erfassungsmethoden zusammen, erlauben es, unser Wissen über Biodiversität und Artgemeinschaften deutlich zu erweitern und die Mechanismen hinter Ökosystemen zu verstehen. ■

Literatur

[1] Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 74:5088–5090
[2] Fox GE,Pechman KR, Woese CR (1977) Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to procaryotic systematics. Int J Syst Evol Microbiol 27:44–57
[3] Hebert PD, Cywinska A, Ball SL et al. (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270:313–321
[4] Dupuis J, Row A, Sperling F (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. Mol Ecol 21:4422–4436
[5] Keller A, Danner N, Grimmer G et al. (2015) Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biol 17:558–566
[6] Junker RR, Keller A (2015) Microhabitat heterogeneity across leaves and flower organs promotes bacterial diversity. FEMS Microbiol Ecol 91:fiv097 (doi: 10.1093/femsec/fiv097)
[7] Bálint M, Schmidt P, Sharma R et al. (2014) An Illumina metabarcoding pipeline for fungi. Ecol Evol 4:2642–2653
[8] Sickel W, Ankenbrand M, Grimmer G et al. (2015) Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC Ecol 15:20
[9] Yu D, Ji Y, Emerson B et al. (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods Ecol Evol 3:613–623
[10] Nilsson R, Ryberg M, Kristiansson E et al. (2006) Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS One 1:e59
[11] Bell K, de Vere N, Keller A et al. (2016) Pollen DNA barcoding: current applications and future prospects. Genome (im Druck)

Korrespondenzadresse:
Dr. Alexander Keller
Lehrstuhl für Tierökologie und Tropenbiologie,
Biozentrum
Universität Würzburg
Am Hubland
D-97074 Würzburg
Tel.: 0931-31-84427
Fax: 0931-31-84352
a.keller@biozentrum.uni-wuerzburg.de

ARBEITSGRUPPE



Alexander Keller, Gudrun Grimmer, Wiebke Sickel und Markus J. Ankenbrand (v. l. n. r.)

Die *Molecular Biodiversity Group* der Universität Würzburg ist eine Nachwuchsforscher-AG, die sich mit ökologischen Artgemeinschaften beschäftigt. Wir analysieren Wirt-Mikroben-Interaktionen von Pflanzen und Insekten sowie Pflanzen-Bestäuber-Interaktionen. Zudem sind wir in der Methodenentwicklung für Metabarcoding aktiv.

CHAPTER 2

Pollen analysis

P.2 Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach

Authors:

Wiebke Sickel*, Markus J. Ankenbrand, Gudrun Grimmer, Andrea Holzschuh, Stephan Härtel, Jonathan Lanzen, Ingolf Steffan-Dewenter, Alexander Keller

*First author

published: 2015

Journal: *BMC Ecology*

Volume: 15, Pages: 20

Sickel *et al. BMC Ecol (2015) 15:20*
DOI 10.1186/s12898-015-0051-y

BMC Ecology

METHODOLOGY ARTICLE

Open Access

Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach

Wiebke Sickel, Markus J Ankenbrand, Gudrun Grimmer, Andrea Holzschuh, Stephan Härtel, Jonathan Lanzen, Ingolf Steffan-Dewenter and Alexander Keller*

Abstract

Background: Meta-barcoding of mixed pollen samples constitutes a suitable alternative to conventional pollen identification via light microscopy. Current approaches however have limitations in practicability due to low sample throughput and/or inefficient processing methods, e.g. separate steps for amplification and sample indexing.

Results: We thus developed a new primer-adapter design for high throughput sequencing with the Illumina technology that remedies these issues. It uses a dual-indexing strategy, where sample-specific combinations of forward and reverse identifiers attached to the barcode marker allow high sample throughput with a single sequencing run. It does not require further adapter ligation steps after amplification. We applied this protocol to 384 pollen samples collected by solitary bees and sequenced all samples together on a single Illumina MiSeq v2 flow cell. According to rarefaction curves, 2,000–3,000 high quality reads per sample were sufficient to assess the complete diversity of 95% of the samples. We were able to detect 650 different plant taxa in total, of which 95% were classified at the species level. Together with the laboratory protocol, we also present an update of the reference database used by the classifier software, which increases the total number of covered global plant species included in the database from 37,403 to 72,325 (93% increase).

Conclusions: This study thus offers improvements for the laboratory and bioinformatical workflow to existing approaches regarding data quantity and quality as well as processing effort and cost-effectiveness. Although only tested for pollen samples, it is furthermore applicable to other research questions requiring plant identification in mixed and challenging samples.

Keywords: DNA barcoding, High throughput sequencing, Illumina MiSeq platform, ITS2, Next generation sequencing, NGS, *Osmia*, Palynology, Pollination ecology

Background

Identification of pollen origin is a central aspect in pollination ecology studies [1–3] and agro-ecological research [4, 5]. Conventional pollen identification utilises light microscopy and discriminates species according to morphological characteristics [6]. This requires expert knowledge for the bioregion and taxa of interest [7], is time-consuming [8] and lacks discriminatory power at lower taxonomic levels [4, 8].

A promising approach to circumvent these issues has been to identify plant species in pollen samples by DNA sequence analysis. This can be done by, for example, cloning amplified PCR products into plasmids and sequencing a subset of clones [8, 9] or sequencing pollen grains of interest [10, 11] or bee crop contents directly [12]. However, this often does not reflect the complete diversity of plant species present, since only a subset of DNA sequences are analysed or only dominant plant taxa can be detected. Recent studies [7, 13–15] have identified

*Correspondence: a.keller@biozentrum.uni-wuerzburg.de
Department of Animal Ecology and Tropical Biology, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany

high throughput sequencing (HTS) approaches based on meta-barcoding as a suitable alternative for existing methods. However, current protocols still suffer from a limited sample throughput [7, 14, 15] and/or practicability issues due to separate steps for PCR amplification and index labelling [13]. We here present a protocol for highly multiplexed pollen sequencing utilising a dual-indexing strategy [16]. An overview of existing methods alongside our new approach is given in Figure 1. We designed meta-barcoding primers suitable for plant identification using the internal transcribed spacer 2 (ITS2) that already incorporate Illumina-specific adapters for high-throughput sequencing as well as new sequencing primers that are added to the sequencing flow cell. The rationale for using ITS2 rather than other genetic markers for plant DNA barcoding in general is provided elsewhere [17] and its applicability regarding meta-barcoding criteria has also been successfully demonstrated [7, 13]. We tested our new approach by sequencing 384 pollen samples collected by two solitary bees species with known different foraging strategies: polylectic *Osmia bicornis* [18] and oligolectic *Osmia truncorum* [19]. Alongside this enhancement of the laboratory method, we updated the reference database used for ITS2 meta-barcoding [7] and added compatibility for the UTAX classification software [20] as a second and alternative strategy beside the RDP classifier [7, 21].

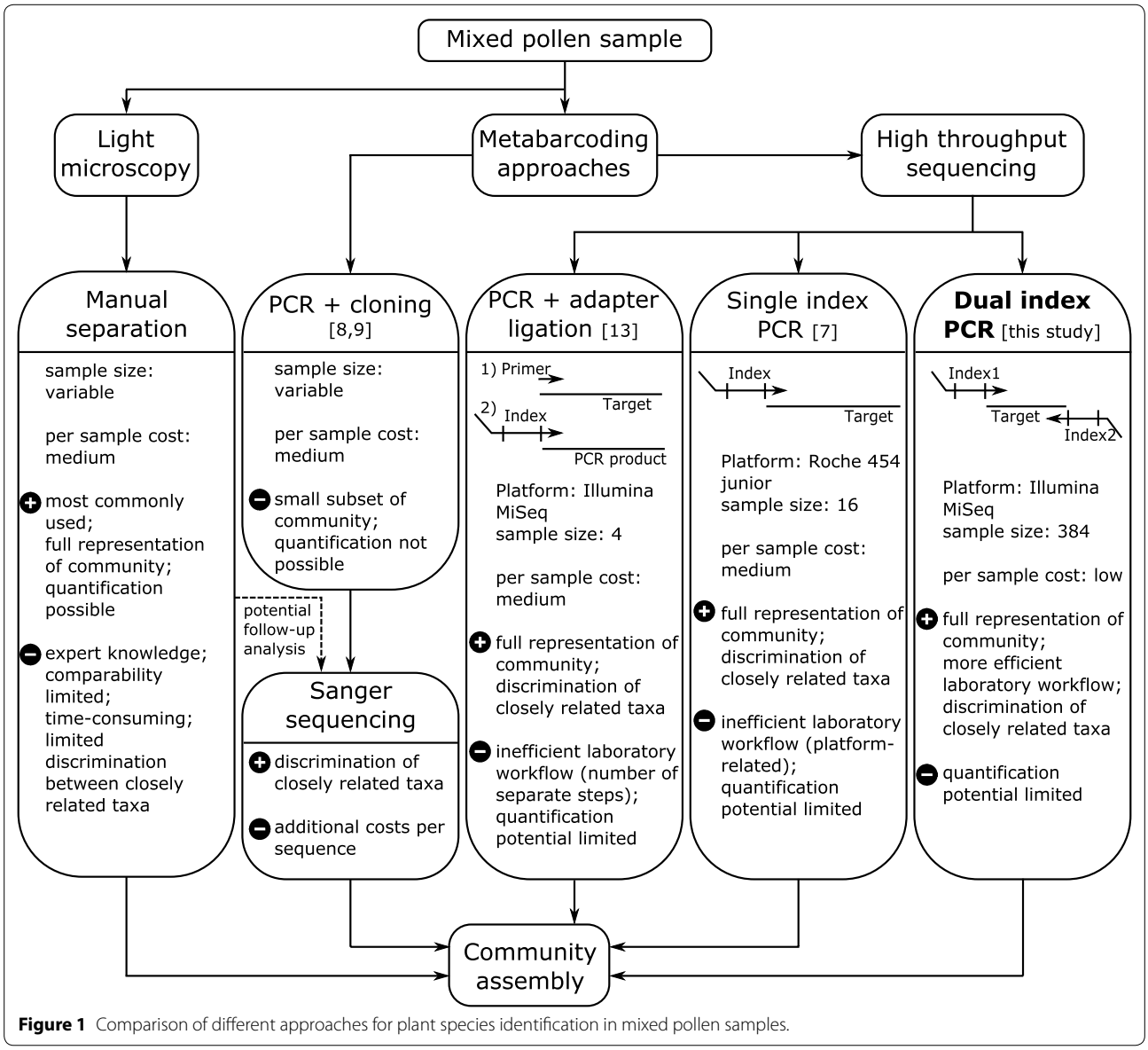


Figure 1 Comparison of different approaches for plant species identification in mixed pollen samples.

Methods

Dual-indexing design

As amplifying primers we used the well-established combination of plant barcoding primers ITS-S2F [17] and ITS4R [22]. These were already used for plant species identification based on meta-barcoding [7] and deliver a fragment of suitable size for MiSeq v2 sequencing using 500 cycles. For MiSeq-conformity, we expanded each of the primers according to the overall oligo scaffold described in Kozich et al. [16]. This scaffold consists of MiSeq-specific adapters, an 8nt index sequence, a 10nt pad as well as a 2nt linker sequence and lastly the amplifying primers. To successfully transfer the scaffold design to ITS2 sequencing, we ensured by minor modifications that the melting temperature (T_m) of the combined pad, linker and amplifying primer was ~65°C (see Additional file of Kozich et al. [16]) enabling the read primers to bind during the later sequencing procedure. In the forward scaffold, we adapted the pad sequence from 5′-TATGGTAATT -3′ to 5′-**CC**TGGT**GC**T**G**-3′ (adapted nucleotides in bold). The pad of the reverse scaffold remained unchanged. Complete sequences of the final oligos were forward: 5′-AATGATACGGCGACCACCGAGATCTACACXXXXXXXX **CC**TGGT**GC**T**G**GT**ATGCGATACTTGGTGTGAAT**-3′ and reverse: 5′-CAAGCAGAAGACGGCATACGAGAT XXXXXXXX AGTCAGTCAG CC**TCCTCCGCTTATTGATATGC**-3′, where adapted nucleotides are denoted in bold and XXXXXXXX indicates the index sequences used for multiplexing. Both primer sequences were thus 32nt long, had a T_m of 64.8°C, a 50% GC content and exhibited low self-complementarity (longest dimer complement: 4 bp). They amplify a total fragment of approximately 470–480 bp, including the complete ITS2 sequence. The actual sequenced part of this fragment covers 350–360 bp (target only) and is thus within the range of 2 × 250 cycles sequencing, leaving some buffer for joining the paired end reads. We used 16 forward index sequences SA501–SB508 and 24 reverse indices SA701–SB712, allowing a total of 384 unique combinations for sample indexing (Additional file of Kozich et al. [16]). With ITS2-specific modifications, it was also necessary to modify the sequencing primers that are added to the MiSeq flow cell. We thus changed read and index primers as follows (adapted nucleotides in bold): Read1: 5′-**CC**TGGT**GC**T**G**GT **ATGCGATACTTGGTGTGAAT**-3′, Read2: 5′-AGTCAGTCAG CC**TCCTCCGCTTATTGATATGC**-3′, Index: 5′-**GCATAT-CAATAAGCGGAGGA**GG CTGACTGACT-3′.

Processing test samples

The newly designed dual-indexing approach was evaluated with mixed pollen samples, collected from nests of the solitary bees *Osmia bicornis* (270 samples), *Osmia truncorum* (111 samples) and other *Osmia spp.* (3 samples) at various sites near Würzburg, Germany from April to September 2013. Different samples originated from pools of two different brood cells from the same nest (likely the same mother bee few days apart). We chose this study system because we wanted to demonstrate that different foraging strategies can be detected using pollen meta-barcoding. We documented flower resources available during the sample period within a 50 m radius (all plant species) and within a 600 m radius (mass-flowering plants only) around the nest sites. This was done to gain information on species identity of flower resources available for bee foraging at the time of sampling (Additional file 1) and to be able to compare them with our sequence data.

DNA from ~0.003 g pollen grains was isolated as described by Keller et al. [7] using the Macherey-Nagel Food Kit (Düren, Germany). PCR was performed in three separate 10 µL reactions in order to avoid PCR bias [23]. Each reaction contained 5 µL 2 × Phusion Master Mix (New England Biolabs, Ipswich, MA, USA), 0.33 µM each of the forward and reverse primers, 3.34 µL PCR grade water and 1 µL DNA template. PCR conditions were as follows: initial denaturation at 95°C for 4 min, 37 cycles of denaturation at 95°C for 40 s, annealing at 49°C for 40 s and elongation at 72°C for 40 s; followed by a final extension step at 72°C for 5 min. Each sample was assigned a different forward/reverse index combination for sample-specific labelling. Triplicate reactions of each sample were combined after PCR and further processed as described in Kozich et al. [16], including between-sample normalization using the SequalPrep™ Normalization Plate Kit (Invitrogen GmbH, Darmstadt, Germany) and pooling of 96 samples. These pools were quality controlled using a Bioanalyzer High Sensitivity DNA Chip (Agilent Technologies, Santa Clara, CA, USA), quantified with the dsDNA High Sensitivity Assay (Life Technologies GmbH, Darmstadt, Germany), and afterwards combined to a single pool containing all 384 samples. This was diluted to 8 pM, denatured and spiked with 5% Phix Control Kit v3 (Illumina Inc., San Diego, CA, USA) according to the Sample Preparation Guide (llumina Inc. 2013). Sequencing was performed on the Illumina MiSeq using 2 × 250 cycles v2 chemistry (Illumina Inc., San Diego, CA, USA).

Data analysis

Raw sequence reads were obtained from the Illumina MiSeq output directly, which includes sample reads already demultiplexed by the MiSeq Reporter v. 2.5.1.3 with perfect index matches only. Forward and reverse reads were joined using the join_paired_ends.py command in QIIME v.1.8.0 [24] using default parameters. Low quality reads were removed (<Q20, <150 bp,

ambiguous base-pairs) with USEARCH v8.0.1477 [25]. Combined reads were taxonomically classified with the RDP classifier [21] as well as the UTAX algorithm and results compared to show that the data is compatible between both alternative analytical strategies. UTAX and RDP were executed for each sample separately.

In the following, we concentrate on UTAX, since the RDP classifier has been used previously for pollen taxonomic assignments [7]. A raw score cut-off at 20 was used, as the UTAX algorithm does currently not provide bootstrap comparable confidence values (but is expected to incorporate these soon, see http://drive5.com/usearch/manual/faq_taxconfs.html, accessed 2015/22/05). These assignment scores are however comparable between reads as long as subsequent analyses do base all upon the same database.

For data analysis, the raw UTAX output was parsed using a self-written perl script, which counts the number of assignments for each taxon and aggregates these into a single table (https://github.com/iimog/meta-barcoding-dual-indexing). This table is converted into a community matrix format, with rows as species and columns representing samples, and a separate file with the taxonomic lineage of each species is also created. These files are directly importable into common statistical software, e.g. *R* v.3.1.2 [26] using the package *phyloseq* v.1.6.1 [27]. To assess sufficiency of the sequencing depth, we created species accumulation curves for each sample using the *vegan* package v2.2-0 [28] in *R* v.3.1.2 [26], excluding taxa accounting for less than 0.1% of sample reads. Additionally, we determined the ten most abundant plant families collected by *O. bicornis* and *O. truncorum*.

Reference database update

Beside the enhancement of the laboratory protocol, we considered it important to address also the actuality and completeness of the reference database. We thus performed an update according to the annotation pipeline described for the ITS2 database [29, 30]. For this, we extracted all available ITS2 sequences belonging to Viridiplantae from GenBank [31] (accessed on 2015/19/01) as described in detail in Koetschan et al. [30]. The taxonomy follows the NCBI taxonomy database [32], which may not perfectly reflect evolutionary status, but is well usable for automatic procedures, due to its integration into the public NCBI framework. Taxonomy was assigned to the sequences by mapping the gi to the NCBI taxid. Taxonomic levels were selected at seven levels (kingdom, phylum, class, order, family, genus, species) using a custom perl script utilizing the NCBI::Taxonomy module by courtesy of F. Förster (doi:10.5281/zenodo.17375). RDP training files, a UTAX database and taxtree were created with a custom perl script (https://github.com/iimog/meta-barcoding-dual-indexing). The database update, scripts and information on how to use it with the RDP classifier or UTAX are provided at http://www.dna-analytics.biozentrum.uni-wuerzburg.de.

Results

Sequencing output and data analysis

In total we obtained 11,624,087 raw ITS2 reads (PhiX excluded), which accounted for an average of 30,271 [standard deviation (SD): 11,373; median: 30,900] reads per sample. After data processing (low-quality <Q20, short reads <150 bp, ambiguous base-pairs), a mean of 15,580 (SD 6,598; median 15,740) reads per sample remained. Species accumulation curves (Figure 2) show that almost all samples were sequenced to saturation after approximately 2,000–3,000 high quality reads. Based on the ratio of raw to high quality reads, this accounts for approximately 4,000–6,000 raw reads required. Per sample pollen in bee brood cells originated from between one and 85 different plant species (Figure 2). Five per cent of samples (19) yielded an output of less than 2,000 reads (minimum saturation threshold, Figure 2), which were removed prior to further analysis. Raw sequences are accessible via the EBI-SRA with the project accession number PRJEB8640.

Reference database update

Our previously published database contained 73,853 reference sequences of 37,403 unique plant species [7]. The updated version now contains 182,505 plant sequences from 72,325 different species. This is an increase by factor 2.47 (147% additional) for sequences and 1.93 (93% additional) for unique species. In comparison with the original reference set [7], with these data 80.1% (original 53.1%) of the plant species and 90.4% (original 75%) of the genera in Bavaria, Germany, where our test samples originate from, were covered (data retrieved from http://bayernflora.de; accessed on 2015/01/24). Correspondingly, for plant species in the USA, the database covers 66.5–79.1% (median 76.1%) of species and 73.8–87.3% (median 84.9%) of genera, depending on the US state (data retrieved from the BISON project; http://bison.usgs.ornl.gov; accessed on 2015/04/02). In both cases, Bavaria and USA, missing species are likely rare or endemic to specific regions. A comparison of numbers of genera per order covered in the old and updated database versions can be found in the Additional file 2: Table S1.

Test samples

Regarding our samples, taxonomic classification (after filtering out rare taxa below 0.1%) identified 650 different plant taxa, of which 617 could be classified taxonomically to plant species level, belonging to 288 genera, 71

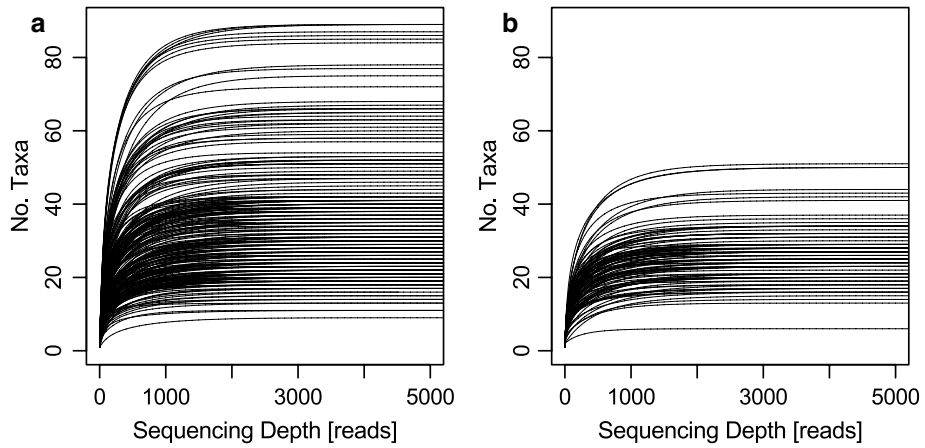


Figure 2 Species accumulation curves. **a** *Osmia bicornis* samples; **b** *Osmia truncorum* samples. The x-axis was limited to 5,000 reads as the saturation of all samples was below this threshold. The y-axis was limited to 90 taxa in both plots to obtain the same scale. Taxa accounting for less than 0.1% of total sample reads were excluded.

families, 37 orders and nine classes. The remaining 33 taxa (5%) could not be classified at the species level. Of these, 17 taxa could still be classified at genus level and another seven at the family level. Nine taxa remained that could not be classified even to family level. These belonged to the Sapindales, Fagales and Microthamniales (one taxon each) or remained unclassified (six taxa). At the genus level, RDP and UTAX taxonomic assignments agreed in ~90% of all read classifications, thus both classifiers yielded comparable results.

For both *Osmia* species together, approximately 50% of documented plant genera (<50 m: all plants, 50–600 m: only mass-flowering plants) were detectable within the sequencing data and contributed with ~75% to all quality-filtered reads. The two bee species differed clearly in foraging patterns as visible through plant families predominantly collected (Figure 3), as well as in the number of plant species with *O. bicornis* collecting up to 85 plant species and *O. truncorum* collecting up to 50 plant species per brood cell (Figure 2). The ten most abundant plant families collected by *O. bicornis* were Brassicaceae (27.07%), Ranunculaceae (16.98%), Aceraceae (11.62%), Fagaceae (10.86%), Juglandaceae (7.16%), Papaveraceae (5.91%) Fabaceae (5.40%), Asteraceae (4.89%), Rosaceae (3.59%) and Plantaginaceae (2.62%). *O. truncorum* pollen was dominated by Asteraceae (92.92%), and only Caprifoliaceae (1.51%) and Brassicaceae (1.14%) contributed more than 1% to the overall collection. The Asteraceae collected by *O. truncorum* contained a wide spectrum of plant genera, with 58 genera being detected, the ten most abundant of which were *Picris, Jacobaea, Tanacetum, Artemisia, Achillea, Tripleurospermum, Inula, Cota, Leucanthemum* and *Crepis* (Figure 3).

Discussion

High throughput sequencing (HTS) has been shown to be successful and valuable for taxonomic assessment of mixed pollen samples [7, 13, 15]. The drawbacks of existing protocols were the low number of samples processed simultaneously or inefficient multistep library preparations. Recent developments in sequencing technologies allow far larger multiplexing, given the enormous throughput already available with desktop NGS devices. Highly multiplexed sample processing has already been established for bacterial assessments using dual-indexing approaches with the MiSeq sequencer [16]. It was the goal of this study to transfer this knowledge to the field of plant meta-barcoding, in our specific case of pollen samples.

By adapting the primer design to the ITS2 region, modifying the oligo scaffold design, and adjusting the sequencing primers to be compatible with the MiSeq device, we successfully established a fast pollen DNA meta-barcoding routine with high multiplexing capabilities. For our test samples, the newly designed primers were used to sequence 384 mixed pollen samples collected by solitary bees with a single sequencing run. In the original bacterial dual-indexing protocol [16], the potential for higher multiplex rates than 384 samples is suggested depending on required throughput to assess the diversity. Our sequencing results indicate that for pollen samples at least a depth of 2,000–3,000 high quality reads per sample should be reached to identify all taxa within the sample (plateau reached, Figure 2), which was comparable for the two bee species under study. However, this is of course highly dependent on number of plant species in the samples, which may be dependent on

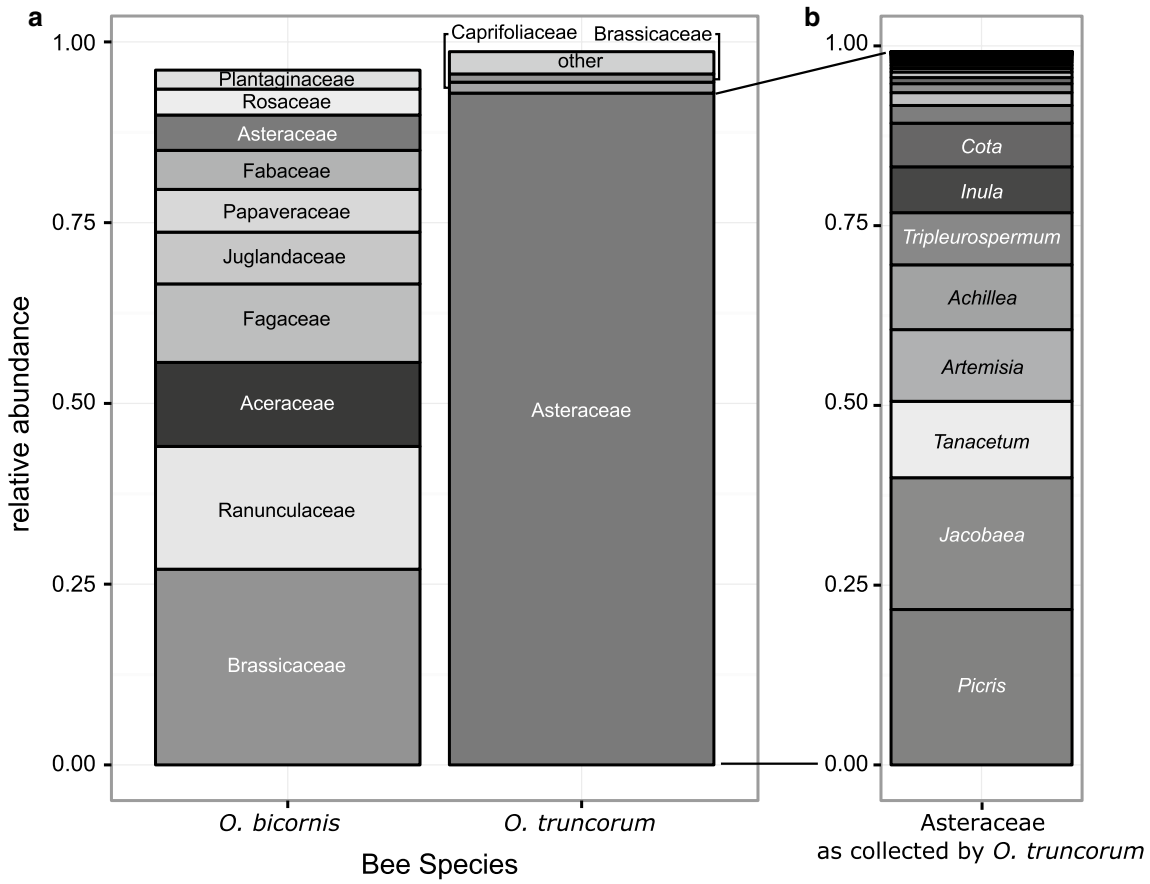


Figure 3 Pollen spectrum of the two bee species. **a** Ten most abundant families as collected by the bee species *O. bicornis* and *O. truncorum*. For *O. truncorum* 'other' include the families Apiaceae, Rosaceae, Fabaceae, Ranunculaceae, Plantaginaceae, Juglandaceae and Amaranthaceae. **b** Plant genera detected within the Asteraceae collected by *O. truncorum*. For visualisation reasons, only the eight most abundant genera are labelled. Please note that Aceraceae is now included within Sapindaceae.

sample origin, foraging behaviour and the biodiversity of the ecosystem of interest, but may serve nonetheless as a guideline for higher multiplex rates. Additional index combinations for more samples are provided in the Additional files alongside the protocol for the bacterial dual-index approach [16].

Beside our dual-indexing strategy, another HTS-based approach has been recently proposed. There, PCR amplification and index labelling were conducted in separate steps [13], which is time and labour-intensive and introduces a further step where errors may be introduced. In our protocol, PCR amplification and sample indexing occur simultaneously, which is highly practical and requires no special reagents, such as additional expensive library preparation kits or adapter ligation chemicals. In our protocol, the complete workflow accounts for less than USD 20.00 for materials per sample, when processing 384 samples simultaneously. This is much lower than conventional pollen analysis under the light microscope, which can reach several hundred USD per sample.

Most plant taxa detected could be successfully classified using the already shown RDP classifier [7, 21], but also the recently developed UTAX algorithm [25]. Due to the missing confidence values for taxonomic assignments in UTAX version 8.0 (announced for version 8.1, http://drive5.com/usearch/manual/faq_taxconfs.html, accessed 2015/22/05), we compared the classifications to the RDP output as well as the documented flower resources. UTAX and RDP showed high agreement between taxonomic classifications, thus both may be used arbitrarily.

Approximately half of the genera found flowering near the nest sites were detected in the pollen samples. This is attributable to bee foraging preferences, where not all available resources might be used, especially for the oligolectic *O. truncorum*. Secondly, about three quarters of the reads were assigned to plant genera documented near the nesting sites (<50 m: all plant species, 50–600 m:

mass-flowering plants only). As bees are expected to forage also further away, the remaining reads are attributable to pollen collected from undocumented plants or misclassifications.

According to our expectation, pollen composition patterns were very different for the oligolectic and the polylectic bee species (Figure 3). *O. truncorum* samples were dominated by Asteraceae, whereas *O. bicornis* samples showed a wide pollen spectrum. Our data correspond to flower preferences and foraging strategies known for these species [18, 19]. This supports the high quality of information obtained by pollen meta-barcoding, as already intensively evaluated in another study [7]. It is noteworthy that even very rare taxa could be detected, which is of special interest in the oligolectic *O. truncorum* and might be overlooked in light microscopy assessment of pollen samples.

We would like to point out that abundance data obtained from molecular approaches should in general be interpreted with care and only as relative abundance (divided by total number of reads in the sample to account for varying library sizes). Contradicting results exist concerning the suitability of pollen meta-barcoding for quantification purposes, with Keller et al. [7] and Kraaijeveld et al. [14] finding a positive significant correlation between genera by light microscopy and meta-barcoding, whilst Richardson et al. [13] were not able to find such a connection. Due to the different steps in the workflow, e.g. dilutions and PCR, biases can be introduced, leading to skewed data and over- or underrepresentation of certain taxa. PCR bias is considered to be a random process and can be accounted for by performing replicate PCR reactions for each sample [23], which are pooled subsequently. We followed this approach in this study likewise to Keller et al. [7] to avoid PCR bias as far as possible. This may explain some of the discrepancy between studies, although a recent study indicated that PCR replicates might not be necessary in pollen meta-barcoding [14]. The reduced amount of individual processing steps of direct indexing, (as performed here and in both studies identifying positive correlation [7, 14]) further reduces additional risks to introduce unwanted effects in comparison with the study using adapter ligation that shows no correlation [13].

In this study, samples of the same bee species show high consistency in abundance patterns of major taxa, which are easily biologically explainable. A good compromise for most studies investigating foraging patterns might be to not use direct count data, but conservatively categorising plant taxa into 'abundant' and 'rare' based on a threshold, as proposed by Keller et al. [7]. Where more detail is needed, a subset of samples may also be analysed in parallel by light microscopy for evaluation purposes [7, 13, 14].

One major advantage of pollen meta-barcoding is that no expert knowledge on pollen morphology is required for taxonomic assignment. Additionally, species level assignment is possible even for closely related plant taxa. However, successful taxonomic assignment critically depends on the quality of the reference database. Our target marker was the ITS2 region, but other genetic markers might also be considered for plant species identification using meta-barcoding, e.g. *trn*L [14, 15] or *rbcL* plus *trnH-psbA* [8, 9]. The described dual indexing approach [16] can also be applied to other genetic markers, provided some considerations are taken into account as described for ITS2 in this study. On the laboratory side of the workflow, firstly target and thereby primer choice should be appropriate for universal amplification and plant species identification based on DNA sequence data. The amplified fragment should be of the appropriate size for the chosen MiSeq sequencing chemistry, e.g. no longer than ~480–490 bp for 2 × 250 v2 sequencing kits, allowing for some overlap between forward and reverse reads. Given these conditions are met, primer design can be performed following the guidelines from Kozich et al. [16] including the required modifications to the various oligonucleotides. However, as mentioned before, successful plant species identification relies to a large degree also on the underlying reference database and bioinformatical classification algorithm. For most alternative markers comprehensive reference databases are currently lacking and thus taxonomic classifications are mainly performed by a BLAST search [33] against sequences downloaded from GenBank [8, 9, 13–15], locally managed alternative databases [9] and/or newly acquired DNA sequences [8, 9]. BLAST searches are based on local alignments that may only use parts of each sequence (e.g. conserved regions) for classification, lack a hierarchy classification procedure and results can be difficult to interpret [7, 17] especially when results show hits for multiple, different taxa. Setting up locally managed databases is time- and labour-intensive a well as costly and makes it difficult to compare independent studies with one another. In the case of the ITS2 region, we benefitted from the already established ITS2 database [30], which contains annotated and trimmed ITS2 sequences from species worldwide and can be publicly accessed, improving overall comparability across studies.

Although Chen et al. [17] reported high identification accuracies with ITS2 as a genetic marker, some plant taxa could not be identified in recent studies on pollen meta-barcoding [7, 13]. These included the families Salicaceae, Lamiaceae [13] and Vitaceae [7] and the genera *Lonicera* [13], *Heracleum*, *Carduus*, *Phacelia*, *Convolvulus* and *Helianthus* [7], although they had been identified with microscopic pollen analysis. In

this study, we could detect all of these taxa. Failure to detect these families and genera with DNA sequence data was most likely due to incompleteness of the reference databases in these studies. Richardson et al. [13] used in total only 2,628 reference sequences, that described about half of the locally occurring plant species. In the case of Keller et al. [7], we were able to directly compare the database then (73,853 sequences) and now (182,505 sequences), which revealed that for each of those plant taxa more reference sequences were included after the database update presented here (Additional file 3: Table S2). This explains the positive detection for those plant taxa in this study in contrast to earlier studies and again highlights the importance of a current and comprehensive reference database for meta-barcoding purposes.

Our test samples comprised only pollen samples collected by bees, but in general ITS2 meta-barcoding can be applied to plant identification in other research fields where mixed samples are encountered, such as diet analysis of herbivores [34, 35] and in palaeo-ecology [36–38]. Furthermore, high-throughput DNA analysis of mixed plant samples can also prove valuable in food safety issues [39], honey quality analysis [8, 9] as well as allergen load assessment [14]. For such applications, alteration of the provided protocol for library preparation and sequencing is not needed, although the DNA extraction process may require alternative kits or adapted protocols specific for the material of interest.

Conclusions

We have successfully transferred a high-throughput technique for bacterial community sequencing to pollen meta-barcoding, which now enables labour- and cost-effective analysis of up to 384 mixed pollen samples simultaneously, thereby omitting drawbacks of previously established methods. We furthermore enhanced the database used for plant taxa identification based on HTS data. Additionally, our method should be easily adaptable to sample analysis of mixed plant origin in other research fields.

Availability of supporting data

The data set supporting the results of this article are in the EBI-SRA repository, under the project accession number PRJEB8640. Data on regional flora has been retrieved from http://bayernflora.de for Bavaria (accessed on: 2015/01/24) and from http://bison.usgs.ornl.gov/ for the USA (accessed on 2015/04/02). The database update, scripts and information on how to use it with the RDP classifier or UTAX are provided at http://www.dna-analytics.biozentrum.uni-wuerzburg.de and https://github.com/iimog/meta-barcoding-dual-indexing.

Additional files

Additional file 1: Plant species documented near solitary bee nest sites.

Additional file 2: Table S1. Comparison of the number of genera per order for all orders.

Additional file 3: Table S2. Comparison of the number of sequences per group for selected taxonomic groups.

Abbreviations

HTS: high throughput sequencing; ITS2: internal transcribed spacer 2; T_m: melting temperature.

Authors' contributions

WS designed the new primers, participated in laboratory work, undertook data analysis and drafted the manuscript. MJA performed the database update, scripted the workflow with RDP classifier and UTAX and performed taxonomic classification. GG performed most of the laboratory work. AH, SH and ISD participated in the study design. AH and JL provided the pollen samples. AK conceived the study, performed bioinformatic processing and helped drafting the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was financially supported by the DFG Collaborative Research Center 1047, Insect Timing. MJA was further supported by a PhD fellowship from the Graduate School of Life Sciences of the University of Würzburg. We are highly grateful to the staff from the Department of Human Genetics (University of Würzburg, Germany), especially S. Rost for granting us access to the Illumina MiSeq device. Further thanks go to F. Förster (Department of Bioinformatics, University of Würzburg, Germany) for help with the ITS2 database update. We also thank P. Schloss (University of Michigan, USA) for initial advice on primer design.

Compliance with ethical guidelines

Competing interests

The authors declare that they have no competing interests.

Received: 15 April 2015 Accepted: 25 June 2015
Published online: 22 July 2015

References

1. Carvell C, Westrich P, Meek WR, Pywell RF, Nowakowski M (2006) Assessing the value of annual and perennial forage mixtures for bumblebees by direct observation and pollen analysis. Apidologie 37:326–340
2. Köppler K, Vorwohl G, Koeniger N (2007) Comparison of pollen spectra collected by four different subspecies of the honey bee Apis mellifera. Apidologie 38:341–353
3. Behl M, Horn H, Schwabe A (2008) Analysis of pollen loads in a wild bee community (Hymenoptera: Apidae)—a method for elucidating habitat use and foraging distances. Apidologie 39:456–467
4. Williams NM, Kremen C (2007) Resource distributions among habitats determine solitary bee offspring production in a mosaic landscape. Ecol Appl 17:910–921
5. Krupke CH, Hunt GJ, Eitzer BD, Andino G, Given K (2012) Multiple routes of pesticide exposure for honey bees living near agricultural fields. PLoS One 7:e29268
6. Mullins J, Emberlin J (1997) Sampling pollens. J Aerosol Sci 28:365–370
7. Keller A, Danner N, Grimmer G, Ankenbrand M, von der Ohe K, von der Ohe W et al (2015) Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biol 17:558–566
8. Galimberti A, De Mattia F, Bruni I, Scaccabarozzi D, Sandionigi A, Barbuto M et al (2014) A DNA barcoding approach to characterize pollen collected by honeybees. PLoS One 9:e109363

9. Bruni I, Galimberti A, Caridi L, Scaccabarozzi D, De Mattia F, Casiraghi M et al (2015) A DNA barcoding approach to identify plant species in multiflower honey. Food Chem 170:308–315
10. Parducci L, Suyama Y, Lascoux M, Bennett KD (2005) Ancient DNA from pollen: a genetic record of population history in Scots pine. Mol Ecol 14:2873–2882
11. Bennett KD, Parducci L (2006) DNA from pollen: principles and potential. Holocene 16:1031–1034
12. Wilson EE, Sidhu CS, LeVan KE, Holway DA (2010) Pollen foraging behaviour of solitary Hawaiian bees revealed through molecular pollen analysis. Mol Ecol 19:4823–4829
13. Richardson RT, Lin C-H, Sponsler DB, Quijia JO, Goodell K, Johnson RM (2015) Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. Appl Plant Sci 3:1400066
14. Kraaijeveld K, de Weger LA, Ventayol García M, Buermans H, Frank J, Hiemstra PS et al (2015) Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol Ecol Resour 15:8–16
15. Valentini A, Miquel C, Taberlet P (2010) DNA barcoding for honey biodiversity. Diversity 2:610–617
16. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79:5112–5120
17. Chen S, Yao H, Han J, Liu C, Song J, Shi L et al (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One 5:e8613
18. Gathmann A, Tscharntke T (2002) Foraging ranges of solitary bees. J Anim Ecol 71:757–764
19. Praz CJ, Müller A, Dorn S (2008) Host recognition in a pollen-specialist bee: evidence for a genetic basis. Apidologie 39:547–557
20. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996–998
21. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261–5267
22. White TJ, Bruns T, Lee S, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, New York, pp 315–322
23. Fierer N, Hamady M, Lauber CL, Knight R (2008) The influence of sex, handedness, and washing on the diversity of hand surface bacteria. Proc Natl Acad Sci USA 105:17994–17999
24. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336
25. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461
26. R Core Team (2014) R: A language and environment for statistical computing. Vienna, Austria. http://www.R-project.org/
27. McMurdie PJ, Holmes S (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217
28. Dixon P (2003) VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930
29. Keller A, Schleicher T, Schultz J, Müller T, Dandekar T, Wolf M (2009) 5.8S-28S rRNA interaction and HMM-based ITS2 annotation. Gene 430:50–57
30. Koetschan C, Förster F, Keller A, Schleicher T, Ruderisch B, Schwarz R et al (2010) The ITS2 Database III–sequences and structures for phylogeny. Nucleic Acids Res 38(Database issue):D275–D279
31. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J et al (2013) GenBank. Nucleic Acids Res 41(Database issue):D36–D42
32. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K et al (2011) Database resources of the national centre for biotechnology information. Nucleic Acids Res 39(suppl 1):D38–D51
33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
34. Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C et al (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. Front Zool 6:16
35. Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, Pompanon F et al (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. Mol Ecol Resour 9:51–60
36. Gugerli F, Parducci L, Petit RJ (2004) Ancient plant DNA: review and prospects. New Phytol 166:409–418
37. Behling H, Pillar VD, Orlóci L, Bauermann SG (2004) Late Quaternary Araucaria forest, grassland (Campos), fire and climate dynamics, studied by high-resolution pollen, charcoal and multivariate analysis of the Cambará do Sul core in southern Brazil. Palaeogeogr Palaeoclimatol Palaeoecol 203:277–297
38. Davies AL, Tipping R (2004) Sensing small-scale human activity in the palaeoecological record: fine spatial resolution pollen analyses from Glen Affric, northern Scotland. Holocene 14:233–245
39. Woolfe M, Primrose S (2004) Food forensics: using DNA technology to combat misdescription and fraud. Trends Biotechnol 22:222–226

P.3 Standard method for identification of bee pollen mixtures through meta-barcoding

Authors:

Wiebke Sickel*, Markus J. Ankenbrand*, Gudrun Grimmer, Frank Förster, Ingolf Steffan-Dewenter, Alexander Keller

*Equally contributing first authors

in press

Book: *COLOSS BeeBook*

Volume: III

Standard method for identification of bee pollen mixtures through meta-barcoding

Wiebke Sickel[1,*], Markus J Ankenbrand[1,*], Gudrun Grimmer[1], Frank Förster[2], Ingolf Steffan-Dewenter[1], Alexander Keller[1]

[1]Department of Animal Ecology and Tropical Biology, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany

[2]Department of Bioinformatics, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany

[*]equally contributing

Table of contents

Summary

Traditional pollen analysis via light microscopy has limitations in sample throughput as well as taxonomic resolution. Recently, pollen meta-barcoding methods have been developed as alternative approaches, where plant species identification of pollen grains works via DNA sequencing. However, these utilise different genetic markers and sequencing platforms lessening study comparability. We here describe a detailed protocol of the latest development in this field as a standard method for pollen meta-barcoding. It is highly cost-efficient, requires no palynological knowledge, is performable in standard laboratories and profits from a well-established reference database.

Key words

Apis mellifera, BEEBOOK, COLOSS, honey bee, Illumina MiSeq platform, ITS2, laboratory protocol, next generation sequencing, palynology, pollination ecology

Short title: Pollen meta-barcoding

1 Introduction

Pollen analysis is a central part of bee ecology research (Carvell et al. 2006; Köppler et al. 2007; Beil et al. 2008). Identification of plant species origin of bee collected pollen traditionally relies on light microscopy and discrimination based on morphological differences of pollen grains (Mullins & Emberlin 1997). However, this is labour- and time-intensive (Galimberti et al. 2014), requires expert knowledge (Keller et al. 2015) and lacks discriminative power at lower taxonomic levels (Williams & Kremen 2007; Galimberti et al. 2014), which means that pollen from closely related plant species often has to be combined at the family level. Recently, meta-barcoding has emerged as a suitable alternative for pollen analysis (Keller et al. 2015; Kraaijeveld et al. 2015; Richardson et al. 2015; Valentini et al. 2010). However, due to a missing consensus on the best marker for plant species identification and the variety of DNA sequencing platforms available, different methods and protocols exist (e.g. Kraaijeveld et al. 2015; Bruni et al. 2015; Galimberti et al. 2014; Richardson et al. 2015; Keller et al. 2015), which makes it difficult to compare independent studies. Additionally, most protocols suffer from limited sample-throughput, inefficient workflow and/or require additional costly chemicals, e.g. for adapter ligation, (Keller et al. 2015; Kraaijeveld et al. 2015; Richardson et al. 2015; Valentini et al. 2010). We here present a detailed protocol of the method described recently (Sickel et al. 2015) as a research standard that is highly cost-efficient and overcomes those limitations. It is based on ITS2-meta-barcoding, which has been validated for plant barcoding (Chen et al. 2010) and for which a comprehensive database has been established (Koetschan et al. 2010) and recently updated (Ankenbrand et al. 2015). Beside the laboratory process, we also provide information on data processing and analysis.

2 Meta-barcoding protocol

2.1 Required materials

2.1.1 Reagents

- DNA isolation kit suitable for pollen grains (e.g. Macherey-Nagel NucleoSpin Food, Düren, Germany)
- PCR grade water
- Ethanol (96 – 100 %)
- Primers as given in Table 1
- Polymerase with proof-reading ability including dNTPs, GC buffer and co-factors (e.g. 2 x Phusion Master Mix)
- Agarose, suitable buffer (e.g. TAE), intercalating dye (e.g. Midori Green Advance, Biozym Scientific GmbH, Hessisch Oldendorg, Germany), 6 x loading dye, DNA ladder (e.g. FastRuler Low Range DNA Ladder, Life Technologies, Carlsbad, CA, USA)
- SequalPrep™ Normalisation Kit 96 wells (Invitrogen, Carlsbad, CA, USA)
- Bioanalyzer High Sensitivity DNA Chip (Agilent Technologies, Santa Clara, CA, USA)
- dsDNA High Sensitivity Assay (Life Technologies, Carlsbad, CA, USA)
- MiSeq Reagent Kit v2 2 x 250bp (Illumina Inc., San Diego, CA, USA)
- 1N NaOH (stock solution)
- PhiX Sequencing Control v3 (Illumina Inc., San Diego, CA, USA)

2.1.2 Laboratory equipment

- Microlitre pipettes and tips
- Microcentrifuge tubes
- Electronic pestle
- Bead mill
- Incubator
- Vortexer
- Table centrifuge
- 96 well PCR plates and PCR foils
- 96 well plate cooling block
- 96 well plate centrifuge
- Thermal cycler
- Agarose gel former, microwave, gel electrophoresis chamber, UV illuminator
- Bioanalyzer, chip vortexer
- Qubit Fluorometer
- Access to an Illumina MiSeq desktop sequencer with MiSeq Control Software version 2.2 or later

2.2 Pollen acquisition

Pollen sampling should be performed as described in the respective BEEBOOK chapter. For long term storage, we recommend lyophilisation before freezing at -80 °C.

2.3 Laboratory workflow

2.3.1 DNA Extraction

For the DNA extraction step, we recommend using the Macherey-Nagel (Düren, Germany) NucleoSpin Food Kit and following the supplementary guidelines for pollen samples, but equivalent extraction procedures may also be comparable. The DNA extraction steps are as follows:

1. Take 2 g of pollen and add 4 mL bidest H_2O

2. Homogenise the sample with an electronic pestle
3. Take 200 µL (~50 mg pollen) of the emulsion and grind it in a bead mill
4. Add 400 µL Buffer CF (preheated to 65 °C) and 10µL Proteinase K and mix carefully
5. Incubate at 65°C for 30 min
6. Centrifuge the mixture for 10 min (>10,000 x *g*)
7. Transfer the supernatant into a new microcentrifuge tube and add 1 vol Buffer C4 and 1 vol ethanol
8. Vortex for 30 s
9. Pipette 700 µL mixture onto a NucleoSpin Food Column placed in a Collection Tube
10. Centrifuge for 1 min at 11,000 x *g*
11. Discard the flow-through
12. Repeat steps 9-11
13. Add 400 µL Buffer CQW onto the spin column
14. Centrifuge for 1 min at 11,000 x g
15. Discard the flow-through
16. Add 700 µL Buffer C5 onto the spin column
17. Centrifuge for 1 min at 11,000 x g
18. Discard the flow-through
19. Add 200 µL Buffer C5 onto the spin column
20. Centrifuge for 2 min at 11,000 x g
21. Place the spin column into a new 1.5 mL microcentrifuge tube
22. Add 100 µL Elution Buffer CE (pre-heated to 70 °C) onto the membrane
23. Incubate for 5 min at room temperature (18-25 °C)
24. Centrifuge for 1 min a 11,000 x g
25. Proceed with amplification or keep frozen until further processing

2.3.2 Amplification

This protocol utilises a dual-indexing strategy (Kozich et al. 2013) amplifying the ITS2 region, using the primers ITS-S2F (Chen et al. 2010) and ITS4R (White et al. 1990). The primer sequences are as follows: forward: 5'-AATGATACGGCGACCACCGAGATCTACAC XXXXXXXX CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT-3'; reverse: 5'-CAAGCAGAAGACGGCATACGAGAT XXXXXXXX AGTCAGTCAG CC TCCTCCGCTTATTGATATGC-3', where XXXXXX indicates the variable index sequences (Table 1). The detailed protocol is described below:

1. Sample index combinations should be planned beforehand according to the scheme in Figure 1
2. Prepare 3 x 10 µL reaction mixes for each sample containing (also see PCR sample design 2.3.2.1 below for details):
    - 5 µL 2 x Phusion Master Mix (New England Biolabs, Ipswich, MA, USA) or equivalent
    - 0.33 µM each of the forward and reverse primers (sample-specific combinations of forward and reverse index sequences)
    - 3.34 µL PCR grade water
    - 1 µL DNA template
3. Carry out the PCR with a programme of:
    - 95 °C for 4 min., then
    - 37 cycles of 95 °C for 40 sec.;
    - 49 °C for 40 sec.;
    - 72 °C for 40 sec. and
    - a final extension at 72°C for 5 min.
4. Combine the triplicate PCR reactions of each sample and mix well.

For quality control purposes, successful amplification can be checked on a 1 % agarose gel using 5 µL of the combined PCR product.

2.3.2.1 96-well PCR sample design

Design 1: Well-equipped laboratories with pipetting robots or 96-channel pipettes can directly fill each well with a different sample and generate three replicates of these. This will result in 4 x 3 replicate 96-well plates according to Figure 1 used for amplification. After amplification one can proceed with 2.3.3. Normalisation.

Design 2: For laboratories with little equipment for automated pipetting, the workflow described above is impractical, since manual pipetting in that format is time-intensive and pipetting errors can be easily introduced. To facilitate the process, we recommend to work with all triplicates but only 24 samples on one 96 well plate (Figure 2). This way, 16 PCR plates will be produced, but pipetting effort is minimized. PCR plate labelling is therefore of utter importance, for example with roman numbers, I – XVI to be able to map the samples back to the scheme in Figure 1. The complete workflow is shown schematically in Figure 2 and described in the following:

1. Prepare two PCR master mixes, each containing one forward primer, corresponding to the samples you want to amplify; each master mix contains:
    - 200 µL 2 x Phusion Master Mix (New England Biolabs, Ipswich, MA, USA) or equivalent
    - 13.2 µL forward primer
    - 133.6 µL PCR grade water
2. Place a new PCR plate into a cooling block

3. Distribute 26 µL of the master mixes into row A (Master Mix 1) and F (Master Mix 2)
4. Add 1 µL of the correct reverse primer
5. Add 3 µL of the correct DNA template
6. Using a pipette set to 10 µL, pipette up and down to mix and distribute 10 µL each into the two rows below: from row A into rows B + C; from row F into rows G +H
7. Seal with a foil, spin down briefly
8. Perform PCR
9. Prepare a 1 % agarose gel
10. After PCR, briefly spin down again
11. Lift the foil carefully and combine the triplicate reactions, pipette up and down to mix
12. For gel electrophoresis, add 1 µL of 6x loading buffer into the so far unused rows D + E
13. Add 5 µL PCR product to the loading buffer
14. Briefly spin down
15. Load the gel, add a DNA ladder
16. Run the gel (e.g. 25 min, 120 V)
17. Check under UV illuminator for successful PCR amplification
18. Freeze PCR product until further processing

2.3.3 Normalisation

To ensure more equalised library sizes, DNA amounts in each PCR product are normalised using the SequalPrep™ Normalisation Kit (Invitrogen, Carlsbad, CA, USA). For 384 samples, four normalisation plates are needed. After normalisation, samples from each plate will be combined in 'plate pools' for the following quality control.

Design 1: Pool the samples of all three replicates together by keeping the sample scheme. Transfer 25 µL of PCR products onto the Normalisation plates. Proceed with the normalisation as described below.

Design 2: For normalisation, PCR plates I – IV; V – VIII; IX – XII and XIII – XVI will be combined to Normalisation Plates 1, 2, 3 and 4. The pipetting scheme is as follows:

1. Thaw the PCR plates
2. Briefly spin down
3. Use four Normalisation plates and add 25 µL of PCR product into the wells following this scheme:

- **Normalisation Plate 1: PCR plates I –IV**
- *PCR plate I:* row **A** → row **A**; row **F** → row **B**
- *PCR plate II:* row **A** → row **C**; row **F** → row **D**
- *PCR plate III:* row **A** → row **E**; row **F** → row **F**
- *PCR plate IV:* row **A** → row **G**; row **F** → row **H**

- Repeat analogous for the other three Normalisation Plates
- Proceed with the normalisation as described below.

Design 1 & 2: Continue for both designs with the normalization:

1. Add 25 µL of Binding buffer
2. Mix by pipetting up and down or seal the plate with foil tape, vortex to mix and briefly centrifuge the plate
3. Incubate for 1 hour at room temperature; alternatively leave to incubate overnight
4. Aspirate liquid from wells, do not scrape the well sides
5. Add 50 µL Wash buffer, mix by pipetting up and down
6. Completely aspirate the buffer from wells, you may need to invert and tap the plate on paper towels
7. Add 20 µL of Elution buffer
8. Mix by pipetting up and down or seal the plate with foil tape, vortex and briefly spin down
9. Incubate for 5min at room temperature
10. Combine 5 µL of each sample (plate-wise) in a new microcentrifuge tube, mix well
11. Prepare 1:10 dilutions of each plate pool

2.3.4 Quality control and quantification

Quality control is performed on a Bioanalyzer High Sensitivity DNA Chip (Agilent Technologies, Santa Clara, CA, USA) to ensure that the correct fragment size (peak at approximately 450bp; target plus adapters) has been amplified. Additionally, libraries are quantified using the dsDNA High Sensitivity Assay on the Qubit fluorometer (both Life Technologies GmbH, Darmstadt, Germany) in order to combine the four plate pools equimolarly to the final sequencing library. We recommend preparing three independent concentration measurements per plate pool.

2.3.4.1 Bioanalyzer

1. Prepare a Bioanalyzer Chip according to the protocol
2. Allow all reagents to equilibrate to room temperature
3. If not ready, prepare a gel-dye mix:
4. Add 15 µL of the dye concentrate (blue lid) to a gel matrix vial (red lid)
5. Vortex well and spin down, transfer to spin filter

6. Centrifuge at 2240 x g for 10 min
7. Protect solution from light, store at 4 °C, use within 6 weeks
8. Put a new chip on the chip priming station
9. Pipette 9 µL gel-dye mix into the well marked with a white 'G'
10. Close the chip priming station, with the plunger at position 1mL
11. Press plunger until held by the clip
12. Wait for 60 s then release clip
13. After 5 s slowly pull back the plunger to the 1mL position
14. Open the priming station, pipette 9 mL gel-dye mix in the wells marked with black 'G's
15. Pipette 5 µL marker (green lid) into all sample wells and the ladder wells
16. Pipette 1 µL of ladder (yellow lid) in the well marked with a ladder symbol
17. In each sample well, pipette 1 µL of sample (concentrated and diluted Plate pools) or 1 µL marker (unused wells)
18. Put the chip horizontally in the adapter and vortex for 1 min at 2400 rpm
19. Run the chip within 5 min
20. The samples are of sufficient quality, if the electropherograms show a single peak at approximately 450bp; this peak can be rather wide due to different lengths of the ITS2 region, a minor peak shortly after the lower marker is acceptable and corresponds to left-over primer dimers, which will not interfere with sequencing

2.3.4.2 Quantification

21. Measure concentrations of plate pools with the dsDNA High Sensitivity Assay on the Qubit Fluorometer
22. Mix 1 x n µL Qubit reagent with 199 x n µL Qubit buffer (working solution)
23. For each measurement, mix 180-199 µL working solution with 1-20 µL sample
24. Vortex and incubate at room temperature for 2 min
25. Combine plate pools to final library equimolarly, starting with the least concentrated library of which take 20 µL
26. Quantify the final pool and dilute to 2 nM, if final pool contains less than 2nM proceed without dilution

2.3.5 Sequencing

For library dilution, we follow the Illumina Sample Preparation Guide for a 2 nM library, with some modifications. In order to increase read quality, 5 % PhiX control is added to the sample library. Additionally, the reagent cassette of the sequencing kit (e.g. Illumina MiSeq Reagent Kit v2 2x250bp) is spiked with the custom Read1, Read2 and index primers (for primer sequences, see Table 1).

2.3.5.1 Sample library

1. Remove Buffer HT1 from freezer
2. Prepare a fresh dilution of 0.15 N NaOH (less than a week old)
3. Mix 5µL of the sample library with 5 µL of 0.15 N NaOH
4. Vortex briefly and centrifuge at 280 x g for 1 min
5. Incubate at room temperature for 5 min
6. Add 990 µL Buffer HT1 (10 pM library)
7. Mix 480 µL of 10 pM library and 120 µL Buffer HT1 (8 pM library)

2.3.5.2 PhiX control

1. Thaw PhiX control at room temperature
2. Mix 2 µL 10 nM PhiX control with 3 µL H2O (4 nM PhiX)
3. Add 5 µL 0.15 N NaOH
4. Vortex briefly and centrifuge at 280 x g for 1 min
5. Incubate at room temperature for 5 min
6. Add 990 µL Buffer HT1 (20 pM PhiX)
7. Mix 375 µL of 20 pM PhiX and 225 µL Buffer HT1 (12.5 pM PhiX)
8. Mix 570 µL 8 pM library with 30 µL 12.5 pM PhiX

2.3.5.3 Preparing reagent cassette and loading the sample

1. Remove the reagent cassette from the freezer
2. Place in water bath, do not fill higher than maximum water line
3. Prepare 3 µL each of Read1, Read2 and index primers in new microcentrifuge tubes
4. Remove cassette from water bath, dry with paper towel
5. Invert the cassette several times to mix
6. Inspect wells, make sure all reagents are thawed and there are no precipitates
7. Gently tap the cassette on the bench to remove air bubbles
8. With a 1000 µL pipette tip, break the foils over wells 12-14 and well 17
9. With a 100 µL pipette set to 75 µL, transfer the read and index primers to the following wells of the reagent cartridge: Read1 → Well 12; Index → Well 13; Read2 → Well14, mix well by pipetting up and down
10. Load 600 µL of the spiked library to well 17
11. Load the cassette, PR2 bottle and flow cell as prompted by the instrument
12. Sequence

3 Bioinformatics

3.1 Required software

- up to date Linux or Unix-based OS
- fastq-join, version 1.01.759, (Aronesty 2011), if necessary add location to your system PATH
- usearch, version 8.0.1477, (Edgar 2010), , if necessary add location to your system PATH
- RDPclassifier, version 2.10.2, (Wang et al. 2007), installed to <path_to_RDPTools>

3.2 Classification

3.2.1 Reference database

1. Download reference datasets and training data of Viridiplantae for UTAX or RDPclassifier from http://www.dna-analytics.biozentrum.uni-wuerzburg.de/molecular_biodiversity_group/downloads or https://github.com/iimog/meta-barcoding-dual-indexing.

Alternatively a reference dataset can specifically created and used to train a classifier, if only a limited set of taxa is of interest (not recommended, but faster). Detailed instructions and scripts are available at: https://github.com/iimog/meta-barcoding-dual-indexing. The steps are:

1. Download/create a fasta file containing ITS2 sequences with gene identifier (gi) as header (e.g. from the ITS2-database (Schultz et al. 2006)
2. Assign taxonomy based on the NCBI TaxID (Federhen 2012) of the gi using the supplied scripts
3. Create specific training files for the classifier of choice using the supplied scripts

3.2.2 Preparation and classification of sequencing data

The sequence reads created in step 2.3.5 have to be joined, quality filtered and classified. This can be automatically done with the script *classify_reads.pl* at https://github.com/iimog/meta-barcoding-dual-indexing. For this purpose

1. copy all R1 and R2 fastq files into a single folder
2. copy reference database folder (utax_trained and/or rdp_trained) from 3.2.1 to this folder
3. navigate on the shell to this folder

4.a execute UTAX based classification (fast):

```
perl classify_reads.pl --out results *.fastq\
 --utax-db utax_trained/viridiplantae_all_2014.utax.udb\
--utax-taxtree utax_trained/viridiplantae_all_2014.utax.tax
```

Alternatively you may:

4.b execute RDP based classification together (slow):

```
perl classify_reads.pl --out results *.fastq\
 --noutax\
 --rdp --rdp-jar <path_to_RDPTools>/classifier.jar\
 --rdp-train-propfile rdp_trained/its2.properties
```

This performs the following steps in an automatic procedure:

1. Join the paired reads using fastq-join (Aronesty 2011)
2. Perform Q20 quality filtering and length filtering with usearch (Edgar 2010) and the fastq_filter subcommand (-fastq_truncqual 19, -fastq_minlen 150)

3.a If specified, run usearch (Edgar 2010) with the utax subcommand and training data from step 3.2.1

3.b If specified, run RDPclassifier (Wang et al. 2007) with the training data from step 3.2.1

4. Discard assignments below a bootstrap/rawscore threshold
5. Count the number of reads per taxon of each sample
6. Aggregates the taxon counts for each sample in a common matrix
7. Separates the taxonomic information from the counts

This procedure will end with the following files: a otu_table.txt, a tax_table.txt (one out_table and one tax_table for rdp and utax each) and a mapfile.tsv file for further analysis with phyloseq (McMurdie & Holmes 2013). In addition also the results of the intermediate steps are retained in the subfolders joined, filtered, count and utax or rdp. Those can be used for troubleshooting, archiving or further analyses.

4. Data analysis

4.1 Required software

- up to date R distribution (R Core Team 2014)
- R package: phyloseq (McMurdie & Holmes 2013); https://joey711.github.io/phyloseq

4.2 Prepare sample meta-data

The generated "mapfile.tsv" is already structured in a format that is adequate to import the sample information into R. This is the file where sample meta-information must be deposited. For example continuous vectors like "altitude" or "temperature" or categorical factors as "bee species" or "site" can be used. For this, open the file with your preferred text-editor or spreadsheet application and add columns according to the sampling design. Save the file again in tab-separated format.

4.3 Importing data

The data generated in 3. can be directly imported into R as a phyloseq class object. This allows a variety of analytical procedures and is recommended. However, other software tools handling community datasets may be equally well used for the task of analyses. The following are R scripts, that can be directly used on the console:

```
1.  library(phyloseq)                    # load the package
2.  setwd("<path_to_data>")              # set the folder where data is located
3.  data <- otu_table(read.table("utax_otu_table.txt"), taxa_are_rows=T)
                                         # import community data, replace utax  with
                                         rdp if adequate.
4.  data.tax <- tax_table(as.matrix(read.table("utax_tax_table.txt", fill=T, header=T,
    sep="\t", row.names=1)))            # import taxonomy information of pollen
5.  data.map <- import_qiime_sample_data("mapfile.tsv")    # import sample meta-data
6.  data <- merge_phyloseq(data.otu, data.tax, data.map)   # create phyloseq object
```

Relativize and filter rare taxa below 0.1 %. This is recommended but not necessary.

```
7.  data.rel = transform_sample_counts(data, function(x) x/sum(x))
8.  otu_table(data)[otu_table(data.rel)<0.001]<-0
9.  otu_table(data.rel)[otu_table(data.rel)<0.001]<-0
10. data = prune_taxa(taxa_sums(data)>0, data)
11. data = prune_taxa(taxa_sums(data)>0, data)
```

After completion of the tasks above, the dataset is in a condition where individual analyses can be started. The tutorials at the repository of phyloseq ((McMurdie & Holmes 2013); https://joey711.github.io/phyloseq) provide a good starting point for this.

4.4 Recommended packages for further analysis

Whilst phylseq provides basic tools suited for most purposes, the modularity of R packages allows a variety of more and deeper analyses. It is not possible to discuss all the features here, yet we provide a list some of the major packages relevant for community ecology and pollination studies:

- vegan: comprehensive community ecology package
- picante: phylogenetic diversity indices
- bipartite: interaction network ecology
- edgeR: tests and logFC to investigate differential distributions of taxa between samples

5. Acknowledgements

This work was financially supported by the DFG Collaborative Research Center 1047, Insect Timing.
MJA was further supported by a grant of the German Excellence Initiative to the Graduate School of Life Sciences of the University of Würzburg.
We are grateful to the members of the Departments of Animal Ecology and Tropical Biology; Bioinformatics; and Human Genetics, University of Würzburg, for constructive input on the design of the workflow. We additionally thank the Department of Human Genetics, especially S. Rost, for granting us access to the Illumina MiSeq device.

6. References

Ankenbrand, M.J. et al., 2015. The ITS2 database V -- Twice as much. *Molecular Biology and Evolution*, doi: 10.1093/molbev/msv174.

Aronesty, E., 2011. ea-utils: "Command-line tools for processing biological sequencing data." Available at: http://code.google.com/p/ea-utils.

Beil, M., Horn, H. & Schwabe, A., 2008. Analysis of pollen loads in a wild bee community (Hymenoptera : Apidae) – a method for elucidating habitat use and foraging distances. *Apidologie*, 39, pp.456–467.

Bruni, I. et al., 2015. A DNA barcoding approach to identify plant species in multiflower honey. *Food Chemistry*, 170, pp.308–315.

Carvell, C. et al., 2006. Assessing the value of annual and perennial forage mixtures for bumblebees by direct observation and pollen analysis. *Apidologie*, 37(3), pp.326–340.

Chen, S. et al., 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PloS one*, 5(1), p.e8613.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), pp.2460–2461.

Federhen, S., 2012. The NCBI Taxonomy database. *Nucleic acids research*, 40(Database issue), pp.D136–43.

Galimberti, A. et al., 2014. A DNA barcoding approach to characterize pollen collected by honeybees. *PloS one*, 9(10), p.e109363.

Keller, A. et al., 2015. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. *Plant Biology*, 17(2), pp.558–566.

Koetschan, C. et al., 2010. The ITS2 Database III--sequences and structures for phylogeny. *Nucleic Acids Research*, 38(Database issue), pp.D275–D279.

Köppler, K., Vorwohl, G. & Koeniger, N., 2007. Comparison of pollen spectra collected by four different subspecies of the honey bee Apis mellifera. *Apidologie*, 38, pp.341–353.

Kozich, J.J. et al., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, 79(17), pp.5112–5120.

Kraaijeveld, K. et al., 2015. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Molecular Ecology Resources*, 15, pp.8–16.

McMurdie, P.J. & Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one*, 8(4), p.e61217.

Mullins, J. & Emberlin, J., 1997. Sampling pollens. *Journal of Aerosol Science*, 28(3), pp.365–370.

R Core Team, 2014. R: A language and environment for statistical computing.

Richardson, R.T. et al., 2015. Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. *Applications in Plant Sciences*, 3(1), p.1400066.

Schultz, J. et al., 2006. The internal transcribed spacer 2 database--a web server for (not only) low level phylogenetic analyses. *Nucleic acids research*, 34(Web Server issue), pp.W704–7.

Sickel, W. et al., 2015. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology*, 15(20).

Valentini, A., Miquel, C. & Taberlet, P., 2010. DNA barcoding for honey biodiversity. *Diversity*, 2, pp.610–617.

Wang, Q. et al., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73(16), pp.5261–5267.

White, T.J. et al., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. A. Innis et al., eds. *PCR Protocols: A Guide to Methods and Applications*. New York: Academic Press, pp. 315–322.

Williams, N.M. & Kremen, C., 2007. Resource distributions among habitats determine solitary bee offspring production in a mosaic landscape. *Ecological Applications*, 17, pp.910–921.

Table 1: Primer Sequences with indexes SA501 – SB712 (adapted from Kozich et al. 2013); index sequences indicated in bold

Forward	
Name	**Sequence**
SA501	AATGATACGGCGACCACCGAGATCTACAC **ATCGTACG** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SA502	AATGATACGGCGACCACCGAGATCTACAC **ACTATCTG** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SA503	AATGATACGGCGACCACCGAGATCTACAC **TAGCGAGT** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SA504	AATGATACGGCGACCACCGAGATCTACAC **CTGCGTGT** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SA505	AATGATACGGCGACCACCGAGATCTACAC **TCATCGAG** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SA506	AATGATACGGCGACCACCGAGATCTACAC **CGTGAGTG** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SA507	AATGATACGGCGACCACCGAGATCTACAC **GGATATCT** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SA508	AATGATACGGCGACCACCGAGATCTACAC **GACACCGT** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SB501	AATGATACGGCGACCACCGAGATCTACAC **CTACTATA** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SB502	AATGATACGGCGACCACCGAGATCTACAC **CGTTACTA** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SB503	AATGATACGGCGACCACCGAGATCTACAC **AGAGTCAC** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SB504	AATGATACGGCGACCACCGAGATCTACAC **TACGAGAC** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SB505	AATGATACGGCGACCACCGAGATCTACAC **ACGTCTCG** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SB506	AATGATACGGCGACCACCGAGATCTACAC **TCGACGAG** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SB507	AATGATACGGCGACCACCGAGATCTACAC **GATCGTGT** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
SB508	AATGATACGGCGACCACCGAGATCTACAC **GTCAGATA** CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
Reverse	
Name	**Sequence**
SA701	CAAGCAGAAGACGGCATACGAGAT **AACTCTCG** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SA702	CAAGCAGAAGACGGCATACGAGAT **ACTATGTC** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SA703	CAAGCAGAAGACGGCATACGAGAT **AGTAGCGT** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SA704	CAAGCAGAAGACGGCATACGAGAT **CAGTGAGT** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SA705	CAAGCAGAAGACGGCATACGAGAT **CGTACTCA** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SA706	CAAGCAGAAGACGGCATACGAGAT **CTACGCAG** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SA707	CAAGCAGAAGACGGCATACGAGAT **GGAGACTA** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SA708	CAAGCAGAAGACGGCATACGAGAT **GTCGCTCG** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SA709	CAAGCAGAAGACGGCATACGAGAT **GTCGTAGT** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SA710	CAAGCAGAAGACGGCATACGAGAT **TAGCAGAC** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SA711	CAAGCAGAAGACGGCATACGAGAT **TCATAGAC** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SA712	CAAGCAGAAGACGGCATACGAGAT **TCGCTATA** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SB701	CAAGCAGAAGACGGCATACGAGAT **AAGTCGAG** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SB702	CAAGCAGAAGACGGCATACGAGAT **ATACTTCG** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SB703	CAAGCAGAAGACGGCATACGAGAT **AGCTGCTA** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SB704	CAAGCAGAAGACGGCATACGAGAT **CATAGAGA** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SB705	CAAGCAGAAGACGGCATACGAGAT **CGTAGATC** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SB706	CAAGCAGAAGACGGCATACGAGAT **CTCGTTAC** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SB707	CAAGCAGAAGACGGCATACGAGAT **GCGCACGT** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SB708	CAAGCAGAAGACGGCATACGAGAT **GGTACTAT** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SB709	CAAGCAGAAGACGGCATACGAGAT **GTATACGC** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SB710	CAAGCAGAAGACGGCATACGAGAT **TACGAGCA** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SB711	CAAGCAGAAGACGGCATACGAGAT **TCAGCGTT** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
SB712	CAAGCAGAAGACGGCATACGAGAT **TCGCTACG** AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
Index and Read	
Name	**Sequence**
Read1	CCTGGTGCTG GT ATGCGATACTTGGTGTGAAT
Read2	AGTCAGTCAG CC TCCTCCGCTTATTGATATGC
Index	GCATATCAATAAGCGGAGGA GG CTGACTGACT

Figure 1 Planning scheme for samples and the corresponding index-combinations. Roman numbers indicate PCR plate numbers, bold Arabian numbers on 96 well plates indicate Normalisation plate number.

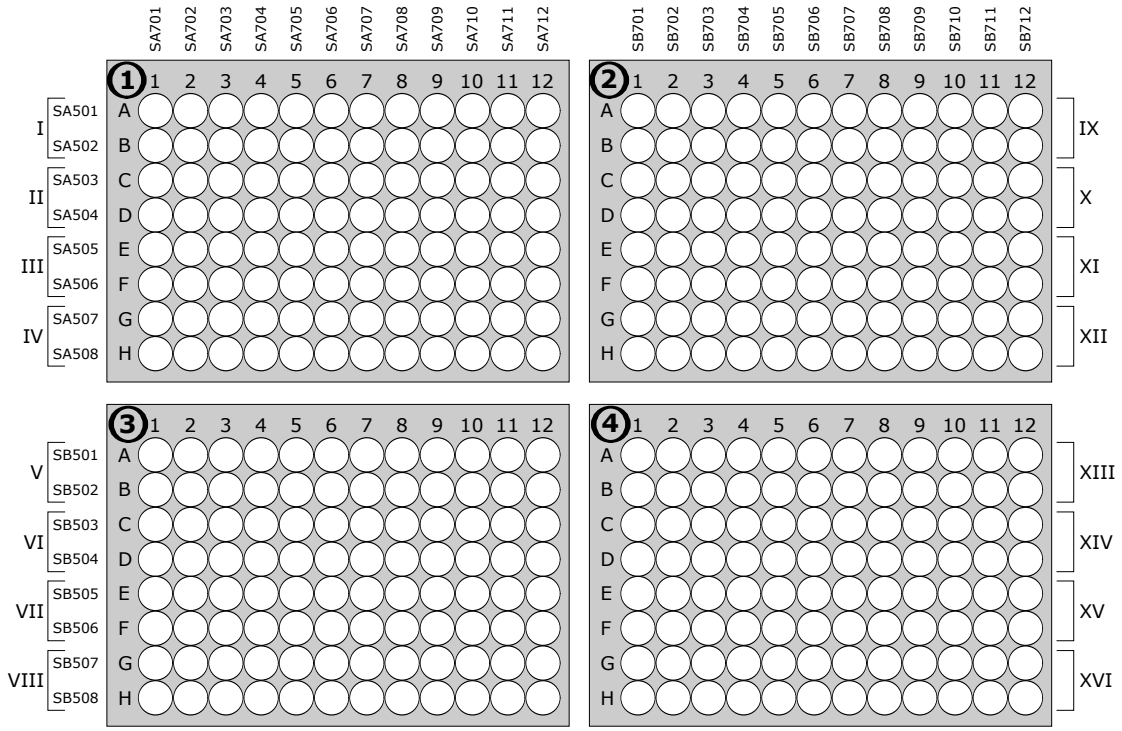
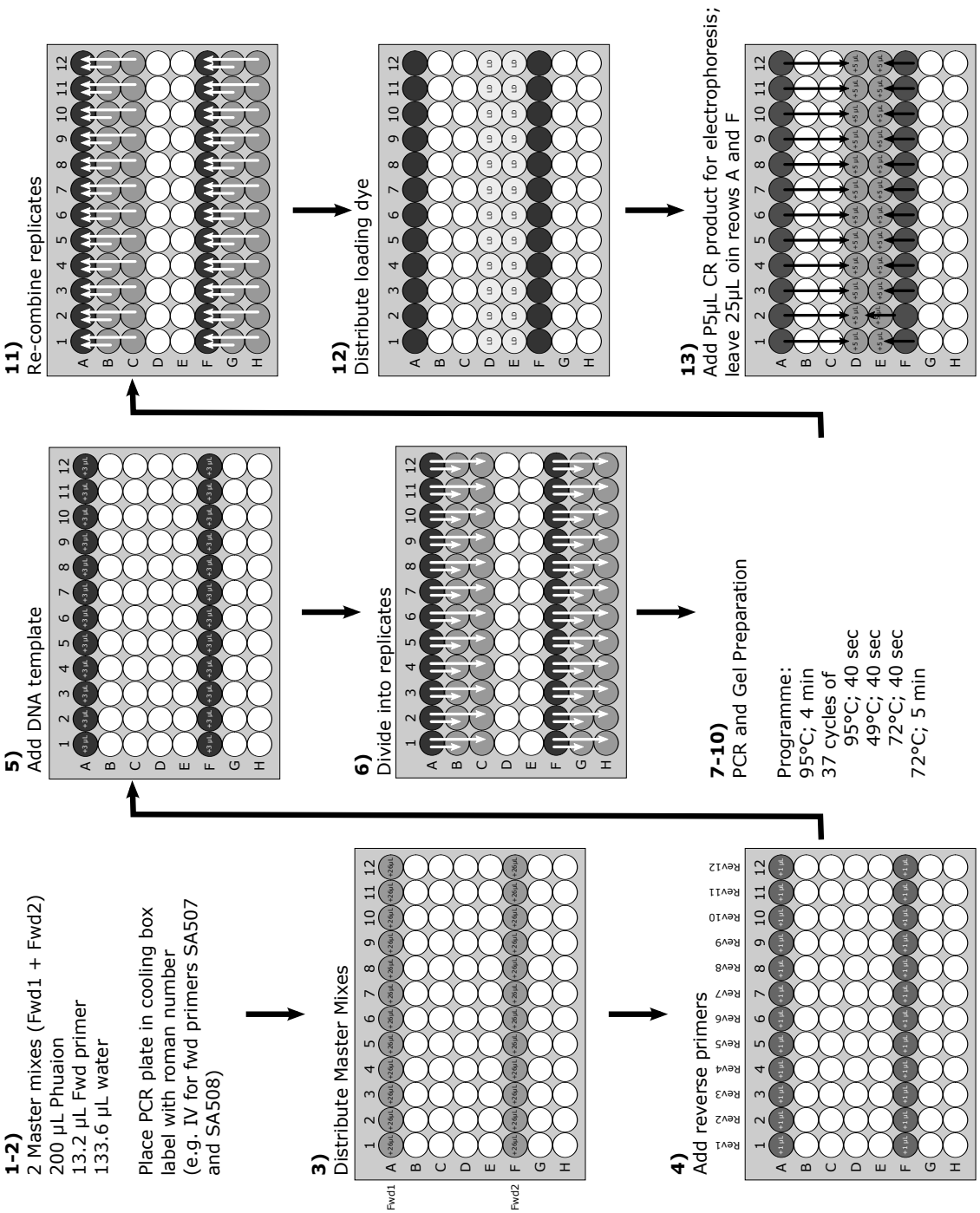


Figure 2 Detailed workflow (schematic), suitable for laboratories with limited access to equipment for automated pipetting. Bold numbers indicated step number of Design 2 in sub-chapter 2.3.2.1



CHAPTER 3

Bacterial communities

P.4 Reptiles as Reservoirs of Bacterial Infections: Real Threat or Methodological Bias?

Authors:

Giulia Zancolli, Dieter Mahsberg, Wiebke Sickel*, Alexander Keller

*Co-author

published: 2015

Journal: *Microbial Ecology*

Volume: 70, Pages: 759-584

Microb Ecol
DOI 10.1007/s00248-015-0618-3

NOTES AND SHORT COMMUNICATIONS

Reptiles as Reservoirs of Bacterial Infections: Real Threat or Methodological Bias?

Giulia Zancolli[1,3] · **Dieter Mahsberg**[1] · **Wiebke Sickel**[1,2] · **Alexander Keller**[1,2]

Received: 25 February 2015 / Accepted: 16 April 2015

Abstract Bacterial infections secondary to snakebites and human pathogens (e.g., *Salmonella*) have been linked to the oral microbiota of snakes and pet reptiles. Based on culture-dependent studies, it is speculated that snakes' oral microbiota reflects the fecal flora of their ingested preys. However, cultured-based techniques have been shown to be limited as they fail to identify unculturable microorganisms which represent the vast majority of the microbial diversity. Here, we used culture-independent high-throughput sequencing to identify reptile-associated pathogens and to characterize the oral microbial community of five snakes, one gecko, and two terrapins. Few potential human pathogens were detected at extremely low frequencies. Moreover, bacterial taxa represented in the snake's oral cavity bore little resemblance to their preys' fecal microbiota. Overall, we found distinct, highly diverse microbial communities with consistent, species-specific patterns contrary to previous culture-based studies. Our study does not support the widely held assumption that reptiles' oral cavity acts as pathogen reservoir and provides important insights for future research.

Keywords 16S rDNA · Oral microbiota · Snakebite · Wound infection · Zoonosis

✉ Giulia Zancolli
giulia.zancolli@gmail.com

[1] Department of Animal Ecology and Tropical Biology, Biocentre, University of Würzburg, Würzburg, Germany

[2] DNA Analytics Core Facility, Biocentre, University of Würzburg, Würzburg, Germany

[3] School of Biological Sciences, Environment Centre Wales, Bangor University, Bangor LL57 2UW, UK

Short Note

Bacterial infections are common complications of wounds secondary to animal bites, and it is established that pathogenic microorganisms recovered from bite wounds reflect the oral flora of the biting animal [1]. It is also determined that the oral flora of the biting animal is influenced by the microbiome of their ingested prey and other foods [1]. Thus, snakes' oral microbiota has been suggested to be fecal in nature, as prey animals often defecate upon being ingested [2]. Microbes frequently identified in the oral cavity of snakes have included *Clostridium* species, pathogenic and non-pathogenic Enterobacteriaceae such as *Salmonella*, *Morganella morganii*, *Proteus*, *Providencia*, and *Escherichia coli* [3–6]. Zoonoses have also been linked to pet reptiles [7]. Nonetheless, those studies were based on culture-dependent techniques, which restrict microbial identification to those bacteria able to grow in cultures, thus providing a biased picture of microbial diversity. However, advances in sequencing technologies have dramatically expanded our ability to profile bacterial communities, revealing an astonishing majority of unculturable microorganisms [8].

In this study, we characterized the oral microbiota of common pet reptiles by means of 16S amplicon sequencing. In particular, we aimed to (i) identify potential pathogens causing bite wound infections and bacteria previously cultured from snakes' mouth cavities and (ii) explore the similarity between the oral microbiota of snakes and the fecal flora of their prey by comparing our results with recently published data from mouse stool samples [9, 10]. We used multiple datasets from different mouse strains [9, 10] to account for potential variability between the fecal microbiota of the mice used to feed the snakes in the present study and that from literature.

We collected swab samples from the oral cavity of the following reptiles: four royal pythons (*Python regius*), one

Published online: 30 April 2015

Table 1 Relative abundances of important bacteria from the oral cavity of captive reptiles and the substrate of the vivaria. Sequences were blasted to search for bacterial taxa identified in past culture-based studies (references in text) isolated from the mouth of snakes and wound infections secondary to snakebites and listed herein with GenBank sequence identification numbers (GI), BLAST percentage of identity, and assigned scientific name. Taxonomic assignment of the most abundant OTUs detected in this study are also reported

	GI	Percentage identity	Scientific name	Substrate	Python regius 1	Python regius 2	Python regius 3	Python regius 4	Acrantophis dumerili	Eublepharis macularius	Trachemys s. scripta 1	Trachemys s. scripta 2
Bacteria known from literature												
Achromobacter denitrificans	Not detected											
Acinetobacter spp.	343206383	100	*Acinetobacter johnsonii*	1.821	0.064	0.117	0.074	0.036	0.553	0.097	13.323	3.532
	631252876	99.21	*Acinetobacter radioresistens*	0.605	0.008	0.125		0.020				0.159
Acinetobacter calcoaceticus	507148019	100	*Acinetobacter calcoaceticus* PHEA-2	0.253				0.008	0.059		0.845	0.072
Acinetobacter lwoffii	631252148	97.63	*Acinetobacter lwoffii*	0.007	0.003				0.005			
Actinomyces spp.	636560534	95.65	*Actinomyces naturae*						0.021	0.006		
Aeromonas spp.	507147982	100	*Aeromonas veronii* B565					0.016				0.338
Arcobacter butzleri	Not detected											
Bacillus spp.	444439662	99.6	*Bacillus pumilus* SAFR-032						0.027	0.006		
	485099110	100	*Bacillus thuringiensis* Bt407	0.239	0.005		0.003		0.032			
Bacteroides fragilis	444439524	99.6	*Bacteroides fragilis* YCH46	0.183								
Bacteroides spp.	631251741	99.6	*Bacteroides nordii*	1.744								
	631251997	100	*Bacteroides sartorii*	0.049						0.084		
Bordetella avium	Not detected											
Bordetella trematum	Not detected											
Bulkolderia spp.	Not detected											
Cedecea neteri	Not detected											
Cetobacterium somerae	Not detected											
Chryseobacterium indologenes	Not detected											
Chryseobacterium massiliae	Not detected											
Chryseobacterium meningosepticum	Not detected											
Citrobacter diversus	Not detected											
Citrobacter freundii	631253147	100	*Citrobacter freundii*	0.555		0.016		0.004		0.006	0.731	0.309
Clostridium spp.	219846899	99.6	*Clostridium disporicum*									0.106
	219857201	94.47	*Clostridium straminisolvens*									0.285
	310975204	98.02	*Clostridium xylanolyticum*	0.274				0.004				0.005
Corynebacterium spp.	265678646	100	*Corynebacterium appendicis*					0.024	0.122			
	343201299	100	*Corynebacterium freneyi*		0.176	0.449	0.071	0.376	0.016		0.219	0.034
	343201411	100	*Corynebacterium pseudodiphtheriticum*					0.012				
	219846842	100	*Corynebacterium riegelii*					0.012			0.609	
	265678670	100	*Corynebacterium tuberculostearicum*		0.005			0.008	0.021	0.071		0.058
Diphtheroid bacillus	Not detected											
Edwardsiella tarda	Not detected											
Enterobacter spp.	Not detected											

Table 1 (continued)

	GI	Percentage identity	Scientific name	Substrate	Python regius 1	Python regius 2	Python regius 3	Python regius 4	Acrantophis dumerili	Eublepharis macularius	Trachemys s. scripta 1	Trachemys s. scripta 2
Enterococcus spp.	631252735	100	*Enterococcus pallens*	1.737	0.028	0.032	0.042		0.271	0.097		
Escherichia spp.	444439587	100	*Escherichia fergusonii* ATCC 35469	3.270					0.032	0.104		0.087
Flavobacterium spp.	343205901	100	*Flavobacterium anhuiense*	0.141					0.021			
	631253066	98.02	*Flavobacterium caeni*			0.036						
	631251641	100	*Flavobacterium saccharophilum*	0.007	0.003	0.077	0.005					0.010
	566085077	100	*Flavobacterium* sp. ARSA-19									0.846
	566085220	98.42	*Flavobacterium* sp. Ruye-71			0.004						
Fusobacterium nucleatum	Not detected											
Fusobacterium varium	Not detected											
Hafnia alvei	Not detected											
Klebsiella spp.	Not detected											
Lactoccus garvieae	Not detected											
Micrococcus spp.	444439747	98.81	*Micrococcus luteus*	0.422	0.008					0.006		0.005
Moraxella spp.	265678369	97.63	*Moraxella caviae*								0.983	0.208
Mycobacterium ulcerans	631252194	100	*Moraxella osloensis*	0.070			0.011					0.029
Morganella morganii	Not detected											
Paracoccus spp.	343202428	100	*Paracoccus solventivorans*	3.481	0.046			0.016	0.474	0.026		0.087
Pasteurella multocida	Not detected											
Pasteurella pneumotropica	Not detected											
Prevotella loescheii	Not detected											
Prevotella oralis	Not detected											
Proteus mirabilis	444439583	100	*Proteus mirabilis* HI4320	0.661						0.091		
Proteus spp.	Not detected											
Providencia spp.	559795321	100	*Providencia burhodogranariea* DSM 19968	0.415								
Pseudomonas aeruginosa	444439513	100	*Pseudomonas aeruginosa* PAO1	0.415			0.252	0.008	8.392			
Pseudomonas spp.	444439519	100	*Pseudomonas brassicacearum* NFM421	0.731	0.003	0.016	0.003			0.006	0.496	0.686
	631252994	100	*Pseudomonas japonica*	0.155	0.010	0.004	0.026				0.991	2.653
	631252999	98.42	*Pseudomonas jinjuensis*									0.348
	631253017	100	*Pseudomonas luteola*	0.865								
Rhizobium radiobacter	Not detected											
Rhodococcus spp.	636560631	99.6	*Rhodococcus equi*	0.338								
Salmonella spp.	Not detected											
Sarcina spp.	Not detected											
Serratia spp.	507148020	100	*Serratia plymuthica*	0.035	0.003							0.005
Shewanella putrefaciens	Not detected											
Shigella spp.	Not detected											
Sphingobacterium paucimobilis	Not detected											

Table 1 (continued)

	GI	Percentage identity	Scientific name	Substrate	Python regius 1	Python regius 2	Python regius 3	Python regius 4	Acrantophis dumerili	Eublepharis macularius	Trachemys s. scripta 1	Trachemys s. scripta 2
Sphingobacterium spiritivorum	631252509	99.21	*Sphingobacterium spiritivorum*								0.690	
Staphylococcus spp.	485099103	100	*Staphylococcus warneri* SG1	0.928	0.089		0.039	0.069	0.165	0.155	0.211	0.271
Stenotrophomonas maltophilia	343200117	100	*Stenotrophomonas maltophilia*		0.049	0.028	0.105	0.012	0.117		1.795	0.077
Streptococcus spp.	Not detected											
Tatumella ptyseos	Not detected											
Tsukamurella spp.	444304034	98.82	*Tsukamurella paurometabola* DSM 20162					0.580				
Vagococcus spp.	636560558	84.43	*Vagococcus penaei*	0.028								
Xanthomonas spp.	444439622	100	*Xanthomonas axonopodis* pv. *citri* str. 306	1.392			0.003			0.065	0.049	
Yersinia enterocolitica	Not detected											
Most abundant OTUs found in this study												
OTU_237	444304263	100	*Polynucleobacter necessarius* subsp. *asymbioticus*	28.646	1.125	1.761	1.485	1.313	8.509	3.880	5.045	0.416
OTU_26	219846259	99.21	*Acidocella aminolytica*	9.077	0.358	0.295	0.315	0.158	0.878	0.589	0.065	0.164
OTU_1	343200563	90.91	*Terrimonas lutea*	0.056	35.175	25.110	50.636	35.135	0.011	0.006		0.111
OTU_2	566085498	92.09	*Chryseobacterium* sp. THMBM1	0.105	34.203	35.872	36.909	36.966	0.021	0.013		0.135
OTU_4	566085064	98.42	*Chryseobacterium* sp. THG 15	0.956	8.107			0.004	61.979	0.006		0.005
OTU_23	265678797	97.63	*Hydrocarboniphaga effusa*	0.429	0.015				10.148	0.136		
OTU_9	631253014	92.09	*Chryseobacterium marinum*	0.028						41.780		0.005
OTU_10	636560426	90.51	*Niastella populi*	0.007	0.003					39.662		0.010
OTU_17	343201778	97.63	*Chryseobacterium pallidum*	0.007	0.003			0.020	0.005		17.141	16.858
OTU_8	253680759	95.26	*Collimonas fungivorans*	0.014	0.005						9.773	15.650

Dumeril's boa (*Acrantophis dumerili*), one leopard gecko (*Eublepharis macularius*), two yellow-bellied sliders (*Trachemys scripta scripta*), and one pooled sample from the substrate of the vivaria. Except for the terrapins, animals were kept within the same facility but in different cages. The snakes were kept on wood fiber litter, the geckos on a mixture of sand and clay. They were all in good health conditions, and samples were taken 2 weeks after the last feeding (except for the terrapins which were fed daily). For each sample, the V4 region of the 16S rDNA gene was sequenced on an Illumina MiSeq using v2 2×250 bp chemistry, strictly accordingly to the protocol by Kozich et al. [11]. Sequences were quality-filtered (>Q20, chimera-checked, length filtered) and clustered into operational taxonomic units (OTUs) using the USEARCH pipeline [12]. For comparisons, OTU abundances were relativized by sequencing depth for each sample. A list of the bacterial taxa of interest was compiled from the literature (see Table 1) and scanned through our data with BLAST [13]. Raw sequences are accessible at the EBI database (http://www.ebi.ac.uk) under project number PRJEB6675.

We obtained a mean of 23,120 reads per sample after quality-filtering and 345 OTUs in total. Potential pathogens and known cultivable bacteria were either absent or rare and mainly isolated from the vivarium substrate (Table 1). Even though rare microbes can potentially lead to problematic infections, the overall frequencies were extremely lower than in previous culture-based studies. Oral samples were dominated by Bacteroidetes (83 %) and Proteobacteria (15 %), and were markedly different between reptile species suggesting a species-specific pattern (Fig. 1). The four pythons had similar community composition dominated by Chitinophagaceae (52 %) and Flavobacteriaceae (36 %) with up to 76 % of the reads assigned to two OTUs. In the boa, members of the Weeksellaceae (62 %), Sinobacteraceae (13 %), Pseudomonadaceae (8 %), and Methylophilaceae (7 %) were the most abundant, with 62 % of the reads assigned to only one OTU. The snake microbial communities were clearly distinct from the fecal microbiota of mice typically dominated by Firmicutes, Bacteroidetes, and Verrucomicrobia, and the most common fecal bacteria (i.e., *Bacteroides*, *Alistipes*, and *Prevotella*) were absent in the oral samples. Similar to the snakes, the gecko's oral microbiota was dominated by Weeksellaceae (41 %) and Chitinophagaceae (39 %) but also Helicobacteraceae and Rhizobiaceae, albeit in lesser abundance (6 %), and the majority of reads was assigned to two OTUs (Table 1). The two terrapins were similar in composition and showed more diverse and evenly distributed communities. The microbial diversity of the substrate was the most varied with one OTU representing 29 % of all reads.

Zoonoses (e.g., salmonellosis) and infections secondary to snakebites have been linked to reptile-associated microorganisms [3, 6, 7, 14, 15]. Contrary to previous culture-based studies, our high-throughput sequencing of the oral cavity of captive snakes (and other common pet reptiles) revealed an extremely low incidence of potential human pathogens which were mostly recovered from the vivarium substrate and not from the oral samples. We cannot rule out that the bacteria recovered from bite wounds can come from the biting animal; yet, our data show that they are found in the environment and not in the mouth cavities. Importantly, the oral communities were well structured, with predominant taxa not found in the substrate, suggesting host-specificity.

It has been suggested that the oral microbiota of snakes may reflect the fecal flora of their prey, which frequently

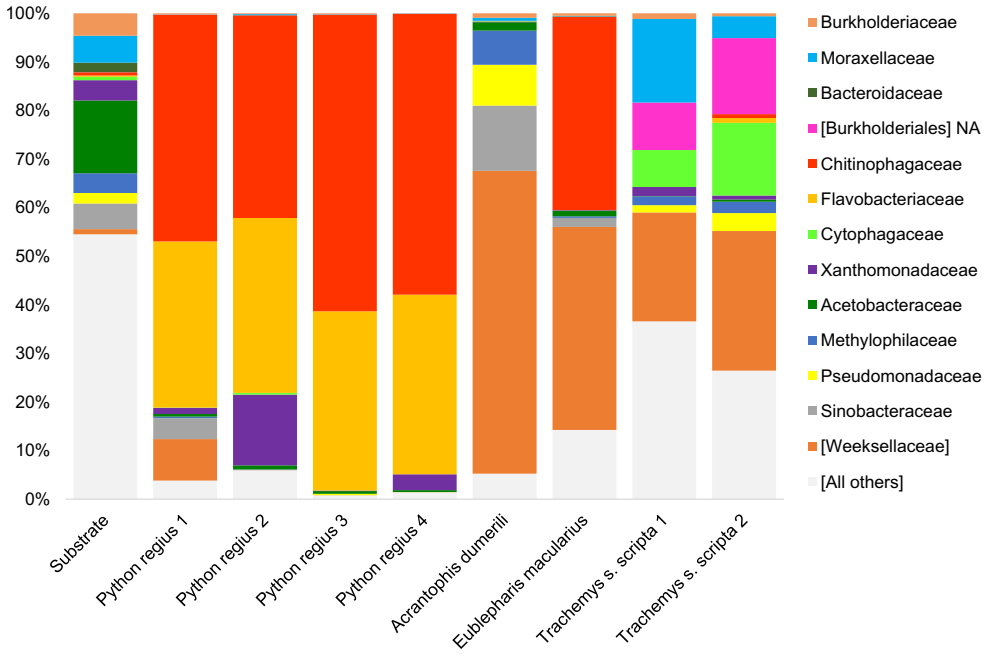


Fig. 1 Relative abundances of bacterial families from the oral cavity of captive reptiles and the vivarium substrate. The most represented bacterial families are displayed in *different colors*; rare taxa are grouped as "*[All others]*" and displayed in *white*

defecate upon being ingested [2]. However, the microbial communities in the mouth cavities of our snakes were markedly distinct from those identified in mouse stool samples using similar molecular techniques [9, 10]. In particular, the most abundant genera found in mouse stool were completely absent. It is worth noting that our snakes were fed upon dead mice.

Other studies also contradict the suggestion that snake oral floras are a result of fecal contamination from prey. For instance, no differences were observed between wild snakes (which feed on live prey) and captive snakes (fed with frozen mice) [4, 16]. Similarly, Goldstein et al. [3] did not find obvious differences between the oral microbiota of adult garter snakes and neonates before their first meal. Ecological communities are dynamic systems, which respond to different environmental conditions by modifying their species composition and population [17]. Oral microbial communities in particular are constantly altered by the host activities, e.g., eating and tongue movements inside and outside the buccal cavity. Saliva and other oral secretions, including snake venom [18], possess antimicrobial activities, and the microbial commensals in the oral cavity can limit the growth of invasive species, including pathogens [19]. It is thus unlikely that snakes' oral microbiota reflects that of their preys' feces, and our data strongly support this.

Our data also suggest that the vivarium substrate is not a major source of oral microbiota, as most of the taxa recovered from the oral samples were restricted to this habitat. Interestingly, the pythons, albeit in separate cages, showed a strongly uniform pattern, which however differed largely from the boa despite being kept within the same room and fed with the same food. Overall, we observed large and consistent differences between reptile taxonomic groups suggesting a role of host phylogeny and/or diet (snakes: carnivores, gecko: insectivore, terrapins: omnivores) in shaping the host-associated microbial community.

This pilot study offers relevant insights for future investigations for better understanding to which extent host phylogeny and diet play a role in determining the assembly of the oral microbiota using reptiles as a model taxon. Furthermore, extending the analysis to venomous snakes will be of uttermost interest considering that venom can exert antibacterial activity [18]. Metagenomic sequencing could also elucidate the biological functions of the oral microbiota, especially in snakes where bacteria may play a central role in producing enzymes and biomolecules for aid in prey digestion.

Acknowledgments This work was supported by the University of Würzburg, Land Bavaria, and a Career Development Fellowship by the Graduate School of Life Sciences, University of Würzburg (to GZ). We thank Martin Llewellyn and Wolfgang Wüster for the constructive comments on an earlier version of the manuscript.

References

1. Abrahamian FM, Goldstein EJC (2011) Microbiology of animal bite wound infections. Clin Microbiol Rev 24:231–246
2. Goldstein EJC, Citron DM, Gonzalez H, Russell FE, Finegold SM (1979) Bacteriology of rattlesnake venom and implications for therapy. J Infect Dis 140:818–821
3. Goldstein EJC, Agyare EO, Vagvolgyi AE, Halpern M (1981) Aerobic bacterial oral flora of garter snakes: development of normal flora and pathogenic potential for snakes and humans. J Clin Microbiol 13:954–956
4. Blaylock RSM (2001) Normal oral bacterial flora from some southern African snakes. Onderstepoort J Vet Res 68:175–182
5. Ferreira Junior RS, Siqueira AK, Campagner MV, Salerno T, Soares TCS, Lucheis SB, Paes AC, Barraviera B (2009) Comparison of wildlife and captivity rattlesnakes (Crotalus durissus terrificus) microbiota. Pesq Vet Bras 29:999–1003
6. Lam KK, Crow P, Ng KHL, Shek KC, Fung HT, Ades G, Grioni A, Tan KS, Yip KT, Lung DC, Que TL, Lam TSK, Simpson ID, Tsui KL, Kam CW (2010) A cross-sectional survey of snake oral bacterial flora from Hong Kong, SAR, China. Emerg Med J 28:107–114
7. Warwick C, Lambiris AJL, Westwood D, Steedman C (2001) Reptile-related salmonellosis. J R Soc Med 94:124–126
8. Rappé MS, Giovannoni SJ (2003) The uncultured microbial majority. Annu Rev Microbiol 57:369–394
9. Gu S, Chen D, Zhang JN, Lv X, Wang K, Duan LP, Nie Y, Wu XL (2013) Bacterial community mapping of the mouse gastrointestinal tract. PLoS One 8:e74957
10. Krych L, Hansen CHF, van den Berg FWJ, Nielsen DS (2013) Quantitatively different, yet qualitatively alike: a meta-analysis of the mouse core gut microbiome with a view towards the human gut microbiome. PLoS One 8:e62578
11. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79:5112–5120
12. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
14. Arroyo O, Bolaños R, Muñoz G (1980) The bacterial flora of venoms and mouth cavities of Costa Rican snakes. Bull Pan Am Health Organ 14:280–285
15. Garg A, Sujatha S, Garg J, Acharya NS, Chandra Parija S (2009) Wound infections secondary to snakebite. J Infect Dev Ctries 3:221–223
16. Fonseca MG, Moreira WMQ, Cunha KC, Ribeiro ACMG, Almeida MTG (2009) Oral microbiota of Brazilian captive snakes. J Venom Anim Toxins incl Trop Dis 15:54–60
17. Morin PJ (1999) Community ecology. Blackwell Science, Malden
18. Shebl RI, Mohamed AF, Ali AE, Amin MA (2012) Antimicrobial profile of selected snake venoms and their associated enzymatic activities. Br Microbiol Res J 2:251–263
19. Avila M, Ojcius DM, Yilmaz Ö (2009) The oral microbiota: living with a permanent guest. DNA Cell Biol 28:405–411

P.5 Bacterial Diversity and Community Structure in Two Bornean *Nepenthes* Species with Differences in Nitrogen Acquisition Strategies

Authors:

Wiebke Sickel*, T. Ulmar Grafe, Ivonne Meuche, Ingolf Steffan-Dewenter, Alexander Keller
*First author

published: 2016

Journal: *Microbial Ecology*

Volume: 71, Pages: 938-953

Microb Ecol
DOI 10.1007/s00248-015-0723-3

PLANT MICROBE INTERACTIONS

Bacterial Diversity and Community Structure in Two Bornean *Nepenthes* Species with Differences in Nitrogen Acquisition Strategies

Wiebke Sickel[1] · **T. Ulmar Grafe**[2] · **Ivonne Meuche**[2] · **Ingolf Steffan-Dewenter**[1] · **Alexander Keller**[1]

Received: 20 December 2014 / Accepted: 21 December 2015

Abstract Carnivorous plants of the genus *Nepenthes* have been studied for over a century, but surprisingly little is known about associations with microorganisms. The two species *Nepenthes rafflesiana* and *Nepenthes hemsleyana* differ in their pitcher-mediated nutrient sources, sequestering nitrogen from arthropod prey and arthropods as well as bat faeces, respectively. We expected bacterial communities living in the pitchers to resemble this diet difference. Samples were taken from different parts of the pitchers (leaf, peristome, inside, outside, digestive fluid) of both species. Bacterial communities were determined using culture-independent high-throughput amplicon sequencing. Bacterial richness and community structure were similar in leaves, peristomes, inside and outside walls of both plant species. Regarding digestive fluids, bacterial richness was higher in *N. hemsleyana* than in *N. rafflesiana*. Additionally, digestive fluid communities were highly variable in structure, with strain-specific differences in community composition between replicates. Acidophilic taxa were mostly of low abundance, except the genus *Acidocella*, which strikingly reached extremely high levels in two *N. rafflesiana* fluids. In *N. hemsleyana* fluid, some taxa classified as vertebrate gut symbionts as well as saprophytes were enriched compared to *N. rafflesiana*, with saprophytes constituting potential competitors for nutrients. The high variation in community structure might be caused by a number of biotic and abiotic factors. Nitrogen-fixing bacteria were present in both study species, which might provide essential nutrients to the plant at times of low prey capture and/or rare encounters with bats.

Keywords *Nepenthes* · Carnivorous plants · Next-generation sequencing · 16s rDNA · Plant-microbe interactions

Introduction

Carnivorous plants of the species-rich genus *Nepenthes* (Nepenthaceae, Caryophyllales) grow in nutrient-depleted, acidic soils across South East Asia [1] with centres of diversity in Borneo, Sumatra and the Philippines [2, 3]. These tropical plants derive nutrients, e.g. nitrogen, usually from arthropod prey, which they catch and digest in highly adapted pitcher traps [1]. The traps consist of the pitcher lid, the peristome, i.e. the rim of the pitcher, and the digestive zone, filled with an acidic and often viscous fluid [4, 5]. The inner pitcher wall above the digestive fluid is commonly referred to as the waxy zone, as it is often covered with epicuticular wax [6–8]. The different parts have distinct functions in prey attraction, capture and retention, for example, extra-floral nectar production in the lid and peristome [9, 10] or anisotropic orientation of the inner walls together with easily detachable wax crystals which clog insect adhesive pads [6, 11]. The fluid is involved in prey retention due to viscoelasticity in some *Nepenthes* species [5, 12] and prey digestion, which involves low pH and secretion of digestive enzymes [13, 14]. Prey introduction has been found to trigger fluid acidification [14], digestive enzyme expression [15] and nutrient uptake via the activation of ammonium transporter genes [4]. Fluid acidification is thought to be

Electronic supplementary material The online version of this article (doi:10.1007/s00248-015-0723-3) contains supplementary material, which is available to authorized users.

✉ Alexander Keller
a.keller@biozentrum.uni-wuerzburg.de

[1] Department of Animal Ecology and Tropical Biology, Biocenter, University of Würzburg, 97074 Würzburg, Germany

[2] Faculty of Science, University Brunei Darussalam, Tungku Link, Gadong, BE 1410, Brunei

Published online: 20 January 2016

induced by ammonium, since introduction of ammonium-containing substances activates proton secretion [13] as well as the expression of plasma-membrane H+-ATPase [14]. However, there seem to exist different mechanisms in different *Nepenthes* species, especially for pH decrease in pitchers, as demonstrated for *Nepenthes rafflesiana* by Bauer et al. [16].

The two *Nepenthes* species, *N. rafflesiana* and *Nepenthes hemsleyana*, differ in their nutrient sequestration strategy, with the former being a 'typical' pitcher plant relying on arthropod capture and the latter having a dual strategy of obtaining nutrients both from arthropod capture and the faeces of Hardwicke's Woolly Bats, *Kerivoula hardwickii*, which roost in the pitchers during daytime [17]. This bat-plant mutualism has only recently been reported [17] and up to now remains to be studied in more detail. However, it seems that, in parts of its range, *K. hardwickii* bats are quite dependent on *N. hemsleyana* as day roosts and only occasionally choose less suitable pitchers of other *Nepenthes* species, for example *Nepenthes bicalcarata* [18]. Preliminary genetic data indicate that *N. hemsleyana* and *N. rafflesiana* (the typical form) are direct sister taxa (M. Scharmann, pers. communication). The two species have only recently been recognised as separate, with *N. hemsleyana* previously being referred to as the elongate form of *N. rafflesiana* [19, 20]. The two species differ in pitcher morphology and subsequently in prey capture, retention strategies and suitability as roosting sites for bats [17, 19, 21–24]. Prey attraction mechanisms are often retarded in *N. hemsleyana* [19], although it retains the ability to digest insect prey [17, 22].

Despite *Nepenthes* plants having been studied for over a century [25], surprisingly little is known about any associations with microorganisms [26]. Regarding the role of microbes in the digestion process, it has been proposed that bacteria might support prey digestion [13, 27, 28], but some reports seem to refute that idea. For example, a recent study has found that the fluid of closed and newly opened pitchers is sterile and does not support microbial growth [29]. Furthermore, *Nepenthes* species are equipped with a variety of endogenous digestive enzymes, including aspartic proteases [15] and chitinases [30], indicating that pitcher plants do not require microbial support in prey breakdown. Other endogenous proteins secreted into the pitcher fluid, such as plumbagin derivatives, have antimicrobial activities [31–33], which might inhibit bacterial growth in the fluid. However, the presence of endogenous enzymes in the plant does not preclude supportive action from bacterial enzymes. Higashi et al. [13] suggested that enzymes from both the plant and bacteria occurring in the pitchers are involved in prey breakdown. Additionally, there is strong evidence supporting a diverse and functional microbial community in mature pitchers with potential to provide services for the plant. Firstly, high bacterial diversity [13, 34], complex bacterial profiles [35] as well as active bacterial enzymes [26, 36, 37] have been reported in the fluid. Secondly, Chou et al. [38] analysed bacterial communities in three *Nepenthes* species—*Nepenthes ampullaria*, *Nepenthes gracilis* and *Nepenthes mirabilis*—in their natural habitat and found that even unopened pitchers harbour bacterial assemblages, but that the community structure differed between closed and opened pitchers. They concluded that bacterial communities in *Nepenthes* pitchers can be endogenous or introduced, with introduced taxa being potential competitors for nutrients. In the study on *Nepenthes* fluid bacteria [38], only 900 bacterial 16s ribosomal DNA (rDNA) sequences (30 per sample) were analysed; thus, their results likely did not represent the complete bacterial community. Introduction of bacteria by prey has also been proposed for the convergent, but phylogenetically very distantly related Northern pitcher plants (*Sarracenia* spp., Sarraceniaceae, Ericales) [39–41], which also capture prey in pitcher traps [25]. However, it was proposed that the identified bacteria could also be beneficial to the plant host through fixation of atmospheric nitrogen and subsequent provision of additional nitrogenous compounds [39], which is also supported by modelling studies [42].

In this study, we analysed complete bacterial communities in two *Nepenthes* species by applying culture-independent high-throughput sequencing technology. We collected samples from natural field sites and analysed bacterial diversity, community structure and taxa distribution. Since the various parts of a pitcher can be considered to be very different with regard to morphological surface structure, exposure to the surroundings and chemical properties, we were interested whether these properties also affect bacterial assemblages. Additionally, we expected distinct bacterial communities in the digestive fluids of the two species, mainly affected by their different diets. We hypothesised that the microbiota of *N. rafflesiana* would contain large proportions of insect-associated taxa. In contrast, vertebrate gut symbionts were expected to be present in *N. hemsleyana* fluid, introduced by the faeces of *K. hardwickii*.

Materials and Methods

Sample Design

Samples were collected near Labi, Brunei Darussalam, Borneo in June 2013. One aerial pitcher of five different plants each of *N. rafflesiana* (the typical form) and *N. hemsleyana* were sampled (Table 1). Each pitcher was fully developed, open for approximately 4–6 weeks and contained captured prey, which was not characterised further. *N. rafflesiana* grew in an open disturbed lowland heath forest (4° 38' 50.55 N, 114° 30' 31.38 E), and *N. hemsleyana* grew in a peat swamp with heath forest mosaic along a small open stream valley (4° 35' 06.92 N, 114° 30' 35.10 E). The two sites were 6.7 km

Table 1 Sample list including index allocation and obtained sequence reads

Site	Sample ID	Species	Tissue	Remarks	Index 1	Sequence index 1 (5′–3′)	Index 2	Sequence index 2 (5′–3′)	Sequence reads obtained
Labi 39	R1-p	*N. rafflesiana*	P	Pitcher fully exposed to the sun: third leaf from top; ~1 m above ground; peristome width 51 mm; pitcher length 89.5 mm	SA712	TCGCTATA	SA501	ATCGTACG	35,177
	R1-i	*N. rafflesiana*	I		SA711	TCATAGAC	SA501	ATCGTACG	34,777
	R1-o	*N. rafflesiana*	O		SA710	TAGCAGAC	SA501	ATCGTACG	28,641
	R1-l	*N. rafflesiana*	L		SA709	GTCGTAGT	SA501	ATCGTACG	25,544
	R1-F	*N. rafflesiana*	F		SA707	GGAGACTA	SA501	ATCGTACG	23,667
	R2-p	*N. rafflesiana*	P	Pitcher fully exposed to the sun: third leaf from top; ~1.5 m above ground; peristome width 52.8 mm; pitcher length 106.7 mm	SA712	TCGCTATA	SA502	ACTATCTG	39,102
	R2-i	*N. rafflesiana*	I		SA711	TCATAGAC	SA502	ACTATCTG	36,567
	R2-o	*N. rafflesiana*	O		SA710	TAGCAGAC	SA502	ACTATCTG	48,296
	R2-F	*N. rafflesiana*	F		SA707	GGAGACTA	SA502	ACTATCTG	40,844
	R3-p	*N. rafflesiana*	P	Pitcher fully exposed to the sun: fifth leaf from top; ~1.5 m above ground; peristome width 55.6 mm; pitcher length 127.2 mm	SA712	TCGCTATA	SA503	TAGCGAGT	76,517
	R3-i	*N. rafflesiana*	I		SA711	TCATAGAC	SA503	TAGCGAGT	54,969
	R3-o	*N. rafflesiana*	O		SA710	TAGCAGAC	SA503	TAGCGAGT	49,590
	R3-F	*N. rafflesiana*	F		SA707	GGAGACTA	SA503	TAGCGAGT	34,465
	R4-p	*N. rafflesiana*	P	Pitcher fully exposed to the sun: third leaf from top; ~1.8 m above ground; peristome width 55.0 mm; pitcher length 98.5 mm	SA712	TCGCTATA	SA504	CTGCGTGT	49,581
	R4-i	*N. rafflesiana*	I		SA711	TCATAGAC	SA504	CTGCGTGT	37,347
	R4-o	*N. rafflesiana*	O		SA710	TAGCAGAC	SA504	CTGCGTGT	25,991
	R4-l	*N. rafflesiana*	L		SA709	GTCGTAGT	SA504	CTGCGTGT	28,146
	R4-F	*N. rafflesiana*	F		SA708	GTCGCTCG	SA504	CTGCGTGT	17,378
	R5-p	*N. rafflesiana*	P	Pitcher fully exposed to the sun: third leaf from top; ~1.5 m above ground; peristome width 54.3 mm; pitcher length 92.5 mm	SA712	TCGCTATA	SA505	TCATCGAG	80,029
	R5-i	*N. rafflesiana*	I		SA711	TCATAGAC	SA505	TCATCGAG	70,490
	R5-o	*N. rafflesiana*	O		SA710	TAGCAGAC	SA505	TCATCGAG	55,381
	R5-F	*N. rafflesiana*	F		SA708	GTCGCTCG	SA505	TCATCGAG	23,332
Labi 31	H1-p	*N. hemsleyana*	P	With bat on the 26th and 27th June. Pitcher in full shade: fourth leaf from top; ~1.2 m above ground; peristome width 39 mm; pitcher length 193 mm	SA706	CTACGCAG	SA501	ATCGTACG	20,067
	H1-i	*N. hemsleyana*	I		SA705	CGTACTCA	SA501	ATCGTACG	17,730
	H1-o	*N. hemsleyana*	O		SA704	CAGTGAGT	SA501	ATCGTACG	22,978
	H1-F	*N. hemsleyana*	F		SA703	AGTAGCGT	SA501	ATCGTACG	42,735
	H2-p	*N. hemsleyana*	P	No bat known to roost here. Pitcher 3/4 in shade: third leaf from top; ~1.5 m above ground; peristome width 36 mm; pitcher length 182 mm	SA706	CTACGCAG	SA502	ACTATCTG	35,586
	H2-i	*N. hemsleyana*	I		SA705	CGTACTCA	SA502	ACTATCTG	31,729
	H2-o	*N. hemsleyana*	O		SA704	CAGTGAGT	SA502	ACTATCTG	16,958
	H2-l	*N. hemsleyana*	L		SA707	GGAGACTA	SA502	ACTATCTG	22,148
	H2-F	*N. hemsleyana*	F		SA703	AGTAGCGT	SA502	ACTATCTG	52,593
	H3-p	*N. hemsleyana*	P	Two bats inside pitcher on 27th June. Pitcher 4/5 in shade: second leaf from top; ~1 m above ground;	SA706	CTACGCAG	SA503	TAGCGAGT	7,475
	H3-i	*N. hemsleyana*	I		SA705	CGTACTCA	SA503	TAGCGAGT	32,116
	H3-o	*N. hemsleyana*	O		SA704	CAGTGAGT	SA503	TAGCGAGT	12,112

Table 1 (continued)

Site	Sample ID	Species	Tissue	Remarks	Index 1	Sequence index 1 (5′–3′)	Index 2	Sequence index 2 (5′–3′)	Sequence reads obtained
	H3-F	*N. hemsleyana*	F	peristome width 42.5 mm; pitcher length 160 mm	SA703	AGTAGCGT	SA503	TAGCGAGT	5,581
	H4-p	*N. hemsleyana*	P	No bat known to roost here. Pitcher 5/6	SA706	CTACGCAG	SA504	CTGCGTGT	13,242
	H4-i	*N. hemsleyana*	I	in shade: fourth leaf from top; ~2.2 m	SA705	CGTACTCA	SA504	CTGCGTGT	22,111
	H4-o	*N. hemsleyana*	O	above ground; peristome width 50 mm;	SA704	CAGTGAGT	SA504	CTGCGTGT	18,215
	H4-F	*N. hemsleyana*	F	pitcher length 218 mm	SA703	AGTAGCGT	SA504	CTGCGTGT	27,065
	H5-p	*N. hemsleyana*	P	No bat known to roost here. Pitcher 2/3 in	SA706	CTACGCAG	SA505	TCATCGAG	9,315
	H5-i	*N. hemsleyana*	I	shade: fifth leaf from top; ~1 m above	SA705	CGTACTCA	SA505	TCATCGAG	20,748
	H5-o	*N. hemsleyana*	O	ground; peristome width 39 mm; pitcher	SA704	CAGTGAGT	SA505	TCATCGAG	8,927
	H5-l	*N. hemsleyana*	L	length 136 mm	SA707	GGAGACTA	SA505	TCATCGAG	27,718
	H5-F	*N. hemsleyana*	F		SA703	AGTAGCGT	SA505	TCATCGAG	41,924

P peristome, *I* inside, *O* outside, *L* leaf, *F* fluid

apart. Two pitchers of *N. hemsleyana* were observed to harbour bats at the time of sampling (Table 1). For sample collection, sterile swabs wet with sterilised water were used to swab the peristome (P), inside (I) and outside (O) of the pitcher, as well as the leaf the pitcher was connected to (L). Fluid (F) was taken with a sterile disposable syringe. Swab heads were cut off to fall into a sterile microcentrifuge tube, and samples were frozen immediately until processing.

DNA Extraction, PCR and Library Preparation for Sequencing

Swab heads were covered in phosphate-buffered saline (PBS, pH 7.4), shaken horizontally for 10 min, swabs removed and centrifuged for 10 min (twice, pellet carried over). For fluids, up to 600 μL were used for DNA extraction. DNA was then isolated using the PowerSoil® DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA) following the manufacturer's instructions, but adding a 5-min incubation step at room temperature after applying the elution buffer. Alongside the actual samples, we processed a laboratory control, which consisted of an empty tube of the DNA extraction kit, to which 60 μL PCR grade water were added and which was then further processed in the exact same way as all other samples.

PCR and library preparation were performed according to a previously published dual indexing approach [43]. This method uses modular primers, consisting of MiSeq specific adapters, 8 nt index sequences, 10 nt pad sequences, 2 nt linker sequences and 16s rDNA amplifying primers [43]. The incorporated amplifying primers were 515f and 806r [44]. The MiSeq adapters ensure sequencing platform conformity; the pad sequence is designed to reach a melting temperature of approximately 65 °C for the pad-linker-primer sequence; the linker lowers self-complementarity and different combinations of forward and reverse index sequences allow sample-specific labelling [43]. The complete primer sequences were, forward: 5′-AATGATACGGCGACCA CCGAGATCTACAC XXXXXXXX TATGGTAATT GT GTGCCAGCMGCCGCGGTAA-3′ and reverse: 5′-CAA GCAGAAGACGGCATACGAGAT XXXXXXXX AGTCAGTCAG CC GGACTACHVGGGTWTCTAAT-3′ [43], where XXXXXXXX indicates the index sequences, which were selected as shown in Table 1.

PCR was performed in triplicate for each sample [45] in 10-μL reactions, each containing 5 μL 2× Phusion® High Fidelity PCR Master Mix (New England Biolabs, Ipswich, MA, USA), 0.33 μM each of forward and reverse primer (Eurofins MWG Operon, Huntsville, AL, USA), 3.34 μL PCR grade water and 1 μL template DNA. PCR conditions comprised an initial denaturation step at 95 °C for 4 min, 35 cycles of denaturation at 95 °C for 40 s, annealing at 55 °C for 30 s and elongation at 72 °C for 1 min, followed by final extension at 72 °C for 5 min. Triplicates of a sample

were pooled, and successful amplification was verified with an agarose gel, using 5 μL of the pool.

The remaining 25 μL were further processed using the SequalPrep™ Normalization Plate Kit (Invitrogen, Carlsbad, CA, USA), eluting in 20 μL, which works both as a PCR product clean-up removing excess primers and nucleotides as well as normalising DNA quantities to 25 ng for each sample. Of the eluate, 5 μL normalised DNA was taken for pooling with samples of other projects and laboratory control samples (pure extraction kit) according to Kozich et al. [43]. This pool was verified for library fragment size with a Bioanalyzer High Sensitivity DNA Chip (Agilent Technologies, Santa Clara, CA, USA) and quantified with the dsDNA High Sensitivity Assay (Life Technologies GmbH, Darmstadt, Germany). The final pool was diluted to 2 nM and further prepared for sequencing following the Illumina Guide for DNA library preparation [46], obtaining a final library of 10 pM. PhiX Control Kit v3 (Illumina Inc., San Diego, CA, USA) was added as a spike-in to ensure high-quality reads. Sequencing was performed on the Illumina MiSeq® Platform (Illumina Inc., San Diego, CA, USA) using 2 × 250 bp v2 MiSeq® chemistry. The cartridge of the reagent kit was additionally supplied with 3 μL each of the custom sequencing and index primers [43].

Bioinformatics and Statistics

Quality control was performed using FastQC v0.11.2 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Forward and reverse reads were joined together with fastq-join v.1.8.0 (https://code.google.com/p/ea-utils/wiki/FastqJoin). Demultiplexing was performed in QIIME v. 1.8.0 [47], which included quality filtering (Phred score > Q20, sequence length > 200 bp). In this step, only reads with complete index sequences were kept. Clusters of operational taxonomic units (OTUs) were built, chimaeras removed and taxonomically classified using the UCLUST [48] and UCHIME [49] algorithms, as implemented in USEARCH v. 7.0.1090 [48]. Chloroplast and mitochondrial 16S rDNA were filtered out. Additionally, overall laboratory control samples suggested Moraxellaceae, Brucellaceae, Oxalobacteraceae, Comamonadaceae and the order Caulobacterales to originate from kits, plasticware and laboratory contamination, which were removed prior to follow-up analyses. Although we cannot rule out that some of these might also be part of the natural microbiota of *Nepenthes*, we applied this conservative filtering, as presence of these taxa as well as their abundance would be highly overestimated in the results and thus not comparable with other bacteria.

The data set was imported into R v.3.0.3 [50] and analysed using the *phyloseq* [51] and *vegan* [52] packages. Sequencing depth was assessed using rarefaction curves. Samples not reaching saturation in the rarefaction were removed from further analyses, as they likely do not represent the whole community present (six leaves). Observed richness and Shannon's H species diversity index [53] were determined with respect to species and tissue. Sample counts were relativised for each sample in subsequent data analysis. We determined the most dominant taxa within the system using the summarize_taxa_through_plots.py command implemented in QIIME [47]. We used detrended correspondence analysis (DCA) [54] based on UniFrac distance [55] to infer similarities in community structure between samples visually. DCA eliminates two common artefacts of other multivariate analyses, such as the arch effect and compression of the ends of the axes, complicating interpretation and producing skewed patterns [54]. We performed environmental fitting of sample data, i.e. plant species and pitcher tissue to infer which variable explained data clustering. Further, we determined OTUs and their relative abundance that co-localised with fluid samples away from the centre to determine bacterial taxa associated with differences in community structure. Additionally, we determined how much of each fluid community was represented by the OTUs of each cluster. Variability in community structure within tissue replicates was further evaluated using analysis of similarity (ANOSIM) [56] with 999 permutations, based on the Bray-Curtis dissimilarity index [57] and grouped by tissue. This was accompanied by looking at the distribution of bacterial families within fluid samples.

Occurrence of certain pre-selected taxa with potential functions (see below) was explored by sub-setting the dataset for each group. These taxa included nitrogen-fixing bacteria: Burkholderiales [41], Rhizobiales [58, 59]; acidophilic taxa: Acidobacteriaceae [60], Acetobacteriaceae [61, 62]; vertebrate gut symbionts: *Mobiluncus*, *Clostridium*, *Anaerococcus* [63], Enterobacteriaceae [64], *Lactobacillus* [65], *Streptococcus* [66]; insect gut symbionts: *Serratia*, *Pantoea*, other Enterobacteriaceae, *Rhodococcus* [39], *Lactococcus* [40], *Chryseobacterium* [41] and bacteria previously identified in *Nepenthes* species: *Xanthobacter*, *Sphingomonas*, *Novosphingobium*, *Kaistia*, *Enterobacter*, *Dyella*, *Acidocella*, *Acidisoma* [38]. For identification of closely related bacterial reference species, OTUs classified as Enterobacteriaceae were additionally run against the NCBI microbial 16s database [67] using the BLASTn option [68] for better resolution between insect and vertebrate gut symbionts. For fluid samples of both *Nepenthes* species, we determined common bacterial OTUs present in at least 90 % of the fluid samples using the compute_core_microbiome.py command in QIIME [47]. Also, we inferred differences in taxa distributions with statistical pairwise comparisons between the two *Nepenthes* species for fluids by calculating the logarithmic fold change and applying the exact test between groups of count libraries, as implemented in the package *edgeR* [69]. The fold change is the ratio of a final value to an initial value. It is usually transformed to the logarithmic fold change (logFC) with the base of 2, so that a

twofold decrease has the logFC value of 1, a twofold decrease subsequently has the logFC value of −1. In our case, the final value corresponds to abundance in *N. rafflesiana*, whereas the initial value corresponds to abundance in *N. hemsleyana*. We only reported genera with statistically significant logFC values, which was determined by fitting negative binomial models to the raw sequence reads and estimating dispersion using the quantile adjusted conditional maximum likelihood method. Then, the exact test was used to determine statistically significant logFCs [69]. This approach has been recommended for microbiome data when analysing differential abundance [70].

Results

Sequencing Output

In total, we obtained 1,611,942 raw reads with quality above Q20 (probability of erroneous base calls within sequence reads 1 in 100), including a laboratory control. This accounted for 30,777 ± 18,581 reads per sample. Reads clustered into 1407 OTUs excluding mitochondria, chloroplasts and laboratory contamination. Most rarefaction curves reached an asymptote (Supplemental File 1), indicating that sequencing effort was sufficient for most samples to characterise the bacterial diversity. Raw sequence data has been deposited at the EBI-SRA with the project accession number PRJEB7957.

Community Structure

Summarised over both species and tissues, Proteobacteria dominated the bacterial community with 49.6 % of all reads, but Actinobacteria (15.8 %), Bacteroidetes (17.2 %) and Firmicutes (8.8 %) were also of considerable abundance. Within Proteobacteria, Alphaproteobacteria were most abundant (31.0 %), followed by Gammaproteobacteria (14.2 %). Beta- and Deltaproteobacteria were rare with 3.8 and 0.5 %, respectively. The most abundant orders of the Proteobacteria were Rhizobiales (14.6 %) and Rhodospirillales (9.8 %), followed by Sphingomonadales (4.8 %, all Alphaproteobacteria), Enterobacteriales (4.7 %), Pseudomonadales (3.6 %), Xanthomonadales (4.2 %, all Gammaproteobacteria) and Burkholderiales (3.2 %, Betaproteobacteria). Within the other phyla, Actinomycetales (13.0 %, Actinobacteria), *Sediminibacterium* (8.6 %, Bacteroidetes) and Bacilli (6.9 %, Firmicutes) were most abundant. A complete list of all detected taxa can be found in Supplemental File 2.

Bacterial Diversity

Alpha-diversity analysis revealed medium bacterial species richness averaging around 100 observed OTUs per sample, which was more or less similar across species and tissues (Fig. 1). For both species, some samples of the pitcher's outer surface showed an exceptional high number of OTUs present resembling the soil control (approximately 750 OTUs, data not shown). *N. hemsleyana* fluids showed higher diversity (Shannon's H index) compared to the other samples of the same species. *N. rafflesiana* fluids showed a slight decrease in diversity compared to the other tissues, but it was not significant (pairwise Wilcox test, uncorrected p values: $p_{inside} = 0.151$, $p_{leaf} = 0.857$, $p_{outside} = 0.095$, $p_{peristome} = 0.056$). The bacterial richness in the fluid of *N. rafflesiana* was significantly lower than in *N. hemsleyana* fluid (Fig. 1a; Wilcoxon rank sum test: $W = 22$, $p = 0.032$); bacterial diversity showed the same trend (Fig. 1b), but was not significantly different ($W = 19$, $p = 0.222$). The outsides of the plant species also seemed to differ in richness and diversity (Fig. 1), but this was not significant (richness: $W = 20$, $p = 0.151$; diversity: $W = 18$, $p = 0.309$). In DCA based on UniFrac distance, 32 and 24.4 % of the total variability between samples were explained on the first two axes (Fig. 2). Most samples clustered closely together, except some fluid and leaf samples, which clustered further away. Pitcher tissue explained sample clustering ($p = 0.001$, 999 permutations), but plant species did not ($p = 0.813$). Fluid samples of both species generally showed high variability between replicates, as inferred by remote clustering of fluid replicates in DCA (Fig. 2). For both plant species, two fluid replicates could be observed that clustered closely together, but far away from the other samples (Fig. 2, marked with # for *N. hemsleyana* and a for *N. rafflesiana*). In the case of *N. hemsleyana* these two replicates corresponded to the pitchers with observed bat roosting (Table 1). Other *N. hemsleyana* fluid replicates were located outside the centre (b in Fig. 2) or at the bottom left together with one outside replicate (c in Fig. 2). The species ordination (data not shown) revealed no clear picture as to which lineages or groups of ecologically similar bacteria (e.g. gut bacteria, nitrogen-fixing taxa) were associated with the respective fluid communities. Thus, we determined all OTUs that co-localised with those samples and combined them in clusters (Table 2 and Supplemental File 3). There were two types of OTUs associated with those samples. Firstly, taxa that could be detected in fluid as well as other samples, but reached higher levels in fluids (e.g. *Acidocella* spp. in cluster 2), and secondly, taxa that were exclusive to fluid samples (e.g. most Chitinophagaceae in cluster 1 and *Treponema* spp. in cluster 2), with the second type occurring more often (Supplemental File 3). Combining all OTUs within their respective clusters showed that these clusters made up a majority of the respective fluid communities (Table 2). Taken together, ordination analysis implied that bacterial communities in *Nepenthes* fluid could be extremely variable. We further assessed this by ANOSIM (Fig. 3), where fluids showed increased dissimilarity ranks compared to all other tissues except leaves (Nemenyi's test, uncorrected p values: $p_{between} < 0.001$, $p_{inside} < 0.001$, $p_{leaf} = 0.766$, $p_{outside} < 0.001$,

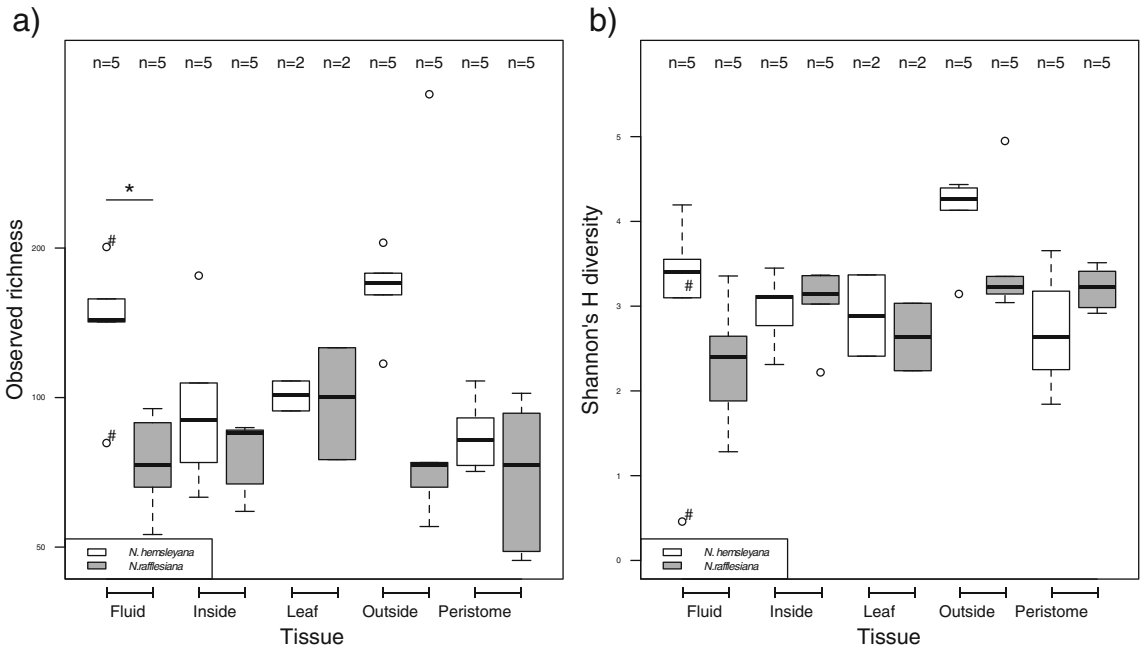


Fig. 1 Bacterial diversity with respect to plant species and tissue. Diversity indices were determined after filtering raw data, but before relativising abundance. **a** Observed species richness, plotted on a logarithmic y-axis for visualisation reasons, **b** Shannon's H diversity. *White N. hemsleyana*, *grey N. rafflesiana*. Order of tissues from left to right: Fluid, Inside, Leaf, Outside, Peristome. *Single dots* indicate outliers lying outside +/−1.5 of the interquartile range. **p* value < 0.05 (Wilcoxon rank sum test), *number sign* indicates *N. hemsleyana* fluid replicates with observed bat roosting, unspecified comparisons were not significant

$p_{peristome} < 0.001$). This was also evident by comparing the distribution of bacterial families within fluid samples, which clearly differed between the different fluid samples (Fig. 4a).

Regarding the taxonomic identities of OTUs within the clusters, some taxa stood out. For example, in cluster 1 (corresponding samples marked with # in Fig. 2), one OTU dominated the fluid community of one of those pitcher replicates. There, it accounted for 93.9 % of all reads and was also moderately abundant in its neighbouring fluid replicate (13.2 %). This taxon could also be detected in other samples, including *N. rafflesiana*, but never at such high levels (6.5 % for two *N. rafflesiana* fluid replicates, but less than 5 % in all other cases). This particular taxon belonged to the Enterobacteriaceae and was identified as being closely related to *Klebsiella* spp. by BLASTn [68], which is found in gastrointestinal tracts [71]. Furthermore, nine OTUs of this cluster belonged to the Chitinophagaceae. Although only detected at low abundance, these OTUs were almost completely exclusive to *N. hemsleyana* fluid (Supplemental File 3). Other Chitinophagaceae in the data set belonged to the genus *Sediminibacterium*, which was quite abundant and could be detected in most of the samples, or were detected almost exclusively in other samples of *N. hemsleyana*. Other OTUs of that cluster belonged to several different lineages, but none stood out. It should be noted that OTUs of other clusters also occurred in these two *N. hemsleyana* fluids.

For the second cluster (corresponding samples marked with a in Fig. 2), we found two OTUs of extremely high abundance levels as well as several low abundance OTUs belonging to the same taxonomic lineage. The OTUs with high abundance both belonged to the genus *Acidocella* (Order: Rhodospirillales), were closely related to one another and together accounted for 66.7 and 72.6 % of all reads in the two fluid communities of this cluster. Including other OTUs belonging to this genus raised these numbers to 74 % in both cases. Generally, *Acidocella* could be detected in six fluid samples as well as all other pitcher tissues except leaves, albeit with highly variable abundance levels (Fig. 4b) and mostly belonging to *N. rafflesiana*. At the same time, it was absent from several samples, also including two *N. rafflesiana* fluid samples. There were several OTUs within cluster 2 that belonged to the genus *Treponema*, all of which were exclusive to one *N. rafflesiana* fluid replicate, within that accounted for 0.7 % of the community. In the whole data set, there was one more OTU assigned to the genus *Treponema*, which was detected on the outside of one *N. hemsleyana* pitcher, accounting for 0.4 % of the community. Some OTUs assigned to *Treponema* were also significantly enriched in *N rafflesiana* fluid, compared to *N. hemsleyana* (Fig. 5d). Additionally, OTUs from cluster 1 contributed to 6.5 % of both fluid communities of those *N. rafflesiana* fluid replicates (a in Fig. 2).

In the other two clusters, no such OTUs or lineages were found. However, OTUs from cluster 3 made up 36.7 % of the respective fluid community of this cluster's *N. hemsleyana* replicate (b in Fig. 2), but that OTUs from cluster 1 and 2 contributed 5.1 and 5.8 % to that community as well. For cluster 4, we found that only 12 out of 55 OTUs belonging to that cluster could be detected in any of the other samples.

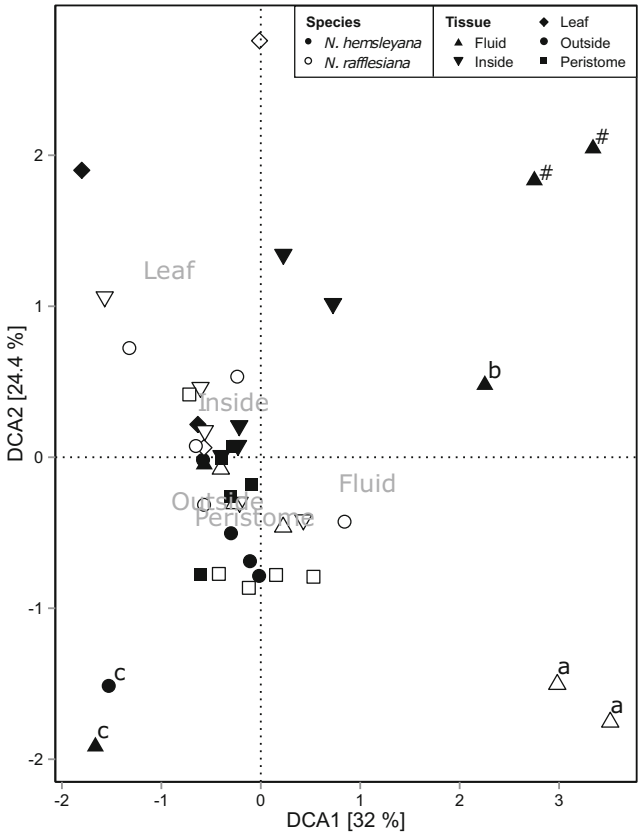


Fig. 2 Detrended correspondence analysis (DCA) based on UniFrac distance. Ordination was performed on filtered and relativised data. *Open symbols N. hemsleyana*, *closed symbols N. rafflesiana*; *shapes* indicate tissue. Tissue had a significant effect on data point location (p value = 0.001, 999 permutations). Tissue designation within the plotting area indicates the centroid of the respective tissue. *Number sign* indicates *N. hemsleyana* fluid replicates with observed bat roosting. *Lower case letters* (*a–c*) indicate samples for which co-localising taxa were investigated

Screening Taxa of Interest

Nitrogen-fixing bacteria, such as Rhizobiales and Burkholderiales were detected in all tissues of both plant species. In the Rhizobiales, *Bradyrhizobium* and *Methylosinus* were most abundant in most samples. Within the Burkholderiales, *Burkholderia* were also occasionally abundant, especially on the inside of *N. hemsleyana*. Of the pre-selected gut symbionts, OTUs belonging to the Enterobacteriaceae were most abundant in the fluids of both *Nepenthes* species. These were identified as being closely related to *Klebsiella*, *Pragia*, *Morganella*, *Escherichia*, *Providencia*, *Pantoea* and *Arsenophonus* spp. using BLASTn [68]. Other gut symbionts could also be recovered in fluids and the other tissues, albeit with low abundances. These included both vertebrate and insect gut symbionts, e.g. *Streptococcus* and *Chryseobacterium. Clostridium* occurred on the inside of both *Nepenthes* species, as well as on the outside and the leaf of *N. hemsleyana*, but not in the fluids. Other pre-selected taxa (*Xanthobacter*, *Sphingomonas*, *Novosphingobium*, *Kaistia*, *Enterobacter* and *Dyella*) were extremely rare or absent. The only exception was *Sphingomonas*, which occurred on all tissues of both *Nepenthes* species. Acidophilic taxa were overall quite rare (Fig. 4b) with the exception of the genus *Acidocella* (Order: Rhodospirillales), which was significantly enriched in *N. rafflesiana* fluid compared to *N. hemsleyana*, as revealed by the logFC (Fig. 5a). Taxa that were significantly

Table 2 Relative amounts of OTUs in clusters contributing to fluid communities

Sample ID	Cluster 1 DCA1 > 2.5 DCA2 < 1.5 (%)	Cluster 2 DCA1 > 2.5 DCA2 < −1.5 (%)	Cluster 3 2 < DCA1 < 3 0 < DCA2 < 1 (%)	Cluster 4 DCA1 < −1.5 DCA2 < −1.5 (%)
H1-F (#)	73.8	0.1	1.5	<0.1
H2-F (b)	5.1	5.8	36.8	<0.1
H3-F (#)	95.5	0.4	<0.1	0
H4-F (c)	<0.1	0.5	<0.1	71.3
H5-F	0.3	0.4	0	0.1
R1-F (a)	6.5	71.7	0	0
R2-F (a)	6.5	90.4	0	<0.1
R3-F	3.0	1.7	0	0.3
R4-F	0.3	<0.1	0	0.8
R5-F	1.2	1.3	0	0.4

Symbols and lower case letters in brackets indicate where the samples are located in Fig. 2

H N. hemsleyana, *R N. rafflesiana*, *F* fluid

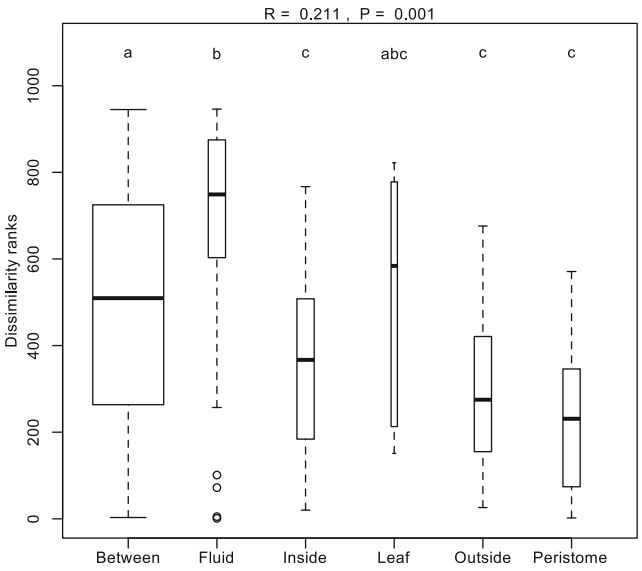


Fig. 3 Analysis of similarity (ANOSIM) for tissues based on Bray-Curtis dissimilarity. From *left to right*: Between, Fluid, Inside, Leaf, Outside, Peristome. *Single dots* indicate outliers lying outside +/−1.5 of the interquartile range. Please note that these do not correspond to sample replicates but to pairwise comparisons between dissimilarity ranks of replicates. *Tissues sharing the same lowercase letter* are not statistically different from one another (Nemenyi's test). Fluids and leaves show increased variability between pitcher replicates

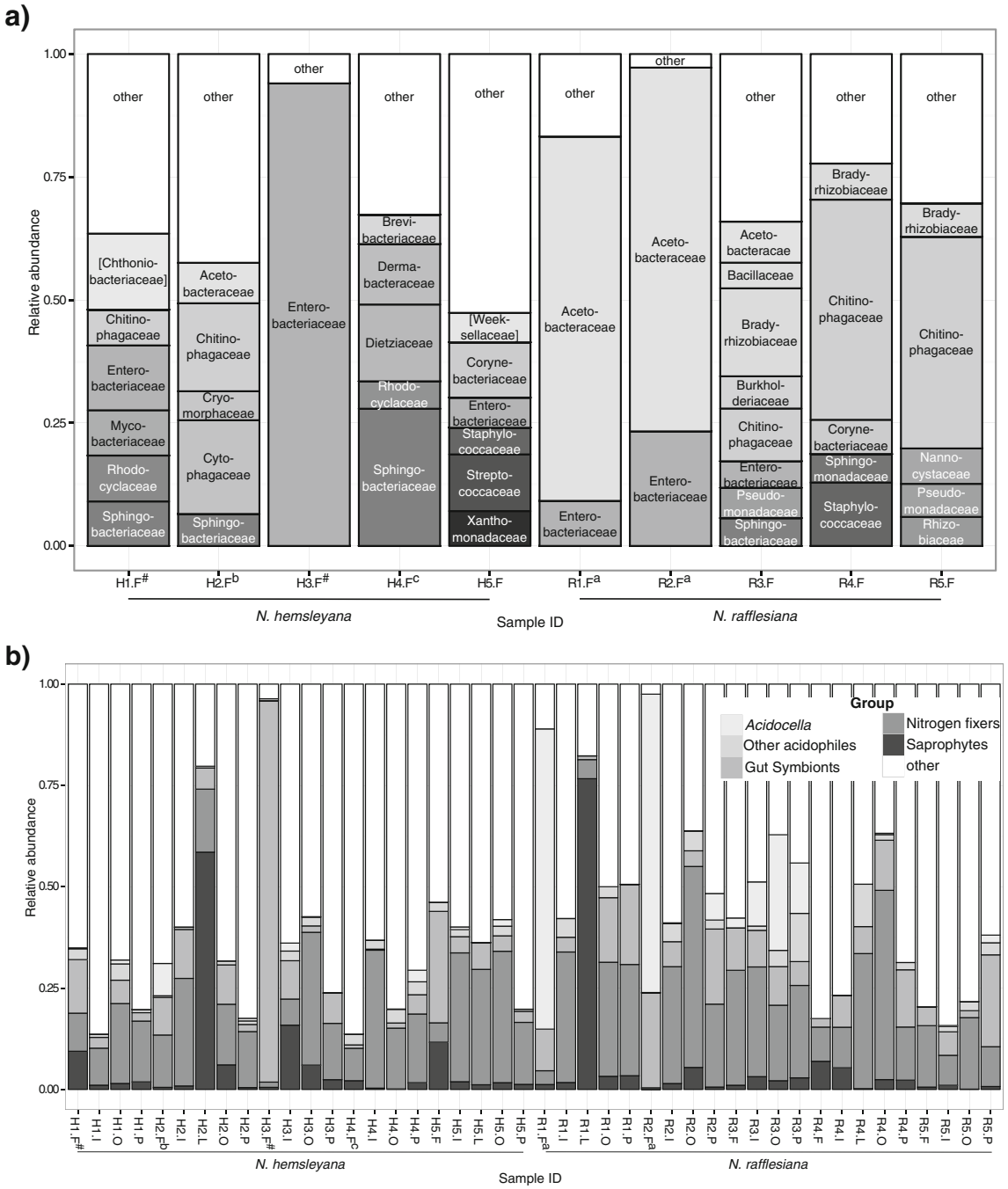
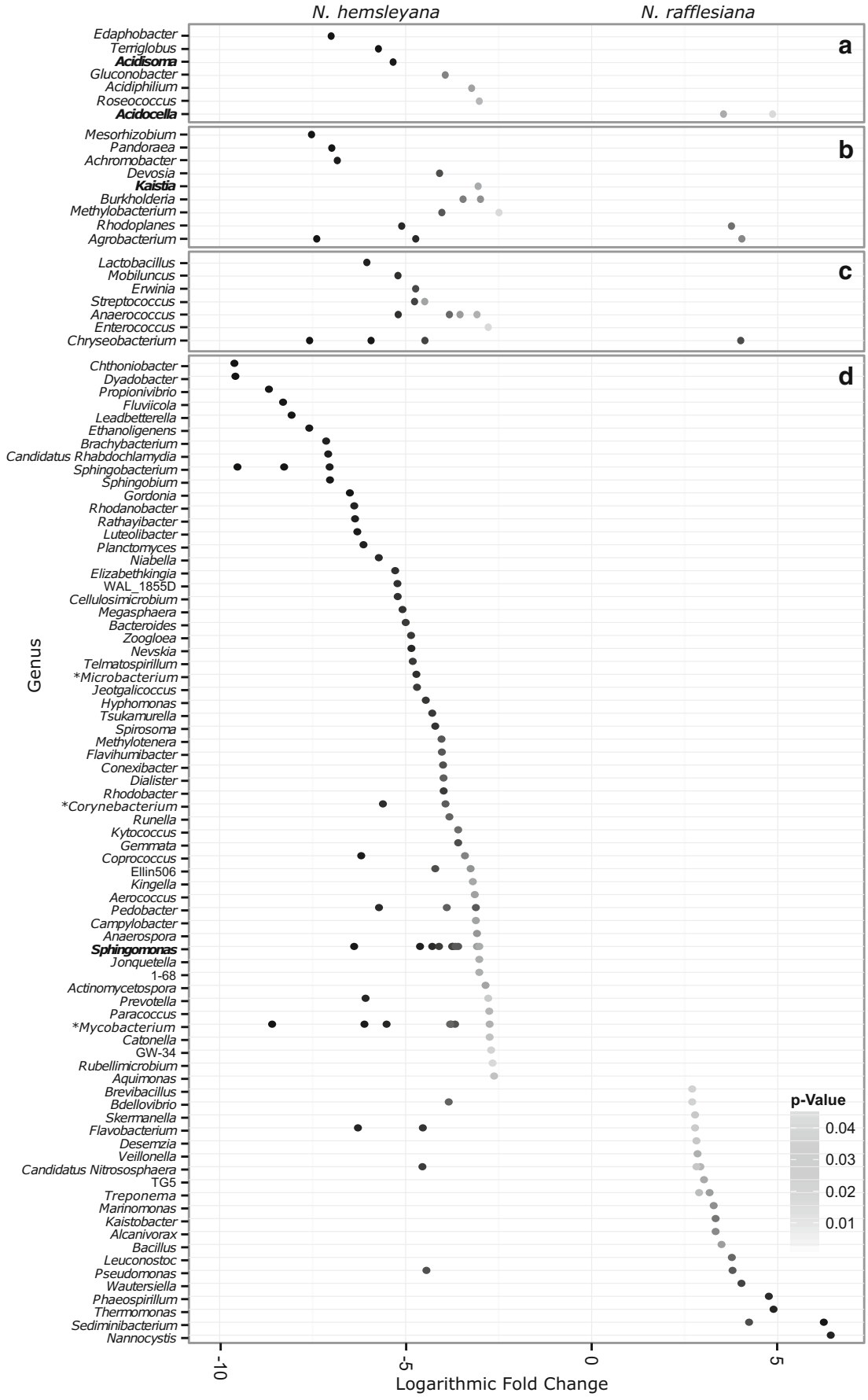


Fig. 4 Community composition. **a** Relative abundance of bacterial families in fluid samples. '*Other*' includes families representing less than 5 % of the sample communities. **b** Relative abundance of taxa grouped into ecologically similar groups (see 'Materials and Methods'). Abbreviations in sample IDs: *H N. hemsleyana*, *R N. rafflesiana*, *F* fluid, *I* inside, *L* leaf, *O* outside, *P* peristome. *Symbols and lower case letters* in the Sample IDs indicate where the samples are located in the ordination plot (Fig. 2), *number sign N. hemsleyana* fluid associated with bat roosting, *a–c* = fluid samples for which co-localising taxa were investigated

more abundant in *N. hemsleyana* fluid belonged to vertebrate gut symbionts, such as *Lactobacillus*, *Mobiluncus* and *Anaerococcus*, but also nitrogen-fixing taxa, e.g. *Mesorhizobium*, *Methylosinus* and *Burkholderia* (Fig. 5b). Some saprophytic taxa, such as *Corynebacterium*, *Sphingobacterium* and *Mycobacterium* were also enriched in *N. hemsleyana* fluid (Fig. 5d). Despite those differences, eight OTUs were common to 90 % of the fluid samples of both *Nepenthes* species. Two of those belonged to the Bradyrhizobiaceae with one OTU further classified as *Bradyrhizobium*, two OTUs belonged to the Enterobacteriaceae (closely related to *Klebsiella* and *Escherichia*) and one OTU each was classified as *Rhodococcus*, *Sphingomonas*, *Methylobacterium* and *Sediminibacterium*.



◀ **Fig. 5** Differences in taxa distributions within the digestive fluids based on logarithmic fold change. *Negative logarithmic fold change* enrichment in *N. hemsleyana* fluid, *positive logarithmic fold change* enrichment in *N. rafflesiana* fluid. *Data point shading* indicates *p* value, with *lighter shading* indicating higher *p* values. For visualisation reasons, only taxa with significantly different distributions are shown (exact test as defined in *edgeR* [69]). The taxa were split by putative ecological/functional roles as identified from literature and described in 'Materials and Methods'. Since assignment to such putative functions was based solely on literature research, functions might not be shared among all taxa within a group and additionally taxa with similar functions might be missing from the groups, respectively. **a** acidophilic taxa: Acidobacteriaceae and Acetobacteraceae; **b** nitrogen-fixing taxa: Burkholderiales and Rhizobiales; **c** vertebrate and insect gut symbionts: *Mobiluncus*, *Clostridium*, *Anaerococcus*, *Lactobacillus*, *Streptococcus*, *Serratia*, *Pantoea*, *Rhodococcus*, *Lactococcus*, *Chryseobacterium*, and other Enterobacteriaceae; **d** other: all remaining taxa with significantly different distributions. *Bold taxa* indicate taxa previously identified in *Nepenthes* by Chou et al. [38]; *asterisk* indicates saprophytic taxa: *Microbacterium*, *Mycobacterium*, and *Corynebacterium*

Discussion

We analysed bacterial communities associated with two *Nepenthes* species, *N. rafflesiana* and *N. hemsleyana*, which differ in their nutrient sequestration strategies. We compared bacterial diversity and community structure between the two *Nepenthes* species as well as different parts of the pitchers. We further inferred potential functional roles of specific bacterial taxa by comparing taxa distribution within the fluids of the two *Nepenthes* species.

Tissue Comparison

Regarding pitcher tissues, we found high similarities in bacterial diversity as well as community structure between most tissues. The outsides showed a somewhat increased bacterial richness and diversity in *N. hemsleyana*, and the leaves seemed rather variable in their community structure. Since these two tissues were exposed to the environment, those differences might be applicable to random colonisation events by environmental bacteria. This indicates that bacteria colonising these pitcher parts do not form a functional association with the plant and thus passively accumulate over time, as previously proposed [34]. The general similarity between the different pitcher parts was unexpected because each pitcher tissue exhibits distinct morphological and chemical surface characteristics, which we expected to affect bacterial community composition, as previously proposed for leaves of different tree species [72]. Within the pitchers, bacterial diversity was lower than in the environmental soil controls, similarly to a recent study on three other *Nepenthes* species [73]. This indicates that some selection mechanisms may exist to control certain groups of bacteria, for example plant pathogens. *Nepenthes* plants produce antimicrobial substances, such as plumbagin and its derivatives, in its fluid and leaves [74]. This substance is probably also expressed in other parts of the pitcher, which might be a mechanism of pathogen defence.

Nepenthes Species Comparison

In contrast to the tissue types discussed above, the digestive fluids showed more striking differences. Here, *N. hemsleyana* showed higher bacterial richness than *N. rafflesiana*. The community structure was highly variable between pitcher replicates of both species. This is in line with the findings of Chou et al. [38]. The difference in richness might be attributed to two opposing mechanisms: introduction of additional taxa in *N. hemsleyana* and suppression of various taxa in *N. rafflesiana*. We have indications for both, as fluid diversity of *N. hemsleyana* was probably increased due to bacteria introduced by bat faeces, which would not occur in *N. rafflesiana*. The two fluid replicates of *N. hemsleyana* with observed bat roosting showed a difference in community structure compared to the other replicates (Fig. 2), indicating that bat faeces introduction alters the bacterial community. Our results show that faeces might introduce bacterial taxa into the pitcher that would otherwise not occur in the fluid, since we found many taxa exclusive to these two samples, most notably a group of several Chitinophagaceae (Supplemental File 3). These might be important gut inhabitants of the insectivorous *K. hardwickii*. Additionally, several taxa associated with mammalian digestive tracts were significantly enriched in the *N. hemsleyana* fluid, including *Lactobacillus* [65, 75–77], *Mobiluncus* and *Anaerococcus* [63]. These genera were absent from *N. rafflesiana* fluid or very rare. However, these are only indications and since we could not control for nutrient source in the natural setting, it is difficult to attribute these differences to bat faeces alone. Thus, the effect of faeces introduction on the bacterial community as opposed to arthropod prey should be assessed in more controlled approaches. However, this increase in bacterial richness might entail increased competition with bacteria for nutrients. Indeed, some taxa with urease and nitrate reduction activities were present, such as *Mycobacterium* [78, 79], *Corynebacterium* [80, 81] and *Sphingobacterium* [82]. The genus *Mycobacterium* also includes animal pathogens [83, 84] and thus could possibly have been introduced by infected bats. Nonetheless, saprophytes would certainly be able to absorb nutrients from bat faeces and urine, so that they may be temporally nutrient competitors for *N. hemsleyana* or mutualists. However, at the end, they might also be digested by the pitcher plant.

The variability in community structure between fluid replicates of both species is in line with previous findings [38, 73], and is already reflected in different profiles of bacterial families (Fig. 4a). However, community variability could not be attributed to specific bacterial taxa or groups of ecologically similar bacteria. Instead, community differences were rather

OTU-specific and influenced by different high and low abundance OTUs as well as sample-specific OTUs. Thus, *Nepenthes* pitchers seem to constitute highly complex systems regarding their bacterial communities, with strain-specific differences. This can have several reasons, which probably all interact in shaping the bacterial community of *Nepenthes* pitcher fluid. Firstly, horizontal transfer of an endogenous community (see [38]) and subsequent diversification of taxa between unrelated plant individuals, which could be tested by comparing intra- and inter-individual differences in fluid microbiota. Secondly, the life history of a pitcher probably plays a big role in shaping the bacterial community, influenced by types of prey caught and time since last prey capture. Additionally, bacteria might also be passively introduced from the environment. Thus, the exogenous bacterial community [38] can be very different between different pitchers. Thirdly, *Nepenthes* pitchers, especially the fluids, could be interpreted as micro-ecosystems with high selective pressure due to high fluid pH [16, 85], but also highly dynamic conditions within the pitcher, due to prey capture and subsequent release of different nutrients. Fourthly, an alternative explanation has been given by Takeuchi et al. [73] who suggested that it is not particular bacterial taxa that are selected for but rather function-related traits. This is supported by the fact that we could detect members of groups with putative ecological functions in most samples.

As already mentioned, one major influence on bacterial community composition is probably the fluid pH, which changes during the digestion process [13] or due to ageing [16] and has been shown to be a major predictor of bacterial community structure [86, 87]. The influence of fluid pH seemed to be further supported by the distribution of one particular bacterial genus, *Acidocella* within *N. rafflesiana* fluids. *Acidocella* spp. could be extremely abundant with up to 74 % of the entire sample community, but at the same time were very variable in their relative abundance and even absent from some samples. Bacteria of this genus are acidophilic, fast growing and some produce acid from a variety of substrates [88, 89]. *Acidocella* has been found in other *Nepenthes* species [38] and also in close associations with other plants in acidic environments [89, 90], indicating a common association of that genus with a variety of plants, but the nature of such an association remains unclear. Considering the high variability of *Acidocella* distribution, its occurrence might also simply be affected by fluid pH or be even only occasional. Further analyses of a potential relationship between *Nepenthes* and *Acidocella* have to be undertaken in future studies.

Apart from *Acidocella*, other acidophilic taxa occurred at low abundance levels in both *Nepenthes* species, which has also been observed in other *Nepenthes* species [73]. This was really surprising, since the fluids of both species are highly acidic (pH = 2.7 for *N. hemsleyana* and pH = 2.1 for *N. rafflesiana* [85]) and can be as low as 1.95 in *N. rafflesiana* [16]. A possible explanation for this is that most bacteria living in the fluid are already adapted to be at least acid-tolerant, but do not belong to 'typical' acidophilic lineages, such as Acidobacteriaceae and Acetobacteraceae. Each of those strains might exhibit particular ecological characteristics, giving them certain roles in the fluid community.

Interestingly, one *N. rafflesiana* fluid sample with extreme *Acidocella* levels (a in Fig. 2), was also associated with several OTUs identified as *Treponema*. Some *Treponema* spp. are termite gut symbionts [91, 92], and thus this association might give an indication of the type of prey recently caught, which is interesting, because we sampled only aerial pitchers. Similarly, we found many Chitinophagaceae in the *N. hemsleyana* fluids with observed bat roosting (# in Fig. 2), which were almost exclusive to these pitchers. Other Chitinophagaceae, apart from *Sediminibacterium*, were mostly detected in other *N. hemsleyana* samples, with a few exceptions. Thus, these Chitinophagaceae might in some way be associated with the *Nepenthes*-bat mutualism.

An important finding is that the fluids of both species as well as the other pitcher parts additionally contained considerable amounts of putative nitrogen-fixing bacteria, accounting for 14.6 % (Rhizobiales) and 3.2 % (Burkholderiales) of all detected bacteria in the complete data set. Genera within those two families included *Bradyrhizobium*, *Methylosinus* and *Burkholderia*. Additionally, according to the common OTU analysis, those bacterial groups were also represented in the core microbiota of the fluid samples. *Burkholderia* and Rhizobiaceae were also found in the fluids of Northern pitcher plants of the genus *Sarracenia* [41], but so far only *Kaistia* spp. have been found in other *Nepenthes* species [38]. Such bacteria are common rhizosphere symbionts [59] converting atmospheric nitrogen to ammonia, which is readily absorbed by the plant (reviewed in e.g. [93]), or promoting plant growth in other ways, for example by producing phytohormones [94]. Bacterial nitrogen fixation has been shown to occur on submerged leaf surfaces of *Sarracenia* [95] and Ecuadorian bromeliads [96], although the nitrogen fixation rates were rather low [95, 96] and it remains unclear whether the host plants actually absorb the resulting ammonia [95, 96]. The detection of putative nitrogen fixers even in the extremely acidic fluids of *N. rafflesiana* and *N. hemsleyana*, indicates that strains exist that can tolerate such challenging environments. This finding is very interesting, and the nitrogen-fixing abilities of these strains should be further investigated by experimental approaches or nitrogen fixation gene analysis. This would provide further evidence for potentially mutualistic relationships between *Nepenthes* plants and fluid bacteria, which has already been proposed previously [4, 13]. Summing up, our results and previous studies in plant phytotelmata [95, 96] indicate that this particular plant-microbe mutualism might be more ubiquitous than previously

thought and might also occur in other nutrient sequestration systems apart from root systems. It is to be speculated that diverse microbial associations may contribute to nitrogen uptake during low-prey periods. In that respect, prey items might serve not only as the main nutrient source, but additionally as a substrate for cultivation of nitrogen-fixing taxa by the plant, which remains to be tested. The apparent acid tolerance of nitrogen-fixing bacteria found in *N. rafflesiana* fluid could be exploited in agricultural settings, potentially as a biofertiliser in acidic soils. *Sphingomonas* spp. were identified both in our study and other *Nepenthes* species [38] and were also identified as one of the OTUs belonging to the common bacteria of the fluids. This points towards a general association with *Nepenthes* pitcher plants, possibly as endophytes [38, 97], a relationship that has been reported for other plants as well [98, 99]. Finally, this great variety of potential symbionts suggests that interactions between the plant and other organisms involve networks of several species [100], in our case between pitcher plants, bats, symbiotic saprophytes and nitrogen-fixing bacteria.

Conclusions

Analysing microbial communities in the two species of tropical pitchers plants, *N. rafflesiana* and *N. hemsleyana*, revealed new insights into associations with microorganisms. We showed that both species harboured diverse bacterial communities in their pitchers, although only the fluids and external surfaces showed differences to the remaining pitcher tissues. We found high variability of bacterial community structure in the digestive fluids, which could not be associated with particular groups of bacteria, but which could be caused by diet, as well as other factors, for example fluid pH and time since last prey capture. These potential factors and their effect on bacterial communities should be investigated in controlled experiments. We detected *Acidocella* spp. at variable abundance levels, which might be commonly associated with a variety of plants. The occurrence of putative nitrogen-fixing bacteria indicated that such bacteria are not restricted to root systems, although nitrogen fixation within pitchers of *Nepenthes* remains to be shown.

Acknowledgments We are grateful to the Universitätsbund Würzburg e.V. for financially supporting this project (Grant number AZ-13-43) and the University of Würzburg/Land Bavaria for covering staff costs. We are furthermore most grateful to the Universiti Brunei Darussalam for logistical and financial support (URG 193) and the Brunei Darussalam Forestry Department for permission to enter forests and for granting the sampling permits (46/JPH/UND/17 PT.1).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no competing interests.

References

1. Juniper BE, Robins RJ, Joel D (1989) The carnivorous plants. Academic Press, London
2. Clarke MC (1997) *Nepenthes* of Borneo. Nat. Hist. Publ
3. Clarke MC (2001) *Nepenthes* of Sumatra and Peninsular Malaysia. Nat. Hist. Publ
4. Schulze W, Frommer WB, Ward JM (1999) Transporters for ammonium, amino acids and peptides are expressed in pitchers of the carnivorous plant *Nepenthes*. Plant J 17:637–646
5. Gaume L, Forterre Y (2007) A viscoelastic deadly fluid in carnivorous pitcher plants. PLoS One 2:e1185. doi:10.1371/journal.pone.0001185
6. Gaume L, Gorb S, Rowe N (2002) Function of epidermal surfaces in the trapping efficiency of *Nepenthes alata* pitchers. New Phytol 156:479–489. doi:10.1046/j.1469-8137.2002.00530.x
7. Riedel M, Eichner A, Jetter R (2003) Slippery surfaces of carnivorous plants: composition of epicuticular wax crystals in *Nepenthes alata* Blanco pitchers. Planta 218:87–97. doi:10.1007/s00425-003-1075-7
8. Riedel M, Eichner A, Meimberg H, Jetter R (2007) Chemical composition of epicuticular wax crystals on the slippery zone in pitchers of five *Nepenthes* species and hybrids. Planta 225:1517–1534. doi:10.1007/s00425-006-0437-3
9. Pavlovič A, Masarovičová E, Hudák J (2007) Carnivorous syndrome in Asian pitcher plants of the genus *Nepenthes*. Ann Bot 100:527–536
10. Moran JA, Clarke CM (2010) The carnivorous syndrome in *Nepenthes* pitcher plants. Current state of knowledge and potential future directions. Plant Signal Behav 5:644–648
11. Gaume L, Perret P, Gorb E et al (2004) How do plant waxes cause flies to slide? Experimental tests of wax-based trapping mechanisms in three pitfall carnivorous plants. Arthropod Struct Dev 33:103–111. doi:10.1016/j.asd.2003.11.005
12. Bonhomme V, Pelloux-Prayer H, Jousselin E et al (2011) Slippery or sticky? Functional diversity in the trapping strategy of *Nepenthes* carnivorous plants. New Phytol 191:545–554
13. Higashi S, Nakashima A, Ozaki H et al (1993) Analysis of feeding mechanism in a pitcher of *Nepenthes hybrida*. J Plant Res 106:47–54
14. An C-I, Fukusaki E, Kobayashi A (2001) Plasma-membrane H + -ATPases are expressed in pitchers of the carnivorous plant *Nepenthes alata* Blanco. Planta 212:547–555
15. An C-I, Fukusaki E, Kobayashi A (2002) Aspartic proteinases are expressed in pitchers of the carnivorous plant *Nepenthes alata* Blanco. Planta 214:661–667. doi:10.1007/s004250100665
16. Bauer U, Willmes C, Federle W (2009) Effect of pitcher age on trapping efficiency and natural prey capture in carnivorous *Nepenthes rafflesiana* plants. Ann Bot 103:1219–1226. doi:10.1093/aob/mcp065
17. Grafe TU, Schöner CR, Kerth G et al (2011) A novel resource-service mutualism between bats and pitcher plants. Biol Lett 7:436–439. doi:10.1098/rsbl.2010.1141
18. Schöner CR, Schöner MG, Kerth G, Grafe TU (2013) Supply determines demand: influence of partner quality and quantity on the interactions between bats and pitcher plants. Oecologia 173:191–202. doi:10.1007/s00442-013-2615-x
19. Clarke C, Moran JA, Lee CC (2011) *Nepenthes baramensis* (Nepenthaceae)—a new species from north-western Borneo. Blumea 56:229–233
20. Scharmann M, Grafe TU (2013) Reinstatement of *Nepenthes hemsleyana* (Nepenthaceae), an endemic pitcher plant from Borneo, with a discussion of associated *Nepenthes* taxa. Blumea 58:8–12. doi:10.3767/000651913X668465

21. Gaume L, Di Giusto B (2009) Adaptive significance and ontogenetic variability of the waxy zone in *Nepenthes rafflesiana*. Ann Bot 104:1281–91. doi:10.1093/aob/mcp238
22. Bauer U, Grafe TU, Federle W (2011) Evidence for alternative trapping strategies in two forms of the pitcher plant, *Nepenthes rafflesiana*. J Exp Bot 62:3683–3692. doi:10.1093/jxb/err082
23. Lim YS, Schöner CR, Schöner MG et al (2015) How a pitcher plant facilitates roosting of mutualistic woolly bats. Evol Ecol Res 16:1–11
24. Schöner MG, Schöner CR, Simon R et al (2015) Bats are acoustically attracted to mutualistic carnivorous plants. Curr Biol 25: 1911–1916. doi:10.1016/j.cub.2015.05.054
25. Mithöfer A (2011) Carnivorous pitcher plants: insights in an old topic. Phytochemistry 72:1678–1682. doi:10.1016/j.phytochem.2010.11.024
26. Takeuchi Y, Salcher MM, Ushio M et al (2011) In situ enzyme activity in the dissolved and particulate fraction of the fluid from four pitcher plant species of the genus *Nepenthes*. PLoS One 6: e25144. doi:10.1371/journal.pone.0025144
27. Amagase S, Mori M, Nakayama S (1972) Digestive enzymes in insectivorous plants. IV Enzymatic digestion of insects by Nepenthes secretion and drosera peltata extract: proteolytic and chitinolytic activities. J Biochem 72:765–767
28. Mayer E, Adlassnig W, Peroutka M, Lichtscheidl IK (2005) Microflora in the traps of pitcher plants. In: VXII Int. Bot. Congr. p 510
29. Buch F, Rott M, Rottloff S et al (2013) Secreted pitfall-trap fluid of carnivorous *Nepenthes* plants is unsuitable for microbial growth. Ann Bot 111:375–383. doi:10.1093/aob/mcs287
30. Eilenberg H, Pnini-Cohen S, Schuster S et al (2006) Isolation and characterization of chitinase genes from pitchers of the carnivorous plant *Nepenthes khasiana*. J Exp Bot 57:2775–2784. doi:10.1093/jxb/erl048
31. Hatano N, Hamada T (2008) Proteome analysis of pitcher fluid of the carnivorous plant *Nepenthes alata*. J Proteome Res 7:809–816. doi:10.1021/pr700566d
32. Hatano N, Hamada T (2012) Proteomic analysis of secreted protein induced by a component of prey in pitcher fluid of the carnivorous plant *Nepenthes alata*. J Proteomics 75:4844–4852. doi:10.1016/j.jprot.2012.05.048
33. Rottloff S, Stieber R, Maischak H et al (2011) Functional characterization of a class III acid endochitinase from the traps of the carnivorous pitcher plant genus, *Nepenthes*. J Exp Bot 62:4639–4647. doi:10.1093/jxb/err173
34. Sota T, Mogi M, Kato K (1998) Local and regional-scale food web structure in *Nepenthes alata* pitchers. Biotropica 30:82–91
35. Yogiara AS, Suhartono MT (2006) A complex bacterial community living in pitcher plant fluid. J Mikrobiol Indones 11:9–14
36. Morohoshi T, Oikawa M, Sato S et al (2011) Isolation and characterization of novel lipases from a metagenomic library of the microbial community in the pitcher fluid of the carnivorous plant *Nepenthes hybrida*. J Biosci Bioeng 112:315–20. doi:10.1016/j.jbiosc.2011.06.010
37. Lüttge U (1964) Untersuchungen zur Physiologie der Carnivoren-Drüsen. Planta 63:103–117
38. Chou LY, Clarke CM, Dykes GA (2014) Bacterial communities associated with the pitcher fluids of three *Nepenthes* (Nepenthaceae) pitcher plant species growing in the wild. Arch Microbiol 196:709–717. doi:10.1007/s00203-014-1011-1
39. Siragusa AJ, Swenson JE, Casamatta DA (2007) Culturable bacteria present in the fluid of the hooded-pitcher plant *Sarracenia minor* based on 16S rDNA gene sequence data. Microb Ecol 54: 324–31. doi:10.1007/s00248-006-9205-y
40. Peterson CN, Day S, Wolfe BE et al (2008) A keystone predator controls bacterial diversity in the pitcher-plant (*Sarracenia purpurea*) microecosystem. Environ Microbiol 10:2257–66. doi:10.1111/j.1462-2920.2008.01648.x
41. Koopman MM, Fuselier DM, Hird S, Carstens BC (2010) The carnivorous pale pitcher plant harbors diverse, distinct, and time-dependent bacterial communities. Appl Environ Microbiol 76: 1851–1860. doi:10.1128/AEM.02440-09
42. Mouquet N, Daufresne T, Gray SM, Miller TE (2008) Modelling the relationship between a pitcher plant (*Sarracenia purpurea*) and its phytotelma community: mutualism or parasitism? Funct Ecol 22:728–737. doi:10.1111/j.1365-2435.2008.01421.x
43. Kozich JJ, Westcott SL, Baxter NT et al (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79:5112–5120. doi:10.1128/AEM.01043-13
44. Caporaso JG, Lauber CL, Walters WA et al (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci 108:4516–4522. doi:10.1073/pnas.1000080107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1000080107
45. Fierer N, Hamady M, Lauber CL, Knight R (2008) The influence of sex, handedness, and washing on the diversity of hand surface bacteria. Proc Natl Acad Sci 105:17994–17999
46. Illumina Inc. (2013) Preparing libraries for sequencing on the MiSeq
47. Caporaso JG, Kuczynski J, Stombaugh J et al (2010) QIIME allows analysis of high- throughput community sequencing data. Nat Methods 7:335–336. doi:10.1038/nmeth0510-335
48. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461
49. Edgar RC, Haas BJ, Clemente JC et al (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27: 2194–2200
50. R Core Team (2014) R: a language and environment for statistical computing
51. McMurdie PJ, Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217. doi:10.1371/journal.pone.0061217
52. Dixon P (2003) VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930
53. Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana
54. Hill MO, Gauch HG Jr (1980) Detrended correspondence analysis: an improved ordination technique. Vegetatio 42:47–58
55. Lozupone CA, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–8235
56. Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. Aust J Ecol 18:117–143
57. Bray JR, Curtis T (1957) An ordination of upland forest communities of Southern Wisconsin. Ecol Monogr 27:325–349
58. Vessey JK (2003) Plant growth promoting rhizobacteria as biofertilizers. Plant Soil 255:571–586
59. Lucy M, Reed E, Glick BR (2004) Applications of free living plant growth-promoting rhizobacteria. Antonie Van Leeuwenhoek 86:1–25. doi:10.1023/B:ANTO.0000024903.10757.6e
60. Rowe OF, Sánchez-España J, Hallberg KB, Johnson DB (2007) Microbial communities and geochemical dynamics in an extremely acidic, metal-rich stream at an abandoned sulfide mine (Huelva, Spain) underpinned by two functional primary production systems. Environ Microbiol 9:1761–1771. doi:10.1111/j.1462-2920.2007.01294.x
61. Kersters K, Lisdiyanti P, Komagata K, Swings J (2006) The family Acetobacteraceae: the genera *Acetobacter, Acidimonas, Asaia, Gluconacetobacter, Gluconobacter,* and *Kozakia*. In: Dworkin

M, Falkow S, Rosenberg E, et al. (eds) Prokaryotes Vol. 5 Proteobacteria Alpha Beta Subclasses, Third Edit. Springer, pp 163–200

62. Belova SE, Pankratov TA, Detkova EN et al (2009) *Acidisoma tundrae* gen. nov., sp. nov. and *Acidisoma sibiricum* sp. nov., two acidophilic, psychrotolerant members of the Alphaproteobacteria from acidic northern wetlands. Int J Syst Evol Microbiol 59:2283–2290. doi:10.1099/ijs.0.009209-0
63. Lee JE, Lee S, Sung J, Ko G (2011) Analysis of human and animal fecal microbiota for microbial source tracking. ISME J 5:362–365
64. Ley RE, Hamady M, Lozupone CA et al (2008) Evolution of mammals and their gut microbes. Science 320(80-):1647–1651
65. Endo A, Irisawa T, Futagawa-Endo Y et al (2013) *Lactobacillus faecis* sp. nov., isolated from animal faeces. Int J Syst Evol Microbiol 63:4502–4507
66. O'Shea EF, Gardiner GE, O'Connor PM et al (2009) Characterization of enterocin- and salivaricin-producing lactic acid bacteria from the mammalian gastrointestinal tract. FEMS Microbiol Lett 291:24–34
67. Sayers EW, Barrett T, Benson DA et al (2011) Database resources of the national centre for biotechnology information. Nucleic Acids Res 39:D38–D51
68. Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410. doi:10.1016/S0022-2836(05)80360-2
69. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140
70. McMurdie PJ, Holmes S (2014) Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol 10: e1003531. doi:10.1371/journal.pcbi.1003531
71. Bagley S (1985) Habitat association of *Klebsiella* species. Infect Control 6:52–58
72. Redford AJ, Fierer N (2009) Bacterial succession on the leaf surface: a novel system for studying successional dynamics. Plant-Microbe Interact 58:189–198
73. Takeuchi Y, Chaffron S, Salcher MM et al (2015) Bacterial diversity and composition in the fluid of pitcher plants of the genus *Nepenthes*. Syst Appl Microbiol 38:330–339. doi:10.1016/j.syapm.2015.05.006
74. Raj G, Kurup R, Hussain AA, Baby S (2011) Distribution of naphthoquinones, plumbagin, droserone, and 5-O-methyl droserone in chitin-induced and uninduced *Nepenthes khasiana*: molecular events in prey capture. J Exp Bot 62:5429–5436. doi:10.1093/jxb/err219
75. Morotomi M, Yuki N, Kado Y et al (2002) *Lactobacillus equi* sp. nov., a predominant intestinal *Lactobacillus* species of the horse isolated from faeces of healthy horses. Int J Syst Evol Microbiol 52:211–214
76. Roos S, Karner F, Axelsson L, Jonsson H (2000) *Lactobacillus mucosae* sp. nov., a new species with in vitro mucus-binding activity isolated from pig intestine. Int J Syst Evol Microbiol 50: 251–258
77. Walter J, Britton RA, Roos S (2011) Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. Proc Natl Acad Sci 108:4645–4652
78. Howard ST, Byrd TF (2000) The rapidly growing mycobacteria: saprophytes and parasites. Microbes Infect 2:1845–1853
79. Oram DM, Avdalovic A, Holmes RK (2004) Analysis of genes that encode DtxR-like transcriptional regulators in pathogenic and saprophytic corynebacterial species. Infect Immunol 72:1885–1895
80. Abe S, Takayama KI, Kinoshita S (1967) Taxonomical studies on glutamic acid-producing bacteria. J Gen Appl Microbiol 13:279–301
81. Pitcher D, Soto A, Soriano F, Valero-Guillén P (1992) Classification of coryneform bacteria associated with human urinary tract infection (group D2) as *Corynebacterium urealyticum* sp. nov. Int J Syst Bacteriol 42:178–181
82. Yabuuchi E, Kaneko T, Yano I et al (1983) *Sphingobacterium* gen. nov., *Sphingobacterium spiritivorum* comb. nov., *Sphingobacterium multivorum* comb. nov., *Sphingobacterium mizutae* sp. nov., and *Flavobacterium indologenes* sp. nov.: glucose-nonfermenting gram-negative rods in CDC groups IIK-2 and IIb. Int J Syst Evol Bacteriol 33:580–598
83. Anes E, Kühnel MP, Bos E et al (2003) Selected lipids activate phagosome actin assembly and maturation resulting in killing of pathogenic mycobacteria. Nat Cell Biol 5:793–802. doi:10.1038/ncb1036.NATURE
84. Cosma CL, Sherman DR, Ramakrishnan L (2003) The secret lives of the pathogenic mycobacteria. Annu Rev Microbiol 57:641–676. doi:10.1146/annurev.micro.57.030502.091033
85. Bazile V, Le Moguédec G, Marshall DJ, Gaume L (2015) Fluid physico-chemical properties influence capture and diet in *Nepenthes* pitcher plants. Ann Bot 115:705–716. doi:10.1093/aob/mcu266
86. Lauber CL, Hamady M, Knight R, Fierer N (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol 75:5111–5120
87. Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci 103:626–631
88. Kimoto KI, Aizawa T, Urai M et al (2010) *Acidocella aluminiidurans* sp. nov., an aluminium-tolerant bacterium isolated from *Panicum repens* grown in a highly acidic swamp in actual acid sulfate soil area of Vietnam. Int J Syst Evol Microbiol 60: 764–768
89. Jones RM, Hedrich S, Johnson DB (2013) *Acidocella aromatica* sp. nov.: an acidophilic heterotrophic alphaproteobacterium with unusual phenotypic traits. Extremophiles 17:841–850
90. Bragina A, Maier S, Berg C et al (2012) Similar diversity of Alphaproteobacteria and nitrogenase gene amplicons on two related *Sphagnum* mosses. Front Microbiol 2:1–10
91. Dröge S, Rachel R, Radek R, König H (2008) *Treponema isoptericolens* sp. nov., a novel spirochaete from the hindgut of the termite *Incisitermes tabogae*. Int J Syst Evol Microbiol 58:1079–1083. doi:10.1099/ijs.0.64699-0
92. Graber JR, Leadbetter JR, Breznak JA (2004) Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. Appl Environ Microbiol 70:1315–1320. doi:10.1128/AEM.70.3.1315
93. Richardson AE, Barea J-M, McNeill AM, Prigent-Combaret C (2009) Acquisition of phosphorous and nitrogen in the rhizosphere and plant growth promotion by microorganisms. Plant Soil 321:305–339
94. Boiero L, Perrig D, Masciarelli O et al (2007) Phytohormone production by three strains of *Bradyrhizobium japonicum* and possible physiological and technological implications. Appl Microbiol Biotechnol 74:874–880
95. Prankevicius AB, Cameron DM (1991) Bacterial dinitrogen fixation in the leaf of the northern pitcher plant (*Sarracenia purpurea*). Can J Bot 69:2296–2298. doi:10.1139/b91-289
96. Bermudes D, Benzing DH (1991) Nitrogen fixation in association with Ecuadorean bromeliads. J Trop Ecol 7:531–536
97. Bhore SJ, Komathi V, Kandasamy KI (2013) Diversity of endophytic bacteria in medicinally important *Nepenthes* species. J Nat Sci Biol Med 4:431–434
98. Videira SS, de Araujo JLS, da Silva RL et al (2009) Occurrence and diversity of nitrogen-fixing *Sphingomonas* bacteria associated with rice plants grown in Brazil. FEMS Microbiol Lett 293:11–19. doi:10.1111/j.1574-6968.2008.01475.x

99. Chen B, Shen J, Zhang X et al (2014) The endophytic bacterium, *Sphingomonas* SaMR12, improves the potential for zinc phytoremediation by its host, *Sedum alfredii*. PLoS One 9: e106826. doi:10.1371/journal.pone.0106826

100. Blatrix R, Djiéto-Lordon C, Mondolot L et al (2012) Plant-ants use symbiotic fungi as a food source: new insight into the nutritional ecology of ant-plant interactions. Proc R Soc B 279:3940–3947. doi:10.1098/rspb.2012.1403

Part III.

Discussion

Overview

The studies presented above have given an overview of meta-barcoding as a method (Publication P.1) and described example applications of meta-barcoding in pollen analysis (Publications P.2 and P.3) and host-microbe association studies (Publications P.4 and P.5). These now form the basis for a critical evaluation of meta-barcoding in high-throughput biodiversity assessments, which will include the advantages and disadvantages of the method as well as recommendations for the application of meta-barcoding to specific research questions and an outlook of future developments.

Advantages of meta-barcoding

Meta-barcoding has been rapidly adopted in different fields of biological research. The main reasons for this are that species resolution can be higher than in traditional species identification based on morphological characters; species assemblages can be studied easily without prior separation of individuals; large-scale studies have become affordable due to the high throughput of sequencing platforms that allows sample multiplexing and the avoidance of cultivation especially for microbiology.

Taxonomic assignment. Meta-barcoding circumvents two central problems commonly encountered in traditional species identification. Firstly, taxonomy is assigned via DNA sequence analysis, thus limited resolution of morphological species identification (Williams and Kremen 2007; Galimberti et al. 2014; Handelsman and Smalla 2003) is not an issue. It is also highly beneficial in the analysis of processed and degraded material, which is the case in palaeo-ecological studies (Murray et al. 2012; Behling et al. 2004; Turner et al. 2013), diet analysis (Soininen et al. 2009; Valentini et al. 2009; Pompanon et al. 2012), food safety and quality issues (Woolfe and Primrose 2004; Galimberti et al. 2014; Bruni et al. 2015) and analysis of traditional Chinese medicine (Yip et al. 2007; Li et al. 2011; Coghlan et al. 2012).

Since taxonomy is assigned via DNA-barcoding, experts in taxonomy are not required for most applications of meta-barcoding, which is especially advantageous considering the approaching shortage of taxonomists (Gaston and May 1992; Carvalho et al. 2007). For meta-barcoding, knowledge of basic techniques of molecular biology is sufficient and samples can be prepared with basic molecular biology laboratory equipment (see Publications P.2; P.3; Kozich et al. 2013). For data analysis, some expertise in bioinformatics is needed, but there are already a variety of tools available that can

be used with a basic understanding of command line usage (see Publication P.3; Caporaso et al. 2010; Edgar 2013).

Circumventing manual separation. Secondly, due to the high throughput of sequences in modern sequencing platforms, manual separation of individual specimens (Galimberti et al. 2014; Handelsman and Smalla 2003) is not required. This is highly beneficial in pollen analysis (see Publications P.2 and P.3), because complete assemblages can be processed as one. This is also highly advantageous in bacterial community analysis (see Publications P.4 and P.5), since isolation and cultivation prior to analysis (Handelsman and Smalla 2003) is bypassed. Circumventing manual separation before DNA-barcoding is furthermore exploited in environmental DNA (eDNA) analysis, which is important for biodiversity monitoring (Hajibabaei et al. 2016).

High throughput. The high data output further allows sample multiplexing, which allows scaling up sample sizes per study, thus enabling the adoption of meta-barcoding in large-scale studies. The amount of multiplexing can be estimated in well-studied systems (see Publication P.2), but should always allow a sufficient sequencing output per sample and still be practical concerning the workflow within a laboratory. There is another advantage of meta-barcoding due to the high sequencing throughput, which is the detection of low abundance members of a community. This is the case both for pollen analysis (see Publication P.2) as well as bacterial community analysis (Bent and Forney 2008) and diet analysis (Pompanon et al. 2012).

From sequence to cultivation. Additionally, in bacterial community analysis, meta-barcoding data can be useful in advancing cultivation (Handelsman and Smalla 2003). Taxa, that have so far not been described can be detected and close relatives identified. Such taxa may be identified as potentially major players within an ecosystem or as taxa with potentially interesting characteristics. It is then possible to devise directed cultivation approaches using prior knowledge of close relatives. Further analysis of the taxon in question would thus advance the knowledge of microbial ecology. For example, in publication P.5 the bacterial genus *Acidocella* reached extremely high levels of abundance within the bacterial community. This genus has also been detected in other *Nepenthes* species (Chou et al. 2014), as well as *Sphagnum* mosses (Bragina et al. 2012). It might thus be an interesting genus to cultivate and study in more detail, especially regarding its association *Nepenthes* and *Sphagnum*. Knowing that *Acidocella* species are acidophilic (Kimoto et al. 2010; Jones et al. 2013) means that directed cultivation approaches should be performed on acidic media (Kishimoto et al. 1995).

Disadvantages of meta-barcoding

Despite these advantages, there are some drawbacks of meta-barcoding, especially during data analysis, how the data can be handled and what can be inferred from the data. In the following, five important limitations will be discussed: application of algorithms during data processing, insufficient species resolution and taxonomy assignments, quality of reference databases, abundance estimation and functional profiling.

Data processing. Data processing in meta-barcoding requires several consecutive steps, for example quality filtering, OTU clustering, chimera checking and taxonomic classification. These steps typically rely on algorithms (Wang et al. 2007; Caporaso et al. 2010; Edgar 2010; Edgar et al. 2011; Coissac et al. 2012; Edgar 2013). During sequencing, errors can be introduced and it is of utmost importance to distinguish these from true sequence variety, otherwise diversity estimates will be overestimated dramatically (Coissac et al. 2012). Sequencing errors can be reduced by joining forward and reverse reads, filtering out low-quality sequences, but also by removing sequences that only occur once, so-called singletons. Joining forward and reverse reads generally improves the data, because sequence quality drops towards the end of the read (Cuesta-Zuluaga and Escobar 2016). Filtering out sequences that are below a certain quality threshold (e.g. probability of erroneous base calls within a sequence 1 in 100) is common practice (Cuesta-Zuluaga and Escobar 2016). Removing singletons is performed, because they are assumed to have arisen from sequencing errors (Coissac et al. 2012).

Another step is OTU clustering, which is the formation of so-called operational taxonomic units as a proxy for species-level taxa (Blaxter 2016). OTU clustering is performed before taxonomic classification, because it would be too computationally intensive to run every single sequence obtained against a database for taxonomic classification. OTU clustering algorithms usually rely on sequence alignment and differentiate OTUs based on sequence dissimilarity (Edgar 2013; Cuesta-Zuluaga and Escobar 2016). Following OTU clustering, chimera checking is performed to remove a common PCR artifact. Chimeras are PCR products that originate from at least two different origins. Chimeras can be identified by running the clustered OTUs against a reference database (Edgar et al. 2011; Cuesta-Zuluaga and Escobar 2016). For taxonomic classification one representative sequence per OTU is then run against a reference database (Coissac et al. 2012; Ji et al. 2013).

For these different steps, a variety of tools with different implementations

of the underlying algorithms have already been developed (Cuesta-Zuluaga and Escobar 2016) and further will probably be developed in future. These tools often differ in their outcome and are generally often imperfect (Cuesta-Zuluaga and Escobar 2016). Thus, it is difficult to choose the most desirable tools for data processing.

Insufficient species resolution. DNA-barcoding can be applied to any type of organism and the same is true for meta-barcoding. However, in single specimen DNA-barcoding, sequences of up to 1,000 base pairs can be obtained, because Sanger sequencing is employed . This can cover the complete bacterial 16S rRNA gene or the animal COI gene (cf. Coissac et al. 2012), which is sufficient for resolving species. In meta-barcoding however, read lengths of high-throughput sequencers are shorter, e.g. 2x300bp for Illumina MiSeq (see www.illumina.com, accessed 20/09/2016). This might not be enough to cover sufficient sequence differences for species distinction (Coissac et al. 2012), thus species resolution can be insufficient in meta-barcoding.

This is especially the case for the animal kingdom, because the commonly used COI gene for species identification is longer than current read lengths, although special meta-barcoding primers have been developed based on other genes (Epp et al. 2012; Pompanon et al. 2012). In microbiology, the 16S rRNA gene is only partially sequenced for bacterial meta-barcoding (Cuesta-Zuluaga and Escobar 2016).

Incomplete reference databases. Considering species resolution, and taxonomic classification in general based on DNA-barcoding, it is not only the genetic marker that influences the success of assigning taxonomy to a sequence, but the quality of the underlying database is also essential (see Publications P.1; P.2; P.3; Nilsson et al. 2006; Pompanon et al. 2012). A good database contains reference sequences covering a broad range of organisms that have been obtained from correctly identified specimens. However, some commonly used databases contain incorrectly assigned sequences (Nilsson et al. 2006; Epp et al. 2012; Pompanon et al. 2012), which can lead to misclassifications. The importance of a good reference database has been stressed in the pollen analysis study, where previously undetected taxa could be identified after a database update that included additional taxa (Publication P.2).

Yet, most classification algorithms can deal with incomplete database coverage by assigning taxonomy hierarchically and determining confidence values for each assignment as well (Cuesta-Zuluaga and Escobar 2016). So, in the case of incomplete databases, taxonomic classification will be as-

signed on higher taxonomic levels only. Whilst this can be sufficient in some studies, constant database improvements are called for.

On a different note, incomplete database coverage can be used to identify target taxa for more in depth studies. For example Verrucomicrobia seem to be globally prevalent and abundant (Handelsman and Smalla 2003), as identified using meta-barcoding, but are not well studied (Cao et al. 2015) and thus constitute a focus phylum for future database improvements.

Abundace estimation. Abundance estimation of taxa is highly limited in meta-barcoding. Several studies have assessed the suitability of HTS data for abundance estimation, with contrasting results (Murray et al. 2011; Keller et al. 2014; Kraaijeveld et al. 2015; Richardson et al. 2015). The inaccuracy for abundance estimation is caused by two underlying mechanisms: firstly, different copy numbers per cell of the selected marker gene among different species (Prokopowich et al. 2003; Cuesta-Zuluaga and Escobar 2016) and secondly, introducing unknown and difficult to estimate biases or skews by each step of library preparation (Fierer et al. 2008; Cuesta-Zuluaga and Escobar 2016).

In the case of difference in gene copy numbers, it can be possible to account for when the copy number of the taxon in question is known (Langille et al. 2013). However, this again requires a good underlying database and especially one that contains this exact information. Currently, this is only possible for taxa with completely sequenced genomes (Cuesta-Zuluaga and Escobar 2016).

The unknown bias introduced by sample processing, however, poses a greater challenge. It has been shown that each step of the process, from DNA isolation to sequencing, can influence the outcome of a high-throughput biodiversity study (Fierer et al. 2008; Cuesta-Zuluaga and Escobar 2016), complicating abundance estimation as well as study comparability. It is thus advisable, to create a laboratory workflow with as few steps as possible (see Publication P.2) and stick to the exact same workflow for study comparability. One approach to estimate the skew introduced by sample processing can be to include an internal control early in the process (Smets et al. 2015). Whilst this is certainly a good approach, it is limited to well-studied systems, because it relies on adding genomic DNA of a taxon that is known to not or only very rarely occur in the study system (Smets et al. 2015). In systems that are less well-studied, such as carnivorous plants (see Publication P.5), it is difficult to choose an appropriate organism.

In the raw data, abundance is estimated as sequence counts, which is often

not correlated with actual taxon abundance. Many studies, including publications P.2, P.4 and P.5 relativise the count data for abundance estimation, but this should still be handled with care, since relative abundance is not correlated with biomass when dealing with multicellular organisms (Pompanon et al. 2012). One approach to compare abundance of taxa between samples can be a differential approach (see Publication P.5; Robinson et al. 2010), inherited from gene expression studies. However, this still does not give an estimate of the actual abundance, but can only infer differences in taxon abundance between samples or states. Another approach for better abundance estimates, often taken in microbiology, can be quantitative PCR (qPCR; Redford and Fierer 2009; Hollister et al. 2010; Murray et al. 2011), although this might also be affected by copy number variation of the marker gene. Despite the different approaches to improve abundance estimation based on HTS data, it remains advisable to additionally analyse at least a subset of samples with an alternative method, which is more robust in abundance estimation, as proposed previously (Pompanon et al. 2012; Keller et al. 2015) and in the pollen meta-barcoding establishment (Publication P.2).

Functional profiling. One major drawback in meta-barcoding is that from the taxonomic profiles obtained it is not possible to infer functionality of the taxa detected, which would be the case with metagenomics. This is especially the case in bacterial community analysis. One approach to infer functions of the microbiota is to identify common functionalities of a certain group of bacteria from literature, as has been done in publication P.5. For example, members of the Burkholderiales or Rhizobiales are commonly reported as nitrogen fixing taxa (Vessey 2003; Lucy et al. 2004). Thus, detecting such taxa in a bacterial community can imply that nitrogen fixation is part of the functional profile of the community. However, this can only potentially be the case, for several reasons. Firstly, it is possible that there are taxa within those lineages that are not able to fix nitrogen. Secondly, it is possible that bacteria possess nitrogen fixation genes, but do not express them within a given community. In publication P.5, the bacterial community was screened for putative nitrogen fixers, but it was not possible to conclude whether this function was part of the host-bacteria association.

Another approach for functional profiling is the use of recently developed tools that infer potential functionality based on the 16S profile (Langille et al. 2013; Keller et al. 2014). These tools screen databases of sequenced bacterial genomes for close relatives of the taxa identified through meta-barcoding and infer probable functions based on phylogenetic relationships. Additionally, it is possible to estimate uncertainty of the reported functions (Langille et al. 2013). This approach has an advantage over metagenomics, because it is possible to map functionality to a specific OTU (Langille et al.

2013). Thus, in cases where metagenomics is not an option, this approach is a feasible alternative. However, it is still only a putative functional profile, whilst metagenomics truly characterises the availability of functions in a bacterial community. However, to obtain true functional profiles, i.e. estimate which functionality a bacterial community actually exhibits in a given state, other approaches, such as (meta-)transcriptomics, bioassays or directed experimental approaches are required.

Recommendations for the application of meta-barcoding in ecological research

Considering the pros and cons discussed above, it needs to be considered how meta-barcoding can be applied in ecological research. The two major advantages of meta-barcoding are that species assemblages can be studied and members identified without considering morphological characteristics and without taxonomic expertise. The high throughput of modern sequencing platforms enables large-scale studies, because a high number of samples can be processed simultaneously, e.g. 384 samples in the present studies (see Publications P.2 and P.3; Kozich et al. 2013), whilst working hours and costs per sample are dramatically reduced compared to manual analysis of species assemblages. Additionally, processed and degraded material can be studied and classified successfully.

For ecologists, this means that large-scale biodiversity studies can be performed at reasonable costs, whilst results can be obtained in a timely manner and meta-barcoding can thus be applied to a variety of research questions. Examples where material is degraded or processed and classification based on morphology are very difficult include diet analysis and food safety issues. Analysis of pollen assemblages, for example in agro-ecology and palaeo-ecology, is challenging, because pollen grains need to be manually separated and closely-related species cannot be distinguished by pollen grain morphology. The application of meta-barcoding in such studies thus constitutes a suitable and powerful alternative to traditional approaches.

In microbial ecology, meta-barcoding has become a common approach for bacterial community analysis, because it avoids the need to bring bacteria into culture. Thus, it allows studying complete bacterial communities, rather than introducing bias by cultivation media. Currently, microbial community research focuses on creating an overview of the global diversity and distribution of microbial taxa and it is sufficient to identify dominant and low abundance taxa, rather than estimating their actual abundance.

Another great advantage of meta-barcoding is the potential for sample multiplexing together with automatic sample and data processing, which is highly beneficial in biodiversity monitoring, for example through routine eDNA analysis. This is also beneficial for large-scale studies or when a large amount of samples is collected over a considerable amount of time before they can be processed.

However, if abundance estimation is a central aspect of the research question, meta-barcoding alone might yield skewed data. In cases like these, it remains advisable to combine meta-barcoding with other methods. For example, for bee foraging studies, combining meta-barcoding for plant species identification of the collected pollen with light microscopy of at least a subset of the samples for abundance estimation would exploit the advantages of both methods.

Generally, combining different approaches with meta-barcoding can be highly beneficial. For example, for functional profiling, a subset of samples can be analysed using (meta-)genomics or (meta-)transcriptomics to complement the taxonomic profiles obtained with meta-barcoding. Other complementing methods for functional analysis can include bio-assays of the complete community (Kaur et al. 2009) or of selected bacterial taxa (Sugahara et al. 2015).

Another aspect is that sometimes and for some taxonomic groups, species resolution obtained with meta-barcoding is still not sufficient. This is due to incomplete databases, imperfect marker genes coupled with short read lengths, but also classification algorithms. Whilst there are many studies where the current species resolution is sufficient, for example prime studies of bacterial communities (Junker et al. 2011; Keller et al. 2013; Kueneman et al. 2013), there are applications where the distinction of closely related taxa is of utmost importance, for examples in allergen load assessment (Kraaijeveld et al. 2015) or for the enforcement of conservation efforts (Ji et al. 2013; Hajibabaei et al. 2016). Thus, when devising a study based on meta-barcoding, it is important to determine, whether the expected species resolution that can be obtained is sufficient for the specific purpose.

Future aspects

Some of the above-mentioned drawbacks are not limitations of the method per sé, but rather logisitcal and technological limitations, such as read lengths still being too short for sufficient species resolution or incomplete databases. Thus, with database improvements and technological advances, species

resolution will improve in future. Additionally, more research into copy number variation of marker genes together with an increase in whole genome data will improve abundance estimates and functional prediction with meta-barcoding. Further technological advances and reductions in the costs of high-throughput sequencing will facilitate increased sample sizes, but might also allow simultaneous sequencing of multiple genetic markers, which would facilitate the study of species interactions across kingdoms. For bacterial community studies, a shift from meta-barcoding to metagenomics can be expected with further decreases of sequencing costs, because both taxonomic and functional profiles can be obtained from the same data set.

On a different note, data processing and analysis might always pose a challenge, especially for scientists without bioinformatic training. The existing tools for data processing utilise various algorithms, which might be difficult to understand and implement by the majority of researchers. It can also be expected that these algorithms and tools will be further developed, which will probably lead to improvements, but will remain to be a challenge for untrained scientists. This actually does not only apply to meta-barcoding, but to all HTS applications, including (meta-)genomics and (meta-)transcriptomics. Thus, appropriate training needs to be implemented in the education of future researchers, whilst bioinformatic tools need to be straightforward in their implementation.

It has been mentioned above that each processing step can affect the outcome of meta-barcoding results and that these effects are difficult to estimate. It is thus recommended to develop standard protocols for sampling, library preparation and data processing. This is especially important for the studies within international consortia, but will also facilitate study comparability. Such protocols need to be straightforward to implement and suitable for laboratories with basic equipment only. At the same time, the standard protocols should facilitate the automation of sample processing to make use of the high throughput potential of meta-barcoding.

Conclusion

Meta-barcoding, as every methodology, has its merits and pitfalls, all of which need to be considered at every step of research, from study and sample design, to laboratory and data processing to data analysis. The great powers of meta-barcoding lie in the species identification of individuals within assemblages and the high throughput of sequences and samples. Additionally, since taxonomy is assigned via DNA-barcoding, taxonomy experts are not required, which is highly beneficial concerning the lack of these

(Gaston and May 1992; Carvalho et al. 2007). Challenges are associated with read lengths as well as data processing, the sometimes insufficient species resolution, inaccurate abundance estimation and limited functional profiling. Whilst some of the problems are expected to improve in future it remains advisable to combine meta-barcoding with other methodological approaches. In conclusion, meta-barcoding is at the moment the gold standard for high-throughput analysis of specimen assemblages. The potential to automate meta-barcoding holds great promises for routine biodiversity monitoring, amongst other applications.

Bibliography

Abe, S et al. (1967). “Taxonomical studies on glutamic acid-producing bacteria”. In: *Journal of General and Applied Microbiology* 13.3, pp. 279–301.

Abrahamian, F M and E J C Goldstein (2011). “Microbiology of animal bite wound infections”. In: *Clinical Microbiolgy Reviews* 24, pp. 231–246.

Altschul, Stephen F et al. (1990). “Basic local alignment search tool”. In: *Journal of molecular biology* 215, pp. 403–410.

Amagase, Shizuko et al. (1972). “Digestive enzymes in insectivorous plants. IV. Enzymatic digestion of insects by Nepenthes secretion and drosera peltata extract: proteolytic and chitinolytic activities”. In: *Journal of Biochemistry* 72, pp. 765–767.

An, Chung-Il et al. (2001). “Plasma-membrane H+-ATPases are expressed in pitchers of the carnivorous plant Nepenthes alata Blanco”. In: *Planta* 212, pp. 547–555.

An, Chung-Il et al. (2002). “Aspartic proteinases are expressed in pitchers of the carnivorous plant Nepenthes alata Blanco”. In: *Planta* 214, pp. 661–667.

Anes, Elsa et al. (2003). “Selected lipids activate phagosome actin assembly and maturation resulting in killing of pathogenic mycobacteria”. In: *Nature Cell Biology* 5.9, pp. 793–802.

Ankenbrand, Markus Johannes et al. (2015). “The ITS2 database V – Twice as much”. In: *Molecular Biology and Evolution* 32.11, pp. 3030–3032.

Aronesty, Erik (2011). *ea-utils: ”Command-line tools for processing biological sequencing data”*.

Arroyo, O et al. (1980). “The bacterial flora of venoms and mouth cavities of Costa Rican snakes”. In: *Bulletin of the Pan American Health Organization* 14, pp. 280–285.

Avila, M et al. (2009). “The oral microbiota: living with a permanent guest”. In: *DNA and Cell Biology* 28, pp. 405–411.

Bagley, ST (1985). “Habitat association of Klebsiella species”. In: *Infection Control* 6.2, pp. 52–58.

Balakrishnan, Rohini (2005). “Species concepts, species boundaries and species identification: a view from the tropics.” In: *Systematic biology* 54.4, pp. 689–693.

Bálint, M et al. (2014). “An Illumina metabarcoding pipeline for fungi”. In: *Ecology and Evolution* 4, pp. 2642–2653.

Bauer, Ulrike et al. (2009). “Effect of pitcher age on trapping efficiency and natural prey capture in carnivorous Nepenthes rafflesiana plants”. In: *Annals of Botany* 103, pp. 1219–1226.

Bauer, Ulrike et al. (2011). “Evidence for alternative trapping strategies in two forms of the pitcher plant, Nepenthes rafflesiana”. In: *Journal of Experimental Botany* 62.10, pp. 3683–3692.

Bazile, Vincent et al. (2015). “Fluid physico-chemical properties influence capture and diet in Nepenthes pitcher plants”. In: *Annals of botany* 115, pp. 705–716.

Behling, Hermann et al. (2004). “Late Quaternary Araucaria forest, grassland (Campos), fire and climate dynamics, studied by high-resolution pollen, charcoal and multivariate analysis of the Cambará do Sul core in southern Brazil”. In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 203, pp. 277–297.

Beil, Marion et al. (2008). “Analysis of pollen loads in a wild bee community (Hymenoptera : Apidae) - a method for elucidating habitat use and foraging distances”. In: *Apidologie* 39, pp. 456–467.

Bell, Karen L. et al. (2016). “Pollen DNA barcoding: current applications and future prospects”. In: *Genome* 59, pp. 1–12.

Belova, Svetlana E et al. (2009). “Acidisoma tundrae gen. nov., sp. nov. and Acidisoma sibiricum sp. nov., two acidophilic, psychrotolerant members of the Alphaproteobacteria from acidic northern wetlands”. In: *International journal of Systematic and Evolutionary Microbiology* 59, pp. 2283–2290.

Bennett, K D and L Parducci (2006). “DNA from pollen: principles and potential”. In: *The Holocene* 16.8, pp. 1031–1034.

Benson, Dennis A et al. (2013). “GenBank”. In: *Nucleic Acids Research* 41.Database issue, pp. D36–D42.

Bent, S J and L J Forney (2008). “The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity”. In: *The ISME Journal* 2, pp. 689–695.

Bermudes, David and David H Benzing (1991). “Nitrogen fixation in association with Ecuadorean bromeliads”. In: *Journal of Tropical Ecology* 7, pp. 531–536.

Bhore, S J et al. (2013). “Diversity of endophytic bacteria in medicinally important Nepenthes species”. In: *Journal of Natural Science, Biology and Medicine* 4.2, pp. 431–434.

Binladen, Jonas et al. (2007). “The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing”. In: *PLoS ONE* 2, e197.

Blatrix, Rumsaïs et al. (2012). “Plant-ants use symbiotic fungi as a food source: new insight into the nutritional ecology of ant-plant interactions.” In: *Proceedings of the Royal Society B* 279, pp. 3940–3947.

Blaxter, Mark (2016). “Imagining Sisyphus happy: DNA barcoding and the unnamed majority”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371, p. 20150329.

Blaylock, R S M (2001). “Normal oral bacterial flora from some southern African snakes”. In: *Onderstepoort Journal of Veterinary Research* 68, pp. 175–182.

Boiero, L. et al. (2007). “Phytohormone production by three strains of Bradyrhizobium japonicum and possible physiological and technological implications”. In: *Applied Microbiology and Biotechnology* 74, pp. 874–880.

Bonhomme, Vincent et al. (2011). “Slippery or sticky? Functional diversity in the trapping strategy of Nepenthes carnivorous plants”. In: *New Phytologist* 191, pp. 545–554.

Bragina, A et al. (2012). “Similar diversity of Alphaproteobacteria and nitrogenase gene amplicons on two related Sphagnum mosses”. In: *Frontiers in Microbiology* 2, pp. 1–10.

Bray, J R and T Curtis (1957). “An ordination of upland forest communities of Southern Wisconsin”. In: *Ecological Monographs* 27.4, pp. 325–349.

Bruni, I et al. (2015). “A DNA barcoding approach to identify plant species in multiflower honey”. In: *Food Chemistry* 170, pp. 308–315.

Buch, Franziska et al. (2013). “Secreted pitfall-trap fluid of carnivorous Nepenthes plants is unsuitable for microbial growth”. In: *Annals of Botany* 111, pp. 375–383.

Cao, Hieu X. et al. (2015). “Metatranscriptome analysis reveals host-microbiome interactions in traps of carnivorous Genlisea species”. In: *Frontiers in Microbiology* 6, p. 526.

Caporaso, J Gregory et al. (2010). “QIIME allows analysis of high- throughput community sequencing data”. In: *Nature Methods* 7.5, pp. 335–336.

Caporaso, J Gregory et al. (2011). “Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample”. In: *Proceedings of the National Academy of Sciences* 108.Suppl. 1, pp. 4516–4522.

Carvalho, Marcelo R. de et al. (2007). “Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm”. In: *Evolutionary Biology* 34.3, pp. 140–143.

Carvell, Claire et al. (2006). “Assessing the value of annual and perennial forage mixtures for bumblebees by direct observation and pollen analysis”. In: *Apidologie* 37.3, pp. 326–340.

Chen, Bao et al. (2014). “The endophytic bacterium, Sphingomonas SaMR12, improves the potential for zinc phytoremediation by its host, Sedum alfredii”. In: *PloS one* 9.9, e106826.

Chen, Shilin et al. (2010). “Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species”. In: *PloS one* 5.1, e8613.

Chou, Lee Yiung et al. (2014). “Bacterial communities associated with the pitcher fluids of three Nepenthes (Nepenthaceae) pitcher plant species growing in the wild.” In: *Archives of Microbiology* 196.10, pp. 709–717.

Clarke, C et al. (2011). “Nepenthes baramensis (Nepenthaceae) - a new species from north-western Borneo”. In: *Blumea* 56, pp. 229–233.

Clarke, K R (1993). “Non-parametric multivariate analyses of changes in community structure”. In: *Australian Journal of Ecology* 18, pp. 117–143.

Clarke, M C (1997). “Nepenthes of Borneo”. In: *Natural History Publications*. Malaysia, Sabah, Kota Kinabalu.

Clarke, M C (2001). “Nepenthes of Sumatra and Peninsular Malaysia”. In: *Natural History Publications2*. Malaysia, Sabah, Kota Kinabalu.

Coghlan, Megan L. et al. (2012). “Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns”. In: *PLoS Genetics* 8.4, e1002657.

Coissac, Eric et al. (2012). “Bioinformatic challenges for DNA metabarcoding of plants and animals”. In: *Molecular Ecology* 21, pp. 1834–1847.

Cosma, Christine L et al. (2003). “The secret lives of the pathogenic mycobacteria”. In: *Annual Review of Microbiology* 57, pp. 641–676.

Cuesta-Zuluaga, Jacobo de la and Juan S Escobar (2016). “Considerations for optimizing microbiome analysis using a marker gene”. In: *Frontiers in Nutrition* 3, p. 26.

Davies, Althea L and Richard Tipping (2004). “Sensing small-scale human activity in the palaeoecological record: fine spatial resolution pollen analyses from Glen Affric, northern Scotland”. In: *The Holocene* 14.2, pp. 233–245.

Dixon, Philip (2003). “VEGAN, a package of R functions for community ecology”. In: *Journal of Vegetation Science* 14, pp. 927–930.

Dröge, Stefan et al. (2008). “Treponema isoptericolens sp. nov., a novel spirochaete from the hindgut of the termite Incisitermes tabogae”. In: *International Journal of Systematic and Evolutionary Microbiology* 58, pp. 1079–1083.

Dupuis, J et al. (2012). “Multi-locus species delimitation in closely related animals and fungi: one marker is not enough”. In: *Molecular Ecology* 21, pp. 4422–4436.

Edberg, Stephen C et al. (1986). “Phenotypic characteristics of coliform and noncoliform bacteria from a public water supply compared with regional

and national clinical species". In: *Applied and Environmental Microbiology* 52.3, pp. 474–478.

Edgar, R C (2010). "Search and clustering orders of magnitude faster than BLAST". In: *Bioinformatics* 26.19, pp. 2460–2461.

Edgar, Robert C (2013). "UPARSE: highly accurate OTU sequences from microbial amplicon reads". In: *Nature Methods* 10.10, pp. 996–998.

Edgar, Robert C et al. (2011). "UCHIME improves sensitivity and speed of chimera detection". In: *Bioinformatics* 27.16, pp. 2194–2200.

Eilenberg, Haviva et al. (2006). "Isolation and characterization of chitinase genes from pitchers of the carnivorous plant Nepenthes khasiana." In: *Journal of Experimental Botany* 57.11, pp. 2775–2784.

Endo, A et al. (2013). "Lactobacillus faecis sp. nov., isolated from animal faeces". In: *International Journal of Systematic and Evolutionary Microbiology* 63, pp. 4502–4507.

Epp, Laura S. et al. (2012). "New environmental metabarcodes for analysing soil DNA: Potential for studying past and present ecosystems". In: *Molecular Ecology* 21, pp. 1821–1833.

Federhen, Scott (2012). "The NCBI Taxonomy database." In: *Nucleic acids research* 40.Database issue, pp. D136–43.

Ferreira Junior, R S et al. (2009). "Comparison of wildlife and captivity rattlesnakes (Crotalus durissus terrificus) microbiota". In: *Pesquisa Veterinária Brasileira* 29.999-1003.

Fierer, Noah and R B Jackson (2006). "The diversity and biogeography of soil bacterial communities". In: *Proceedings of the National Academy of Sciences* 103.3, pp. 626–631.

Fierer, Noah et al. (2008). "The influence of sex, handedness, and washing on the diversity of hand surface bacteria". In: *Proceedings of the National Academy of Sciences* 105.46, pp. 17994–17999.

Fonseca, M G et al. (2009). "Oral microbiota of Brazilian captive snakes". In: *Journal of Venomous Animals and Toxins including Tropical Diseases* 15, pp. 54–60.

Fox, G E et al. (1977). "Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to procaryotic systematics". In: *International Journal of Systematic and Evolutionary Microbiology* 27, pp. 44–57.

Galimberti, Andrea et al. (2014). "A DNA barcoding approach to characterize pollen collected by honeybees". In: *PloS one* 9.10, e109363.

Garg, A et al. (2009). "Wound infections secondary to snakebite". In: *Journal of Infection in Developing Countries* 3, pp. 221–223.

Gaston, Kevin J. and Robert M. May (1992). "Taxonomy of taxonomists". In: *Nature* 356, pp. 281–282.

Gathmann, Achim and Teja Tscharntke (2002). “Foraging ranges of solitary bees”. In: *Journal of Animal Ecology* 71, pp. 757–764.

Gaume, Laurence and Bruno Di Giusto (2009). “Adaptive significance and ontogenetic variability of the waxy zone in Nepenthes rafflesiana.” In: *Annals of botany* 104.7, pp. 1281–91.

Gaume, Laurence and Yoel Forterre (2007). “A viscoelastic deadly fluid in carnivorous pitcher plants”. In: *PloS one* 2.11, e1185.

Gaume, Laurence et al. (2002). “Function of epidermal surfaces in the trapping efficiency of Nepenthes alata pitchers”. In: *New Phytologist* 156, pp. 479–489.

Gaume, L et al. (2004). “How do plant waxes cause flies to slide? Experimental tests of wax-based trapping mechanisms in three pitfall carnivorous plants”. In: *Arthropod Structure & Development* 33.1, pp. 103–111.

Gerner-Smidt, Peter et al. (1991). “Reliability of phenotypic tests for identification of Acinetobacter species”. In: *Journal of Clinical Microbiology* 29.2, pp. 277–282.

Goldstein, E J C et al. (1979). “Bacteriology of rattlesnake venom and implications for therapy”. In: *Journal of Infectious Diseases* 140, pp. 818–821.

Goldstein, E J C et al. (1981). “Aerobic bacterial oral flora of garter snakes: development of normal flora and pathogenic potential for snakes and humans”. In: *Journal of Clinical Microbiology* 13, pp. 954–956.

Graber, Joseph R et al. (2004). “Description of Treponema azotonutricium sp. nov. and Treponema primitia sp. nov., the first spirochetes isolated from termite guts”. In: *Applied and Environmental Microbiology* 70.3, pp. 1315–1320.

Grafe, T Ulmar et al. (2011). “A novel resource-service mutualism between bats and pitcher plants”. In: *Biology Letters* 7, pp. 436–439.

Gu, S et al. (2013). “Bacterial community mapping of the mouse gastrointestinal tract”. In: *PloS one* 8, e74957.

Gugerli, Felix et al. (2005). “Ancient plant DNA: review and prospects”. In: *New Phytologist* 166, pp. 409–418.

Hajibabaei, Mehrdad et al. (2016). “A new way to contemplate Darwin’s tangled bank: how DNA barcodes are reconnecting biodiversity science and biomonitoring”. In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 371, p. 20150330.

Handelsman, Jo and Kornelia Smalla (2003). “Conversations with the silent majority”. In: *Current Opinion in Microbiology* 6.3, pp. 271–273.

Harris, J. Kirk et al. (2010). “Comparison of normalization methods for construction of large, multiplex amplicon pools for next-generation sequencing”. In: *Applied and Environmental Microbiology* 76.12, pp. 3863–3868.

Hatano, Naoya and Tatsuro Hamada (2008). “Proteome analysis of pitcher fluid of the carnivorous plant Nepenthes alata”. In: *The Journal of Proteome Research* 7, pp. 809–816.

Hatano, Naoya and Tatsuro Hamada (2012). “Proteomic analysis of secreted protein induced by a component of prey in pitcher fluid of the carnivorous plant Nepenthes alata”. In: *Journal of Proteomics* 75, pp. 4844–4852.

Hebert, P. D. et al. (2003). “Biological identifications through DNA barcodes”. In: *Proceedings of the Royal Society B: Biological Sciences* 270, pp. 313–321.

Higashi, Shiro et al. (1993). “Analysis of feeding mechanism in a pitcher of Nepenthes hybrida”. In: *Journal of Plant Research* 106, pp. 47–54.

Hill, M O and H G Gauch Jr (1980). “Detrended correspondence analysis: an improved ordination technique”. In: *Vegetatio* 42, pp. 47–58.

Hollingsworth, Peter M et al. (2011). “Choosing and using a plant DNA barcode”. In: *PloS one* 6.5, e19254.

Hollister, Emily B et al. (2010). “Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments.” In: *The ISME journal* 4, pp. 829–838.

Holmes, B. et al. (1994). “Evaluation of two BBL crystal systems for identification of some clinically important gram-negative bacteria”. In: *Journal of Clinical Microbiology* 32.9, pp. 2221–2224.

Howard, S T and T F Byrd (2000). “The rapidly growing mycobacteria: saprophytes and parasites”. In: *Microbes and Infection* 2, pp. 1845–1853.

Illumina Inc. (2013). *Preparing libraries for sequencing on the MiSeq*. Tech. rep.

Ji, Yinqiu et al. (2013). “Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding”. In: *Ecology Letters* 16.10, pp. 1245–1257.

Jones, R M et al. (2013). “Acidocella aromatica sp. nov.: an acidophilic heterotrophic alphaproteobacterium with unusual phenotypic traits”. In: *Extremophiles* 17.5, pp. 841–850.

Juniper, B E et al. (1989). *The carnivorous plants*. London: Academic Press.

Junker, R R and A Keller (2015). “Microhabitat heterogeneity across leaves and flower organs promotes bacterial diversity”. In: *FEMS Microbiology Ecology* 91, fiv097.

Junker, R R et al. (2011). “Composition of epiphytic bacterial communities differs on petals and leaves”. In: *Plant Biology* 13, pp. 918–924.

Kaur, Harleen et al. (2009). “Taking ecological function seriously: Soil microbial communities can obviate allelopathic effects of released metabolites”. In: *PLoS ONE* 4.3, pp. 1–6.

Keller, Alexander et al. (2009). “5.8S-28S rRNA interaction and HMM-based ITS2 annotation”. In: *Gene* 430, pp. 50–57.

Keller, Alexander et al. (2014). “Computational integration of genomic traits into 16S rDNA microbiota sequencing studies”. In: *Gene* 549.1, pp. 186–191.

Keller, A et al. (2013). “Diverse microbiota identified in whole intact nest chambers of the red mason bee Osmia bicornis”. In: *PloS one* 8.10, e78296.

Keller, A et al. (2015). “Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples”. In: *Plant Biology* 17.2, pp. 558–566.

Kersters, K et al. (2006). “The family Acetobacteraceae: the genera Acetobacter, Acidimonas, Asaia, Gluconacetobacter, Gluconobacter, and Kozakia”. In: *The Prokaryotes: Volume 5: Proteobacteria: Alpha and Beta Subclasses*. Ed. by Martin Dworkin et al. Third Edit. Springer, pp. 163–200.

Kimoto, K I et al. (2010). “Acidocella aluminiidurans sp. nov., an aluminium-tolerant bacterium isolated from Panicum repens grown in a highly acidic swamp in actual acid sulfate soil area of Vietnam”. In: *International Journal of Systematic and Evolutionary Bacteriology* 60, pp. 764–768.

Kishimoto, N et al. (1995). “Transfer of Acidiphilium facilis and Acidiphilium aminolytica to the Genus Acidecella gen. nov., and Emendation of the Genus Acidiphilium”. In: *Systematic and Applied Microbiology* 18, pp. 85–91.

Koetschan, Christian et al. (2010). “The ITS2 Database III–sequences and structures for phylogeny.” In: *Nucleic Acids Research* 38.Database issue, pp. D275–D279.

Koopman, Margaret M et al. (2010). “The carnivorous pale pitcher plant harbors diverse, distinct, and time-dependent bacterial communities”. In: *Applied and Environmental Microbiology* 76.6, pp. 1851–1860.

Köppler, Kirsten et al. (2007). “Comparison of pollen spectra collected by four different subspecies of the honey bee Apis mellifera”. In: *Apidologie* 38, pp. 341–353.

Kozich, James J et al. (2013). “Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform”. In: *Applied and Environmental Microbiology* 79.17, pp. 5112–5120.

Kraaijeveld, Ken et al. (2015). “Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing”. In: *Molecular Ecology Resources* 15, pp. 8–16.

Krupke, Christian H et al. (2012). “Multiple routes of pesticide exposure for honey bees living near agricultural fields”. In: *PloS one* 7.1, e29268.

Krych, L et al. (2013). “Quantitatively different, yet qualitatively alike: a meta-analysis of the mouse core gut microbiome with a view towards the human gut microbiome”. In: *PloS one* 8, e62578.

Kueneman, Jordan G et al. (2013). “The amphibian skin-associated microbiome across species, space and life history stages”. In: *Molecular ecology* 23.6, pp. 1238–1250.

Lam, K K et al. (2010). “A cross-sectional survey of snake oral bacterial flora from Hong Kong, SAR, China.” In: *Emergency Medicine Journal* 28, pp. 107–114.

Langille, Morgan G I et al. (2013). “Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences.” In: *Nature Biotechnology* 31, pp. 814–821. arXiv: NIHMS150003.

Lauber, Christian L. et al. (2009). “Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale”. In: *Applied and Environmental Microbiology* 75.15, pp. 5111–5120.

Lee, J E et al. (2011). “Analysis of human and animal fecal microbiota for microbial source tracking”. In: *The ISME Journal* 5, pp. 362–365.

Ley, Ruth E et al. (2008). “Evolution of mammals and their gut microbes”. In: *Science* 320, pp. 1647–1651.

Li, Ming et al. (2011). “Identification of herbal medicinal materials using DNA barcodes”. In: *Journal of Systematics and Evolution* 49.3, pp. 271–283.

Lim, Y. S. et al. (2015). “How a pitcher plant facilitates roosting of mutualistic woolly bats”. In: *Evolutionary Ecology Research* 16, pp. 1–11.

Loudon, Andrew H et al. (2014). “Microbial community dynamics and effect of environmental microbial reservoirs on red-backed salamanders (Plethodon cinereus)”. In: *The ISME journal* 8, pp. 830–840.

Lozupone, Catherine A and Rob Knight (2005). “UniFrac: a new phylogenetic method for comparing microbial communities”. In: *Applied and Environmental Microbiology* 71.12, pp. 8228–8235.

Lucy, M et al. (2004). “Applications of free living plant growth-promoting rhizobacteria”. In: *Antonie van Leeuwenhoek* 86, pp. 1–25.

Lüttge, Ulrich (1964). “Untersuchungen zur Physiologie der Carnivoren-Drüsen”. In: *Planta* 63, pp. 103–117.

Mata, Juan Antonio et al. (2002). “A detailed phenotypic characterisation of the type strains of Halomonas species”. In: *Systematic and Applied Microbiology* 25, pp. 360–375.

Mayer, E. et al. (2005). “Microflora in the traps of pitcher plants”. In: *VXII International Botanical Congress*. July, p. 510.

McMurdie, Paul J and Susan Holmes (2013). “phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data”. In: *PloS one* 8.4, e61217.

McMurdie, Paul J. and Susan Holmes (2014). “Waste not, want not: why rarefying microbiome data is inadmissible”. In: *PLoS Computational Biology* 10.4, e1003531. arXiv: 1310.0424.

Mithöfer, Axel (2011). “Carnivorous pitcher plants: insights in an old topic”. In: *Phytochemistry* 72, pp. 1678–1682.

Moran, Jonathan A et al. (2010). “Ion fluxes across the pitcher walls of three Bornean Nepenthes pitcher plant species: flux rates and gland distribution patterns reflect nitrogen sequestration strategies”. In: *Journal of Experimental Botany* 61.5, pp. 1365–1374.

Morin, P J (1999). *Community ecology*. Malden: Blackwell Science.

Morotomi, M et al. (2002). “Lactobacillus equi sp. nov., a predominant intestinal Lactobacillus species of the horse isolated from faeces of healthy horses”. In: *International Journal of Systematic and Evolutionary Microbiology* 52, pp. 211–214.

Mouquet, N. et al. (2008). “Modelling the relationship between a pitcher plant (Sarracenia purpurea) and its phytotelma community: mutualism or parasitism?” In: *Functional Ecology* 22, pp. 728–737.

Mullins, J and J Emberlin (1997). “Sampling pollens”. In: *Journal of Aerosol Science* 28.3, pp. 365–370.

Murray, Dáithí C. et al. (2011). “DNA-based faecal dietary analysis: A comparison of qPCR and high throughput sequencing approaches”. In: *PLoS ONE* 6.10, e25776.

Murray, Dáithí C et al. (2012). “High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens”. In: *Quaternary Science Reviews* 58, pp. 135–145.

Nilsson, R. Henrik et al. (2006). “Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective”. In: *PloS one* 1, e59.

Nord, Carl-Erik et al. (1974). “Evaluation of five test-kits - API , AuxoTab , Enterotube , PathoTec and R/B - for identification of Enterobacteriaceae”. In: *Medical Microbiology and Immunology* 159, pp. 211–220.

Odumeru, Joseph A et al. (1999). “Evaluation of accuracy and repeatability of identification of food-borne pathogens by automated bacterial identification systems”. In: *Journal of Clinical Microbiology* 37.4, pp. 944–949.

Ogden, Rob et al. (2009). “Wildlife DNA forensics - Bridging the gap between conservation genetics and law enforcement”. In: *Endangered Species Research* 9, pp. 179–195.

Oram, Diana Marra et al. (2004). “Analysis of genes that encode DtxR-like transcriptional regulators in pathogenic and saprophytic corynebacterial species”. In: *Infection and Immunology* 72.4, pp. 1885–1895.

O’Shea, Eileen F et al. (2009). “Characterization of enterocin- and salivaricin-producing lactic acid bacteria from the mammalian gastrointestinal tract”. In: *FEMS Microbiology Letters* 291, pp. 24–34.

Parducci, L et al. (2005). “Ancient DNA from pollen: a genetic record of population history in Scots pine”. In: *Molecular Ecology* 14, pp. 2873–82.

Pavlovič, Andrej et al. (2007). “Carnivorous syndrome in Asian pitcher plants of the genus Nepenthes”. In: *Annals of botany* 100.3, pp. 527–536.

Peterson, Celeste N et al. (2008). “A keystone predator controls bacterial diversity in the pitcher-plant (Sarracenia purpurea) microecosystem.” In: *Environmental Microbiology* 10.9, pp. 2257–66.

Pitcher, David et al. (1992). “Classification of coryneform bacteria associated with human urinary tract infection (group D2) as Corynebacterium urealyticum sp. nov.” In: *International Journal of Systematic Bacteriology* 42.1, pp. 178–181.

Pompanon, Francois et al. (2012). “Who is eating what: Diet assessment using next generation sequencing”. In: *Molecular Ecology* 21, pp. 1931–1950.

Prankevicius, Andrew B. and Duncan M. Cameron (1991). “Bacterial dinitrogen fixation in the leaf of the northern pitcher plant (Sarracenia purpurea)”. In: *Canadian Journal of Botany* 69, pp. 2296–2298.

Praz, Christophe J et al. (2008). “Host recognition in a pollen-specialist bee: evidence for a genetic basis”. In: *Apidologie* 39, pp. 547–557.

Prokopowich, Cheryl D et al. (2003). “The correlation between rDNA copy number and genome size in eukaryotes”. In: *Genome* 46.1, pp. 48–50.

R Core Team (2014). *R: A language and environment for statistical computing*.

Raj, Gopan et al. (2011). “Distribution of naphthoquinones, plumbagin, droserone, and 5-O-methyl droserone in chitin-induced and uninduced Nepenthes khasiana: molecular events in prey capture”. In: *Journal of Experimental Botany* 62.15, pp. 5429–5436.

Rappé, M S and S J Giovannoni (2003). “The uncultured microbial majority”. In: *Annual Review of Microbiology* 57, pp. 369–394.

Ratnasingham, Sujeevan and Paul D N Hebert (2007). “BOLD : The Barcode of Life Data System (www.barcodinglife.org)”. In: *Molecular Ecology Notes* 7.3, pp. 355–364.

Redford, Amanda J and Noah Fierer (2009). “Bacterial succession on the leaf surface: a novel system for studying successional dynamics”. In: *Plant Microbe Interactions* 58, pp. 189–198.

Richardson, Alan E et al. (2009). “Acquisition of phosphorous and nitrogen in the rhizosphere and plant growth promotion by microorganisms”. In: *Plant and Soil* 321, pp. 305–339.

Richardson, Rodney T. et al. (2015). “Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem”. In: *Applications in Plant Sciences* 3.1, p. 1400066.

Riedel, Michael et al. (2003). “Slippery surfaces of carnivorous plants: composition of epicuticular wax crystals in Nepenthes alata Blanco pitchers”. In: *Planta* 218, pp. 87–97.

Riedel, Michael et al. (2007). “Chemical composition of epicuticular wax crystals on the slippery zone in pitchers of five Nepenthes species and hybrids”. In: *Planta* 225, pp. 1517–1534.

Robinson, Mark D. et al. (2010). “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data”. In: *Bioinformatics* 26.1, pp. 139–140.

Roos, Stefan et al. (2000). “Lactobacillus mucosae sp. nov., a new species with in vitro mucus-binding activity isolated from pig intestine”. In: *International Journal of Systematic and Evolutionary Microbiology* 50, pp. 251–258.

Rottloff, Sandy et al. (2011). “Functional characterization of a class III acid endochitinase from the traps of the carnivorous pitcher plant genus, Nepenthes”. In: *Journal of Experimental Botany* 62.13, pp. 4639–4647.

Rowe, Owen F et al. (2007). “Microbial communities and geochemical dynamics in an extremely acidic, metal-rich stream at an abandoned sulfide mine (Huelva, Spain) underpinned by two functional primary production systems”. In: *Environmental Microbiology* 9.7, pp. 1761–1771.

Sayers, Eric W et al. (2011). “Database resources of the national centre for biotechnology information”. In: *Nucleic Acids Research* 39.Database Issue, pp. D38–D51.

Scharmann, M. and T.U. Grafe (2013). “Reinstatement of Nepenthes hemsleyana (Nepenthaceae), an endemic pitcher plant from Borneo, with a discussion of associated Nepenthes taxa”. In: *Blumea* 58, pp. 8–12.

Schoch, Conrad L. et al. (2012). “Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi”. In: *Proceedings of the National Academy of Sciences of the United States of America* 109.16, pp. 1–6. arXiv: arXiv:1408.1149.

Schöner, Caroline R et al. (2013). “Supply determines demand: influence of partner quality and quantity on the interactions between bats and pitcher plants”. In: *Oecologia* 173, pp. 191–202.

Schöner, Michael G. et al. (2015). “Bats are acoustically attracted to mutualistic carnivorous plants”. In: *Current Biology* 25, pp. 1911–1916.

Schultz, Jörg et al. (2006). “The internal transcribed spacer 2 database–a web server for (not only) low level phylogenetic analyses.” In: *Nucleic acids research* 34.Web Server issue, W704–7.

Schulze, Waltraud et al. (1999). “Transporters for ammonium, amino acids and peptides are expressed in pitchers of the carnivorous plant Nepenthes”. In: *The Plant Journal* 17.6, pp. 637–646.

Shannon, C. E. and W. Weaver (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press, p. 117.

Shebl, R I et al. (2012). “Antimicrobial profile of selected snake venoms and their associated enzymatic activities”. In: *British Microbiology Research Journal* 2, pp. 251–263.

Shokralla, Shadi et al. (2012). “Next-generation sequencing technologies for environmental DNA research”. In: *Molecular Ecology* 21, pp. 1794–1805.

Sickel, Wiebke et al. (2015). “Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach”. In: *BMC Ecology* 15.20.

Siragusa, Alex J. et al. (2007). “Culturable bacteria present in the fluid of the hooded-pitcher plant Sarracenia minor based on 16S rDNA gene sequence data”. In: *Microbial Ecology* 54.2, pp. 324–331.

Smets, Wenke et al. (2015). “A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing”. In: *PeerJ PrePrints* 3, e1622.

Soininen, Eeva M et al. (2009). “Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures”. In: *Frontiers in Zoology* 6, p. 16.

Sota, T et al. (1998). “Local and regional-scale food web structure in Nepenthes alata pitchers”. In: *Biotropica* 30.1, pp. 82–91.

Stackebrandt, E. and B. M. Goebel (1994). “Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology”. In: *International Journal of Systematic Bacteriology* 44.4, pp. 846–849. arXiv: `/dx.doi.org/10.1099/00207713-44-4-846 [http:]`.

Stackebrandt, Erko et al. (2002). “Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology”. In: *International Journal of Systematic and Evolutionary Microbiology* 52, pp. 1043–1047.

Sugahara, Hirosuke et al. (2015). “Probiotic Bifidobacterium longum alters gut luminal metabolism through modification of the gut microbial community”. In: *Nature* 5.13548, pp. 1–11.

Takeuchi, Yayoi et al. (2011). “In situ enzyme activity in the dissolved and particulate fraction of the fluid from four pitcher plant species of the genus Nepenthes”. In: *PloS one* 6.9, e25144.

Takeuchi, Yayoi et al. (2015). “Bacterial diversity and composition in the fluid of pitcher plants of the genus Nepenthes”. In: *Systematic and Applied Microbiology* 38.5, pp. 330–339.

Turner, Thomas R et al. (2013). “Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants”. In: *The ISME Journal* 7, pp. 2248–2258.

Valentini, Alice et al. (2009). “New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach”. In: *Molecular Ecology Resources* 9, pp. 51–60.

Valentini, Alice et al. (2010). “DNA barcoding for honey biodiversity”. In: *Diversity* 2, pp. 610–617.

Vessey, J Kevin (2003). “Plant growth promoting rhizobacteria as biofertilizers”. In: *Plant and Soil* 255, pp. 571–586.

Videira, Sandy Sampaio et al. (2009). “Occurrence and diversity of nitrogen-fixing Sphingomonas bacteria associated with rice plants grown in Brazil”. In: *FEMS Microbiology Letters* 293, pp. 11–19.

Walter, Jens et al. (2011). “Host-microbial symbiosis in the vertebrate gastrointestinal tract and the Lactobacillus reuteri paradigm”. In: *Proceedings of the National Academy of Sciences* 108.Suppl 1, pp. 4645–4652.

Wang, Qiong et al. (2007). “Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy”. In: *Applied and Environmental Microbiology* 73.16, pp. 5261–5267.

Warwick, C et al. (2001). “Reptile-related salmonellosis”. In: *Journal of the Royal Society of Medicine* 94, pp. 124–126.

White, T J et al. (1990). “Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics”. In: *PCR Protocols: A Guide to Methods and Applications*. Ed. by M A Innis et al. New York: Academic Press, pp. 315–322.

Wiens, John J. and Maria R. Servedio (2000). “Species delimitation in systematics: inferring diagnostic differences between species.” In: *Proceedings of the Royal Society of London - Series B: Biological Sciences* 267, pp. 631–636.

Williams, Neal M. and Claire Kremen (2007). “Resource distributions among habitats determine solitary bee offspring production in a mosaic landscape”. In: *Ecological Applications* 17, pp. 910–921.

Wilson, Erin E et al. (2010). “Pollen foraging behaviour of solitary Hawaiian bees revealed through molecular pollen analysis”. In: *Molecular Ecology* 19, pp. 4823–4829.

Woese, C R and G E Fox (1977). “Phylogenetic structure of the prokaryotic domain: the primary kingdoms”. In: *Proceedings of the National Academy of Sciences USA* 74, pp. 5088–5090.

Woolfe, Mark and Sandy Primrose (2004). “Food forensics: using DNA technology to combat misdescription and fraud”. In: *Trends in Biotechnology* 22.5, pp. 222–226.

Yabuuchi, Eiko et al. (1983). “Sphingobacterium gen. nov., Sphingobacterium spiritivorum comb. nov., Sphingobacterium multivorum comb. nov., Sphingobacterium mizutae sp. nov., and Flavobacterium indologenes sp. nov.: glucose-nonfermenting gram-negative rods in CDC groups IIK-2 and

IIb". In: *International Journal of Systematic and Evolutionary Bacteriology* 33.3, pp. 580–598.

Yip, Pui Ying et al. (2007). "DNA methods for identification of Chinese medicinal materials". In: *Chinese medicine* 2, p. 9.

Yogiara, Antonius Suwanto and Maggy T. Suhartono (2006). "A complex bacterial community living in pitcher plant fluid". In: *Jurnal Mikrobiologi Indonesia* 11.1, pp. 9–14.

Yu, Douglas W. et al. (2012). "Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring". In: *Methods in Ecology and Evolution* 3, pp. 613–623.

Zinger, Lucie et al. (2012). "Two decades of describing the unseen majority of aquatic microbial diversity". In: *Molecular Ecology* 21, pp. 1878–1896.

Abbreviations

ANOSIM analysis of similarity
COI cytochrome oxidase I
DCA detrended correspondence analysis
eDNA environmental DNA
F fluid
HTS high throughput sequencing
I inside
ITS internal transcribed spacer
ITS 2 internal transcribed spacer 2
L leaf
logFC logarithmic fold change
O outside
OTU operational taxonomic unit
P peristome
PBS phosphate-buffered saline
qPCR quantitative PCR
rDNA ribosomal DNA
rRNA ribosomal RNA
Tm melting temperature

List of Figures

List of Tables

"Dissertation Based on Several Published Manuscripts"

Statement of individual author contributions and of legal second publication rights

(If required please use more than one sheet)

Publication (complete reference): **Sickel W**, Grafe TU, Meuche I, Steffan-Dewenter I, Keller A (2016) Bacterial diversity and community structure in two Bornean *Nepenthes* species with differences in nitrogen acquisition strategies, *Microbial Ecology* **71**:4, 938-953, DOI: 10.1007/s00248-015-0723-3

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	AK WS	TUG AK	ISD		
Data Collection	IM	TUG	WS		
Data Analysis and Interpretation	WS	AK			
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	 WS WS WS WS	 AK AK AK AK	 TUG/IM/ISD TUG/IM/ISD TUG/IM/ISD		

Explanations (if applicable):

Publication (complete reference): Keller A, Grimmer G, **Sickel W**, Ankenbrand MJ (2016) DNA-Metabarcoding - ein neuer Blick auf organismische Diversität, *BioSpektrum* **22**:2, 147-150

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	N/A N/A				
Data Collection	N/A				
Data Analysis and Interpretation	N/A				
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	 AK N/A AK AK	 WS/MJA/GG WS/MJA/GG			

Explanations (if applicable): Study Design, Methods Development, Data Collection, Data Analysis & Interpretation not applicable, since this is a review article.

Publication (complete reference): **Sickel W**, Ankenbrand MJ, Grimmer G, Holzschuh A, Härtel S, Lanzen J, Steffan-Dewenter I, Keller A (2015) Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology* **15**:20, DOI: 10.1186/s12898-015-0051-y

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	AK WS/GG	AH/ISD AK/MJA	JL/SH		
Data Collection	WS/GG	MJA	JL		
Data Analysis and Interpretation	WS/MJA	AK			
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	 WS WS WS WS	 AK/MJA MJA/AK MJA/AK MJA/AK	 GG/AH/SH/JL/ISD GG/AH/SH/JL/ISD GG/AH/SH/JL/ISD GG/AH/SH/JL/ISD		

Explanations (if applicable):

Publication (complete reference): Zancolli G, Mahsberg D, **Sickel W**, Keller A (2015) Reptiles as reservoirs of bacterial infections: real threat or methodological bias? *Microbial Ecology* **70**:3, 579-584, DOI: 10.1007/s00248-015-0618-3

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	GZ/AK WS/AK				
Data Collection	GZ/DM	WS			
Data Analysis and Interpretation	GZ	AK			
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	 GZ GZ GZ GZ	 WS/AK/DM WS/AK/DM WS/AK/DM WS/AK/DM			

Explanations (if applicable):

Publication (complete reference): **Sickel W**, Ankenbrand MJ, Grimmer G, Holzschuh A, Härtel S, Lanzen J, Steffan-Dewenter I, Keller A (2015) Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology* **15**:20, DOI: 10.1186/s12898-015-0051-y

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	WS	AK			
2	WS	MJA			
3	WS	AK			
Supplemental File 1	WS	MJA	JL		
Supplemental File 2	MJA				
Supplemental File 3	MJA				

Explanations (if applicable):

Publication (complete reference): Zancolli G, Mahsberg D, **Sickel W**, Keller A (2015) Reptiles as reservoirs of bacterial infections: real threat or methodological bias? *Microbial Ecology* **70**:3, 579-584, DOI: 10.1007/s00248-015-0618-3

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	GZ	AK			
Table 1	GZ	AK			

Explanations (if applicable):

Publication (complete reference): Sickel W, Ankenbrand MJ, Grimmer G, Förster F, Steffan-Dewenter I, Keller Alexander (in press) Standard method for identification of bee pollen mixtures through meta-barcoding. *COLOSS BEEBOOK* Volume III

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	WS	AK			
2	WS	AK			
Table 1	WS				

Explanations (if applicable):

I also confirm my primary supervisor's acceptance.

Wiebke Sickel 04.07.16 Würzburg W. Sickel

Doctoral Researcher's Name Date Place Signature

"Dissertation Based on Several Published Manuscripts"

Statement of individual author contributions and of legal second publication rights

(If required please use more than one sheet)

Publication (complete reference): **Sickel W**, Grafe TU, Meuche I, Steffan-Dewenter I, Keller A (2016) Bacterial diversity and community structure in two Bornean *Nepenthes* species with differences in nitrogen acquisition strategies, *Microbial Ecology* **71**:4, 938-953, DOI: 10.1007/s00248-015-0723-3

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	AK WS	TUG AK	ISD		
Data Collection	IM	TUG	WS		
Data Analysis and Interpretation	WS	AK			
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	 WS WS WS WS	 AK AK AK AK	 TUG/IM/ISD TUG/IM/ISD TUG/IM/ISD		

Explanations (if applicable):

Publication (complete reference): Keller A, Grimmer G, **Sickel W**, Ankenbrand MJ (2016) DNA-Metabarcoding - ein neuer Blick auf organismische Diversität, *BioSpektrum* **22**:2, 147-150

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	N/A N/A				
Data Collection	N/A				
Data Analysis and Interpretation	N/A				
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	 AK N/A AK AK	 WS/MJA/GG WS/MJA/GG			

Explanations (if applicable): Study Design, Methods Development, Data Collection, Data Analysis & Interpretation not applicable, since this is a review article.

"Dissertation Based on Several Published Manuscripts"

Statement of individual author contributions to figures/tables/chapters included in the manuscripts

(If required please use more than one sheet)

Publication (complete reference): **Sickel W**, Grafe TU, Meuche I, Steffan-Dewenter I, Keller A (2016) Bacterial diversity and community structure in two Bornean *Nepenthes* species with differences in nitrogen acquisition strategies, *Microbial Ecology* **71**:4, 938-953, DOI: 10.1007/s00248-015-0723-3

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	WS	AK			
2	WS	AK			
3	WS	AK			
4	WS	AK			
5	WS	AK			
Table 1	WS	AK			
Table 2	WS	AK			
Supplemental File 1	WS	AK			
Supplemental File 2	WS	AK			
Supplemental File 3	WS	AK			

Explanations (if applicable):

Publication (complete reference): Keller A, Grimmer G, **Sickel W**, Ankenbrand MJ (2016) DNA-Metabarcoding - ein neuer Blick auf organismische Diversität, *BioSpektrum* **22**:2, 147-150

Figure	**Author Initials,** Responsibility decreasing from left to right				
1	AK				
2	WS				
3	AK				

Explanations (if applicable):

Publication (complete reference): **Sickel W**, Ankenbrand MJ, Grimmer G, Förster F, Steffan-Dewenter I, Keller A (in press) Standard method for identification of bee pollen mixtures through meta-barcoding. *COLOSS BEEBOOK* Volume III					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	N/A WS/MJA	GG/AK	FF	ISD	
Data Collection	N/A				
Data Analysis and Interpretation	N/A				
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	 WS/MJA/AK WS/MJA/AK N/A WS/MJA/AK	 GG/FF GG/FF GG/FF	 ISD ISD ISD		

Explanations (if applicable): shared first authorship between WS and MJA; Study Design, Data Collection, Data Analysis and Interpretation, Writing of Discussion not applicable, since this is a protocol for a standard method.

The doctoral researcher confirms that she/he has obtained permission from both the publishers and the co-authors for legal second publication.

The doctoral researcher and the primary supervisor confirm the correctness of the above mentioned assessment.

Wiebke Sickel	04.07.16	Würzburg	W. Sickel
Doctoral Researcher's Name	Date	Place	Signature

Dr Alexander Keller	04.07.16	Würzburg	
Primary Supervisor's Name	Date	Place	Signature

Curriculum Vitae

Wiebke Sickel, PhD Student
born 10 March 1989 in Oranienburg, Germany,
unmarried, no children

Main areas of research

Biodiversity, Carnivorous plants, High-throughput sequencing, Microbial Genomics, Microbial Ecology, Plant-microbe interactions

Scientific career and education

planned: 10/2016 - 01/2017	Research stay at Genecology Research Center, Faculty of Science, Health, Education, Engineering, University of the Sunshine Coast, Australia; followed by private stay
since 10/2014	PhD Student, Molecular Biodiversity Group, Department for Animal Ecology and Tropical Biology, University of Würzburg; Thesis title: *High-throughput biodiversity assessment - Powers and limitations of meta-barcoding*, submission: 09/2016
10/2013 - 09/2014	MSc Biology, University of Würzburg, Grade: 1.4; MSc thesis title: *Diet as a potential factor determining microbial communities in Bornean pitcher plants*
05/2013 - 09/2013	Research stay at Leibniz Institute for Zoo and Wildlife Research; MSc thesis title: *Phylogeography of the Marbled Cat (*Pardofelis marmorata*, Martin 1837) in South East Asia*
09/2012 - 09/2013	MSc Forensic and Conservation Genetics, University of Central Lancashire, UK, Grade: Distinction
09/2011 - 05/2012	BSc (Hons) Forensic Science, Robert Gordon University Aberdeen, UK (Double Degree Programme); Grade: First Class; BSc thesis title: *An investigation into the potential use of elemental profiles of soil samples as a method of between-city discrimination for forensic purposes - A SoilFUN Project* (in conjunction with The James Hutton Institute, Aberdeen, UK)
09/2009 - 07/2012	BSc Forensic Science, University of Applied Science Bonn-Rhein-Sieg, Germany; Grade: 1.3
09/2006 - 06/2009	Secondary Education (Abitur), Gymnasium 'Friedlieb Ferdinand Runge', Oranienburg, Germany; Grade: 1.2

Skills

Laboratory techniques	Bacterial cultivation, DNA isolation, DNA quantification, PCR, Sanger Sequencing, High throughput sequencing
IT	R, LaTeX, MS Office, Perl, Unix/Linux, QIIME, USEARCH
Languages	German (mother tongue), English (fluent, Cambridge Certificate in Advanced English)

Research grants

2015	€1,000 for covering sequencing costs, granted by Universitätsbund Würzburg e.V. for the project *Microbiota of the poison dart frog* Oophaga pumilio *and its variability regarding toxicity, biogeography and behaviour*

Scientific activities

since 2015	Member of German Carnivorous Plant Society (Gesellschaft für Fleischfressende Pflanzen im deutschsprachigen Raum e.V.) Member of International Carnivorous Plant Society Member of German Herpetological Society (Deutsche Gesellschaft für Herpetologie und Terrarienkunde)

International Conferences & Workshops

08/2016	11th International Carnivorous Plant Conference, Kew, London, UK; Oral Presentation: *Inventorying carnivorous plant associated bacteria: co-evolution, convergence and function*; Travel grant of German Academic Exchange Service (DAAD), Travel grant of Graduate School of Life Sciences, University of Würzburg, Germany
05/2016	Plant and Pollen Metabarcoding Summerschool 2016 with Pollen BoL Mini-Symposium, Würzburg, Germany; Co-Organiser and Lecturer
10/2015	Eureka! - 10th Symposium of the Graduate School of Life Sciences, Würzburg, Germany; Poster Presentation: *You shall not pass! The Venus flytrap microbiota is resistant against invaders introduced by prey items*; Co-Organiser
06/2015	13th Symposium on Bacterial Genetics and Ecology, Milan, Italy; Poster Presentation: *Bat mutualism or not: microbiota comparison of two closely related Bornean* Nepenthes *species*; Travel grant of Graduate School of Life Sciences, University of Würzburg, Germany
10/2014	Eureka! - 9th International Symposium of the Graduate School of Life Sciences, Würzburg, Germany; Poster Presentation: *Diet determines bacterial diversity and community structure in Bornean pitcher plants*
07/2014	ISFAS 2014 - International Symposium on Research and Conservation Biology in South America, Nürnberg, Germany
09/2013	9th International Conference on Behaviour, Physiology and Genetics of Wildlife, Berlin, Germany

Other workshops

06/2016	Open Access and Copyright in Science, organised by the Graduate School of Life Sciences, University of Würzburg
	Slidewriting and Storylining, organised by the Graduate School of Life Sciences, University of Würzburg
01/2016	Introduction to System Administration, organised by the Bioinformatics Department, University of Würzburg
03/2015	Intercultural Communication, organised by the Graduate School of Life Sciences, University of Würzburg
	Good Scientific Practice, organised by the Graduate School of Life Sciences, University of Würzburg
09/2014	Scientific Writing and Publishing, organised by the Graduate School of Life Sciences, University of Würzburg
10/2013	Leading Skills for Professionals and Students (Führungskompetenzen für Beruf und Studium), organised by Deutsche Bildung AG

Publications

2016

Sickel W, U Grafe, I Meuche, I Steffan-Dewenter, A Keller. Bacterial diversity and community structure in two Bornean *Nepenthes* species with differences in nitrogen acquisition strategies, *Microbial Ecology* **71**:4, 938-953, DOI: 10.1007/s00248-015-0723-3

Keller A, G Grimmer, **W Sickel**, MJ Ankenbrand (2016) DNA-Metabarcoding - ein neuer Blick auf organismische Diversität, *BioSpektrum* **22**:2, 147-150

2015

Sickel W, M Ankenbrand, G Grimmer, A Holzschuh, S Härtel, J Lanzen, I Steffan-Dewenter, A Keller (2015) Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology* **15**:20, DOI: 10.1186/s12898-015-0051-y

Zancolli G, D Mahsberg, **W Sickel**, A Keller (2015) Reptiles as reservoirs of bacterial infections: real threat or methodological bias? *Microbial Ecology* **70**:3, 579-584, DOI: 10.1007/s00248-015-0618-3

in press

Sickel W*, MJ Ankenbrand*, G Grimmer, F Förster, I Steffan-Dewenter, A Keller (in press) Standard method for identification of bee pollen mixtures through meta-barcoding. COLOSS BeeBook, Vol. III

*equally contributing