

Pollen DNA barcoding: current applications and future prospects¹

Karen L. Bell, Natasha de Vere, Alexander Keller, Rodney T. Richardson, Annemarie Gous, Kevin S. Burgess, and Berry J. Brosi

Abstract: Identification of the species origin of pollen has many applications, including assessment of plant–pollinator networks, reconstruction of ancient plant communities, product authentication, allergen monitoring, and forensics. Such applications, however, have previously been limited by microscopy-based identification of pollen, which is slow, has low taxonomic resolution, and has few expert practitioners. One alternative is pollen DNA barcoding, which could overcome these issues. Recent studies demonstrate that both chloroplast and nuclear barcoding markers can be amplified from pollen. These recent validations of pollen metabarcoding indicate that now is the time for researchers in various fields to consider applying these methods to their research programs. In this paper, we review the nascent field of pollen DNA barcoding and discuss potential new applications of this technology, highlighting existing limitations and future research developments that will improve its utility in a wide range of applications.

Key words: DNA metabarcoding, metagenomics, pollen, palynology, high-throughput sequencing, next-generation sequencing.

Résumé : L'identification de l'espèce à l'origine d'un pollen se prête à de nombreuses applications dont la description des réseaux plante–pollinisateur, la reconstruction de communautés de plantes anciennes, l'authentification de produits, la surveillance des allergènes et les enquêtes médico-légales. Cependant, ces applications ont précédemment été limitées à l'identification du pollen par examen microscopique, un processus lent, à faible résolution taxonomique et qui compte peu de praticiens experts. Une alternative est l'identification du pollen par le recours aux codes à barres de l'ADN, une avenue qui permettrait de surmonter plusieurs de ces limitations. De récentes études ont montré qu'il était possible d'amplifier les marqueurs de codage tant chloroplastiques que nucléaires à partir du pollen. Ces récentes validations du métacodage à barres chez le pollen indiquent qu'il est maintenant opportun pour les chercheurs dans divers domaines de considérer l'emploi de ces méthodes dans leurs programmes de recherche. Dans cet article, les auteurs passent en revue le domaine naissant du codage à barres du pollen et discutent des nouvelles applications potentielles de cette technologie en mettant en lumière les limitations existantes ainsi que de futurs développements qui pourraient accroître son utilité dans un grand nombre d'applications. [Traduit par la Rédaction]

Mots-clés : métacodage à barres, métagénomique, pollen, palynologie, séquençage à haut débit, séquençage de nouvelle génération.

Background and potential of pollen DNA barcoding

The ability to identify plant species based on their pollen has a multitude of applications across fields as diverse as pollination biology, forensics, and allergen

monitoring, but in the past, technical limitations have prevented its use in most fields. Traditional techniques of pollen identification rely on a high level of expertise and time-consuming examination of morphological characters on the pollen exine, typically with low taxonomic resolution

Received 30 November 2015. Accepted 8 February 2016.

Corresponding Editor: M. Van der Bank.

K.L. Bell and B.J. Brosi. Emory University, School of Environmental Sciences, Atlanta, GA, USA.

N. de Vere. National Botanic Garden of Wales, Llanarthne, United Kingdom.

A. Keller. Department of Animal Ecology and Tropical Biology, University of Würzburg, Würzburg, Germany.

R.T. Richardson. Department of Entomology, The Ohio State University, Columbus, OH, USA.

A. Gous. Biotechnology Platform, Agricultural Research Council, Pretoria, South Africa; School of Life Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa.

K.S. Burgess. Columbus State University, Columbus, GA, USA.

Corresponding author: Karen L. Bell (email: karen.bell@emory.edu).

¹This paper is part of a special issue entitled Barcodes to Biomes.

Copyright remains with the author(s) or their institution(s). Permission for reuse (free in most cases) can be obtained from RightsLink.

(Rahl 2008). Although some of these issues could potentially be overcome with automated analysis of digital micrographs of pollen (Holt and Bennett 2014), the lack of morphological characters for species-level taxonomic resolution in many plant groups will remain a limitation (Rahl 2008; Salmaki et al. 2008; Khansari et al. 2012). These issues could be resolved using new DNA barcoding technologies, making palynological information available to a broader range of studies.

Over the past decade, the power of DNA barcoding has opened up new fields in taxonomic, ecological, and evolutionary research by facilitating species identification. For animals, DNA barcoding is defined as the sequencing of a standardized barcode marker (the *COI* gene region of the mitochondrial genome) that shows specificity within a species and variability between species (Hebert et al. 2003a, 2003b; Borisenko et al. 2009; Janzen et al. 2009). In plants, three regions of the chloroplast genome (*rbcl*, *matK*, and *trnH-psbA*) as well as the nuclear ribosomal ITS region have been widely used as DNA barcodes, either separately or in combination (Fazekas et al. 2008; CBOL Plant Working Group 2009; Fazekas and Kesanakurti 2009; Chen et al. 2010; Hollingsworth et al. 2011). For the most part, these barcodes yield relatively high species discrimination, particularly at the regional scale, at relatively low cost and have been used to build barcode libraries for local floras to address taxonomic and ecological questions (Burgess et al. 2011; Kesanakurti et al. 2011; de Vere et al. 2012). For example, barcoding efforts targeting a broad sampling of plant taxa have shown that species resolution is ~72% using the standard *rbcl*+*matK* barcode (CBOL Plant Working Group 2009). More recent studies, however, using the same barcode regions for floras of moderate phylogenetic dispersion have shown that up to 92% of the species can be distinguished (Kress et al. 2009; Burgess et al. 2011) and that *rbcl*+*matK* can be used effectively in a number of diverse barcoding applications, including environmental sampling (reviewed in Hollingsworth et al. 2011). ITS2 has shown similar discriminative capabilities, with 92.7% successful identifications in 6600 samples (Chen et al. 2010).

These methods have only recently been applied to the identification of plants based on their pollen. This may seem surprising given the myriad applications of a rapid, standardized pollen species identification method, but this late uptake is likely due to a number of real and perceived technical difficulties. First, the standard DNA barcoding loci are on the plastid genome (CBOL Plant Working Group 2009; Hollingsworth et al. 2009). Plastids are plant organelles with a distinct genome, which can differentiate into several more specialized organelles, in particular chloroplasts (Fujiwara et al. 2010). In most flowering plants this organelle is inherited maternally and is reduced in the male germ line of the pollen grain in many taxa (e.g., Sakamoto et al. 2008) via specific organelle degradation pathways in pollen that have been

described in detail (Matsushima et al. 2011). Following similar observations, suggestions that plastid DNA (ptDNA) was absent from the pollen (e.g., Willerslev et al. 2003) may have discouraged early development of pollen DNA barcoding methods. Several studies have now shown proof-of-concept for amplification of ptDNA from pollen (Galimberti et al. 2014; Hawkins et al. 2015; Kraaijeveld et al. 2015; Richardson et al. 2015a), so this is no longer considered an issue.

Another caveat with pollen is that samples normally occur as mixtures of multiple species, meaning traditional Sanger-based sequencing is of little utility. One work-around is to isolate and sequence individual pollen grains from these mixtures (Matsuki et al. 2007; Aziz and Sauve 2008). This technique can be useful where fine-scale knowledge is required, but it is not practical for large-scale application and underlies the same sorting restrictions as morphology. Another strategy is to use amplicon cloning techniques (e.g., Galimberti et al. 2014), but these are also labour-intensive and not comprehensive with respect to sampling depth. Recent improvements and price reductions in high-throughput sequencing (HTS) are promising in terms of addressing the issue of mixed-species identification (i.e., DNA metabarcoding), and recent studies have demonstrated the potential of these methods (Hawkins et al. 2015; Keller et al. 2015; Kraaijeveld et al. 2015; Richardson et al. 2015b; Sickel et al. 2015). These recent breakthroughs could enable rapid, large-scale species identification of pollen mixtures, with the potential to transform research in a range of fields.

To successfully conduct DNA metabarcoding of pollen, four components are needed: first, an extraction protocol that yields high-quality DNA template for amplification; second, a set of genetic markers that can be successfully amplified across all seed plants (those that produce pollen); third, a database containing reference sequences of the aforementioned genetic markers for the majority of seed plant species, enabling comparison to the sequenced product; and fourth, a HTS method and bioinformatic pipeline that allows the simultaneous identification of several species from a single mixed-species pollen sample. In this paper, we review these four components, evaluating the current state of the art in pollen DNA barcoding, outlining progress on resolving technical issues, and making recommendations for standardizing methodology. We then discuss a range of potential future applications of pollen DNA barcoding.

Component 1: The pollen DNA template

Methodologically, current pollen DNA barcoding approaches can be delineated based on the use of traditional Sanger sequencing or the implementation of more technologically advanced but less established HTS. For both methodologies, the quality of the template is important, and is determined by the pollen collection and DNA isolation methods. Pollen collection methods will

vary depending on the type of study, for example pollination studies would involve collecting pollen directly from a pollinator, while air quality monitoring would involve collecting airborne pollen. According to the type of study, and the pollen collecting method, the quantity of pollen may be low. Methods exist for amplifying DNA from a single pollen grain (Petersen et al. 1996; Matsunaga et al. 1999), allowing DNA barcoding from such samples either through traditional Sanger sequencing or through HTS sequencing of amplicons.

Following pollen sampling, an effective DNA isolation method is required to maximize yield. A key element of DNA extraction is disruption of the pollen exine, which has high structural integrity. Methods of exine disruption include pulverization with bead-beating or TissueLyser devices, either with or without proteinases, and (or) mortar and pestle-based pulverization facilitated by liquid nitrogen freezing. For studies investigating pollens derived from honey samples, sample pulverization is preceded by sample dilution and centrifugation to separate the pollen from the highly viscous honey (Bruni et al. 2015; Hawkins et al. 2015). It may also be possible to use enzymatic techniques to digest the pollen exine. The honey bee gut microbiome includes species that produce pectin-degrading enzymes, enabling the digestion of pollen (Engel et al. 2012). Further investigation is required to assess the potential for such enzymes in DNA extraction from pollen. For DNA extraction, studies to date have largely relied on proprietary kits including the QIAamp DNA Mini Kit (Kraaijeveld et al. 2015), Qiagen DNeasy Plant Mini Kit (Galimberti et al. 2014; Bruni et al. 2015; Hawkins et al. 2015; Richardson et al. 2015a, 2015b), and Machery-Nagel NucleoSpin Food Kit (Keller et al. 2015; Sickel et al. 2015). It is, however, currently unclear to what extent different extraction strategies are comparable. Quantitative studies on DNA metabarcoding of microbiomes have found biases towards specific taxa, depending on the DNA isolation method (Brooks et al. 2015), and similar biases may occur with pollen. The development of a standardized pollen DNA isolation method would ensure comparability between studies.

There are also several unresolved technical issues specific to the DNA barcoding of mixed-species pollen samples. As previously mentioned, the fact that the consensus DNA barcoding markers are on the plastid genome (CBOL Plant Working Group 2009) may cause technical problems for quantitative pollen DNA barcoding. Copy numbers of ptDNA in pollen likely vary amongst species, particularly between those where the ptDNA is inherited maternally and those where it is inherited biparentally or paternally. This variation is poorly understood (Sangwan and Sangwan-Norreel 1987; Corriveau and Coleman 1988; Nagata et al. 1999; Zhang et al. 2003), which severely limits any quantitative inference (i.e., pollen counts of various species) from DNA metabarcoding of mixed-species pollen samples. ITS2

suffers the same issue in terms of quantification, and copy number is related to both ploidy and the number of ribosomal DNA copies in the nuclear genome, which can be highly variable between species, between individuals of the same species, and variation has even been recorded within individuals due to somatic mutations (e.g., Rogers and Bendich 1987).

In addition to difficulties with quantification of pollen mentioned above, the variable ptDNA copy number and variable DNA extraction efficiency may also exacerbate problems of contamination. Contamination can come from various sources, including contaminated samples, laboratory contamination, cross-contamination of samples, and contaminated reagents. The small size of a typical pollen sample means that any trace of contamination could generate a misleading result, particularly in HTS approaches. Standard laboratory hygiene practices should prevent contamination in typical circumstances. For more sensitive applications, methods commonly used in ancient DNA and forensics could also be applied to pollen DNA barcoding to eliminate contamination from the laboratory, reagents, and cross-contamination between samples (e.g., Champlot et al. 2010). Pollen sample decontamination methods could take advantage of the chemical resistance of the pollen exine (Southworth 1974; Kearns and Inouye 1993). A chemical treatment prior to pollen DNA extraction could eliminate contamination from non-pollen plant material and non-cellular DNA, such as PCR products.

Component 2: Genetic markers

The choice of marker and the primer set used for amplification are of great importance for any DNA barcoding endeavor. They dictate the scope of taxonomic recovery, discriminatory power, and sequencing considerations. In general, the universality, across seed plants, of the amplification primers, as well as high inter-specific but low intra-specific variability, are the major requirements for a successful pollen DNA barcode marker. To date, five markers have been employed for pollen DNA metabarcoding, including *rbcl*, *matK*, ITS2, *trnL*, and *trnH-psbA*. See Table 1 for a summary of these markers and their relevant attributes. It is important to use a marker with a high degree of universality across a broad range of taxonomic groups to avoid PCR biases, which can lead to some taxa being preferentially amplified (Shokralla et al. 2012; Taberlet et al. 2012). Trade-offs may occur between primer universality and inter-specific variability, with more universal markers providing lower species-level discrimination (CBOL Plant Working Group 2009). Poor taxonomic resolution of identifications based on single-locus barcodes using the most universal markers (Hollingsworth et al. 2009; Kress et al. 2009; Burgess et al. 2011; Hollingsworth et al. 2011) has resulted in the adoption of multi-locus DNA barcoding approaches within the plant barcoding community (CBOL Plant Working Group 2009; Kress et al. 2009;

Table 1. Summary of information available on plant DNA barcoding markers used in pollen DNA barcoding studies.

Locus	No. of Genbank entries ^b	Reported length (bp)	Studies
<i>rbcLa</i> ^a	155 634	702–883 ^c	Bruni et al. 2015; Galimberti et al. 2014; Hawkins et al. 2015; Richardson et al. 2015a
<i>trnL</i>	198 308	321–447 ^d	Kraaijeveld et al. 2015; Valentini et al. 2010
<i>trnH-psbA</i>	86 828	103–1025 ^c	Bruni et al. 2015; Galimberti et al. 2014
<i>matK</i>	127 990	656–861 ^c	Richardson et al. 2015a
ITS2	243 155	163–311 ^c	Keller et al. 2015; Richardson et al. 2015a, 2015b; Sickel et al. 2015

^a*rbcLa* has previously been denoted *rbcL*, but a new convention was proposed by Dong et al. 2013 to distinguish from newly promoted *rbcLb*.

^bAccessed on 4 November 2015 using the following search parameters: "ITS2 OR internal transcribed spacer 2[All Fields] AND plants[filter]"; "rbcL OR rbc-L or Rubisco [All Fields] AND plants[filter]"; "trnL OR trn-L OR trnL-trnF [All Fields] AND plants[filter]"; "trnH OR trn-H OR trnH-psbA OR psbA-trnH [All Fields] AND plants[filter]"; "matK OR mat-K OR maturase K [All Fields] AND plants[filter]".

^cRange reported in Chen et al. (2010) and subject to primer set used and species amplified.

^dRange reported in Kress et al. (2005) and subject to primer set used and species amplified.

Burgess et al. 2011). Through such approaches, information from multiple markers can be linked to an individual specimen, improving species-level resolution relative to a single-marker analysis (CBOL Plant Working Group 2009; Hollingsworth et al. 2011; Kress et al. 2009; Burgess et al. 2011). Even "unlinked," as in a DNA metabarcoding context, multiple markers can generate better discriminatory power than one. For example, one could imagine a sample with exactly two plant species, which have an identical *rbcL* sequence but different *matK* sequences; these would be easily distinguished bioinformatically. Future efforts to improve the bioinformatics analysis methods could resolve situations where a mixed-species sample includes a set of species with identical *rbcL* sequence and a partially overlapping set of species with identical *matK* markers, enabling automated processes to give more accurate species counts and identities.

A number of studies have assessed the advantages and limits of the various gene regions available for plant DNA barcoding (CBOL Plant Working Group 2009; Ford et al. 2009; Hollingsworth et al. 2009, 2011; Chen et al. 2010). In an effort to standardize DNA barcoding methods, the chloroplast regions *rbcL* and *matK* were selected as barcode markers, based on their relative universality and combined taxonomic resolution (CBOL Plant Working Group 2009; Hollingsworth et al. 2011), although ITS2 has also been recommended as a standard barcode for plant

DNA barcoding based on its higher taxonomic resolution (Chen et al. 2010). The disadvantages of using non-coding regions (e.g., *trnH-psbA*) and multiple-copy nuclear markers (e.g., ribosomal genes) in traditional sequencing technologies have been reviewed (CBOL Plant Working Group 2009; Hollingsworth et al. 2011), although multiple-copy nuclear markers may be less of an issue with HTS methods. A further consideration when using ITS2 as a barcode, however, is the potential for fungal co-amplification (Cheng et al. 2016). While fungal co-amplification can lead to sequencing failure when using Sanger sequencing, the impact is low, but still significant. In HTS DNA metabarcoding studies, fungal contamination will be sequenced alongside the taxa of interest (i.e., plants) and may comprise a significant proportion of the sequencing reads (e.g., Cornman et al. 2015). This will not prevent sequencing and identification of plant species, but it may increase the number of reads required per sample, therefore limiting the number of samples that can be analyzed. A recent study (Cheng et al. 2016) assessed several existing primers for ITS2 amplification along with newly designed primers for their relative universality to plants and relative levels of fungal co-amplification.

The relative advantages and limits of different barcodes are mostly the same for pollen DNA barcoding as for general plant DNA barcoding. The theoretically lower copy number of ptDNA could decrease amplification efficiency of ptDNA barcodes, although this has not been observed in practice (Valentini et al. 2010; Galimberti et al. 2014; Bruni et al. 2015; Hawkins et al. 2015; Kraaijeveld et al. 2015; Richardson et al. 2015a). While five different markers have been employed with relative success in pollen DNA metabarcoding, only data from *rbcL*, *matK*, ITS2, and *trnL* have been compared with microscopic palynology for assessment of qualitative and quantitative consistency, and further comparisons across a broader scope of taxa are needed. Additionally, amplicon fragment size is important in HTS DNA metabarcoding, as the read lengths in many platforms are limited as of this writing (e.g., ~600-bp paired-end reads on the Illumina MiSeq platform; www.illumina.com). While ITS2 and *trnL* have been successfully sequenced via HTS with sufficient overlap for paired-end merging (Kraaijeveld et al. 2015; Richardson et al. 2015b; Sickel et al. 2015), the long amplicon length generated by standard *rbcL* and *matK* primers, as well as the length hyper-variability of *trnH-psbA*, pose a technical limitation, necessitating a redesign of primer pairs for shorter amplicon lengths. Future improvements in sequencing technology are likely to increase read lengths of HTS technologies, and may alleviate this issue in the future.

Component 3: DNA barcoding sequence reference libraries

DNA metabarcoding relies on the completeness and quality of the reference database used for taxonomic as-

signments. Ideally, reference databases would have complete representation of all taxa from a given habitat, with manually curated and verified reference sequences, especially for DNA metabarcoding purposes. This is an ambitious task for most habitats and in most cases is only applicable where a specific list of species of interest is being investigated, and other species can be neglected. A current compromise is to rely on public sequence databases that may contain high coverage of species for a given habitat. For plants, the largest sources for reference sequencing data are the redundant repositories NCBI (Benson et al. 2015), EMBL (Squizzato et al. 2015), and DDBJ (Mashima et al. 2015). There are over 32 million vascular plant nucleotide sequences deposited, although only a fraction of these represent DNA barcoding markers (see Table 1 for the number of sequences associated with each of the standard DNA barcoding markers). Secondary databases obtain their data mostly by searching this raw-data for specific markers and extracting sequences, and improve the quality by different approaches (BOLD, www.boldsystems.org; ITS2-database, Ankenbrand et al. 2015). BOLD hosts additional barcodes beyond these extracted sequences that researchers have deposited directly. For example, BOLD contains 96 744 *matK*, 97 380 ITS2, and 84 132 *rbcL* sequences for plants, and the ITS2-database contains 199 932 ITS2 sequences (access dates: 9 November 2015). Regardless of the database used, all are far from complete where many sequences are missing species-level taxonomic information and the quantity of data per species varies largely. According to Ankenbrand et al. (2015), 72% of known US plant species are represented in the ITS2 database, and many other bioregions are likely less well sampled. An additional drawback for using such databases is that misidentifications, intra-specific variation, sequencing errors, and other issues are present in such repositories and not easily inferable. The consequence is that reads obtained from DNA barcoding may be wrongly assigned in such cases. Errors in reference databases hamper automated species identifications from bioinformatics “pipelines”.

Component 4: Sequencing methodologies and bioinformatics pipelines

Sequencing methodologies for pollen DNA barcoding have changed over time. Early studies typically relied on either direct Sanger sequencing of purified PCR amplicons (Longhi et al. 2009; Wilson et al. 2010), for single-species barcoding, or else, for multi-species samples, Sanger sequencing of amplicon clones randomly selected from PCR products (Galimberti et al. 2014; Bruni et al. 2015). The latter has mostly been superseded by HTS approaches for DNA metabarcoding. More recent HTS-based studies have employed one of three major sequencing library preparation strategies: ligation-based “tagmentation” kits (Richardson et al. 2015a, 2015b), singly indexed barcoded primers (Valentini et al. 2010;

Hawkins et al. 2015; Keller et al. 2015; Kraaijeveld et al. 2015), or dual-indexed barcoded primers (Sickel et al. 2015). The dual-indexing approach described in Sickel et al. (2015), adapted from Kozich et al. (2013), shows great promise for facilitating library preparation for large studies while reducing multiplexing cost and increasing laboratory efficiency. With the large amount of sequencing coverage afforded by HTS platforms, such library preparation techniques allow the sequencing of many sample libraries on a single sequencing run, reducing cost and minimizing sequencing coverage waste. Still, we continue to have a poor idea of sequencing depth needed for sufficient detection of species present in a sample, which will vary with number of pollen grains per sample in addition to a range of other parameters. Thus, the number of samples per HTS reaction should be considered thoughtfully. As a first guideline, Sickel et al. (2015) report 2000–3000 high-quality reads to be adequate to describe bee-collected samples with up to 80 taxa included. Lastly, pollen metabarcoding has been successfully conducted using a variety of platforms including Ion Torrent (Kraaijeveld et al. 2015), Roche 454 (Valentini et al. 2010; Hawkins et al. 2015; Keller et al. 2015), and Illumina (Richardson et al. 2015a, 2015b; Sickel et al. 2015); however, with increased relative throughput, accurate homopolymer sequencing and increasing length capabilities, we see Illumina as the current platform of choice for PCR-based approaches.

Although mixed-amplicon sequencing is currently the most feasible method for mixed-species DNA barcoding, reliance on PCR amplification has some disadvantages. A potential source of error comes from the PCR amplification step (Pompanon et al. 2012). Some species, especially those present in low quantities, can be missed out when a mixed sample is amplified (Hajibabaei et al. 2012; Gibson et al. 2014). Techniques are being developed that avoid the PCR stage altogether, such as using shotgun sequencing with subsequent recovery of DNA barcode markers or even whole chloroplasts (Kane et al. 2012; Taberlet et al. 2012; Ji et al. 2013; Tang et al. 2014). Such methods could improve the accuracy of mixed-species DNA barcoding, both for species identification and quantification. Single-molecule sequencing methods could also eliminate the PCR stage (Roberts et al. 2013). These methods have a high, but random, error rate, which makes them currently less feasible for identifying species based on a DNA marker barcode.

The HTS methods described above produce large quantities of data, which need to be analysed with an efficient and accurate bioinformatics procedure. The development of a standardized bioinformatic pipeline for plant DNA metabarcoding is an important step in the development of pollen metabarcoding, since many laboratories operate on modest budgets and do not have bioinformatics specialists. There are web-based bioinformatics pipelines and sequence reference libraries for microbial

metagenomics that allow convenient, rapid assignment of sequence mixes to taxonomic units (Meyer et al. 2008). Recent, publicly available pipeline developments for the ITS2 marker (Keller et al. 2015; Sickel et al. 2015) represent a major improvement in bioinformatics capability and standardization for plant DNA metabarcoding. The development of bioinformatics pipelines is not trivial, and there are some technical issues that need to be resolved. For example, bioinformatics pipelines need to be robust to errors in reference databases. Best-hit approaches (e.g., Altschul et al. 1990) are problematic when misidentified references are matched or the sequences of interest are not represented in the database, whilst classifiers such as UTX (Edgar 2013) or RDP (Wang et al. 2007) are more resilient in both regards. Specifically, these classifiers hierarchically assign taxonomic information to make a decision about taxonomic identity and are dependent on a confidence threshold for each Linnean taxonomic level. On the other hand, the output of classifiers may contain wrong information when training bases on misclassified or identical sequences belonging to different taxa, since the output does not display alternative assignments in such cases. In every case, run-time and computational power required for this process can be reduced by the pre-clustering of taxonomic units prior to assignments. Commonly used algorithms, such as UCLUST (Edgar 2010) or MOTHUR (Schloss et al. 2009), require constant read length as the input is regarded as aligned, which can be impractical for markers such as ITS2 that show high sequence length variability. Overall, these automatic procedures help to bring information to a massive amount of data, but the resulting assignments require cross-checking between assignment methods. Fully automated identifications are possible, but should be regarded with caution, and checked using any available biological or biogeographic knowledge to assess potential for misidentifications. There is considerable scope for future improvements in this area, including expansion to all of the standard plant barcoding loci, incorporating sequence data from both BOLD and NCBI, and further automated analyses.

Applications of pollen DNA barcoding and metabarcoding

Following the recent methodological developments in the DNA barcoding of pollen and pollen DNA metabarcoding, these methods have been used in novel applications, including pollination biology (Galimberti et al. 2014; Keller et al. 2015; Richardson et al. 2015a; Sickel et al. 2015), food provenance monitoring (Bruni et al. 2015; Hawkins et al. 2015), and airborne allergen monitoring (Kraaijeveld et al. 2015). Several other applications are already feasible with the current technology, but they are yet to be applied. Here we consider a number of potential and realized ecological and socio-economic ap-

plications that are within the fields of ecology, provenance tracking, human health, and palaeobiology.

Plant–pollinator interactions over space and time

The movement of pollen is of importance to the long-term structure and function of plant communities, whether natural or managed (Biesmeijer et al. 2006; Klein et al. 2007; Ricketts et al. 2008; Jordano 2010). Advances in pollen DNA metabarcoding could afford researchers a more highly resolved understanding of pollination biology at broader scales and greater sampling intensities than previously possible, by enabling characterization of complex pollen assemblages collected from either pollinators or plant stigmatic surfaces. Further, pollen metabarcoding is likely to enhance taxonomic resolution and sensitivity for studies investigating ecological phenomena such as plant–pollinator networks, facilitation or competition between plants or pollinators, and plant biogeography.

Additionally, broad patterns in plant–pollinator biology over space and time could be explored. These include application to anthropogenic environmental landscape changes, such as habitat fragmentation (e.g., Steffan-Dewenter 2003; Brosi et al. 2008) and climate change (e.g., Walther et al. 2002; Inouye 2008; Hegland et al. 2009). Such work could particularly take advantage of historical specimen collections, especially of insect pollinators, many of which were carrying pollen when collected. Specimens are typically labeled with descriptors such as date and place of collection, the collector, and sometimes their association, such as the plants on which they were collected (Pennisi 2000). The use of DNA metabarcoding to track changes in plant communities, and plant–pollinator interactions, over time, is likely particularly valuable in terms of its potential to generate results with conservation implications, such as community reference states for ecological restoration projects.

Ancient pollen DNA barcoding

The study of fossil pollen has various applications in paleoecology, archaeology, as well as anthropology, including the reconstruction of ancient plant communities (Jørgensen et al. 2012a; Pedersen et al. 2013), the study of population dynamics in single plant species (Parducci et al. 2005; Magyari et al. 2011), and the investigation of biodiversity with the aim of endemic species conservation (Wilmshurst et al. 2014) inter alia. Mathewes (2006) also showed how the study of ancient pollen plays an important role in forensics (see Forensic palynology section, below). Applying ancient DNA methods has the potential to improve efficiency and accuracy of ancient pollen identifications over traditional microscopy-based methods.

Most ancient DNA barcoding studies have used sedimentary ancient DNA (sedaDNA) to provide complementary data to macrofossil identification and classic

palynology (Jørgensen et al. 2012a; Parducci et al. 2013; Pedersen et al. 2013). A significant amount of pollen is likely present in the samples taken, together with plant fragments and anything else present in the sediment. Parducci et al. (2013) showed that *sedaDNA* metabarcoding is better at detecting plants that produce restricted amounts of pollen, taxa that are difficult to identify with palynology, and rare plants. It is not certain whether the result is from pollen, or DNA from other parts of plants taken from the sediment sample. Within pollen grains, DNA can be preserved for millennia if environmental conditions are suitable. The presence of DNA in fossil pollen can be confirmed with specific dyes, and Suyama et al. (1996) were able to extract DNA from pollen 150 000 years old.

This genetic tool could be used to identify pollen samples that have been either left unidentified or have been partially identified microscopically but warrant further investigation, such as in forensic cases (Bell et al. 2016) and paleobotany (Suyama et al. 1996). Challenges remain in applying DNA metabarcoding to ancient pollen DNA. Degradation due to hydrolysis and oxidation in metabolically inactive cells complicates amplification of ancient DNA, making it very hard to obtain amplification products that exceed 500 base pairs (bp) (Paabo 1989; Willerslev and Cooper 2005). Therefore, paleoecologists have adapted standard barcoding methods, especially through the use of “mini-barcodes” comprised of shorter amplicons (Jørgensen et al. 2012a, 2012b; Parducci et al. 2013; Wilmschurst et al. 2014). The most commonly used plant mini-barcode is the P6 loop of the *trnL* intron in the chloroplast (Taberlet et al. 2007). In many ancient applications, using multiple markers could potentially provide a much clearer picture of past vegetation. For example, the *trnL* mini-barcode provides accurate resolution at family level (Taberlet et al. 2007). Similar levels of resolution have been noted for mini-barcodes based on *rbcl* (Little 2014). Plant mini-barcode marker development is a key need in terms of increasing the accuracy of taxon identification in ancient DNA. Additionally, comparative sequence databases need to be expanded to include now-extinct plants, which is admittedly challenging.

Food quality and provenance monitoring

Pollen DNA metabarcoding also has the potential to be widely applied to studies of food provenance and quality. Pollen is a nearly ubiquitous environmental biomarker and most foodstuffs (and other products) are likely to contain pollen that can be used to trace the geographic and potentially temporal provenance of products. The clearest application is in tracing the geographic and botanical origins of honey, given its derivation from flowers. Honey is a high-value nutritional product and its taste, food quality, and safety differ depending on the plants the honeybees have foraged upon (Crane 1975). Product labeling guidelines therefore often require the

floral source of commercially sold honey to be declared (Bruni et al. 2015). Honeys labeled as monofloral differ from multifloral honeys by the dominance of nectar and pollen from a single plant species. Honeys are classified as monofloral if the pollen content of one species is greater than 45% (Anklam 1998). Monofloral honeys often have higher commercial value and are therefore prone to fraudulent adulterations and incorrect labeling (Persano Oddo and Bogdanov 2004). Food safety and quality is also of concern as pollen from poisonous plants can sometimes be found within honey, for example *Atropa belladonna* (Bruni et al. 2015). Hepatotoxic pyrrolizidine alkaloids (PAs) have been detected in honeys after bees have foraged on plants within the Boraginaceae (Edgar et al. 2002). Poisoning by “mad honey” has been documented, caused by eating honey containing grayanotoxins arising from *Rhododendron* spp. (Koca and Koca 2007). Understanding the botanical profile of honey is therefore very important to ensure that products are of high quality, and safe for the consumer (Olivieri et al. 2012).

As with other applications, DNA barcoding applications hold considerable promise over traditional melissopalynology, the examination of pollen found within honey using light microscopy (Louveaux et al. 1978; Bruni et al. 2015). Laube et al. (2010) successfully used real-time PCR to identify different plant species frequently found within Corsican honey, but this method requires an a priori knowledge of the species likely to be found. Galimberti et al. (2014) and Bruni et al. (2015) amplified the *rbcl* and *trnH-psbA* plastid markers to identify the floral composition of honey from the Italian Alps. Their method used cloning to sequence individual amplicons, which places limits on the depth of sequencing that can be achieved. Valentini et al. (2010) trialed the use of pyrosequencing amplicons of the *trnL* (UAA) intron to characterize two commercial honeys. Hawkins et al. (2015) used the *rbcl* marker and 454 pyrosequencing to characterize nine honeys from domestic beekeepers in the UK. They showed that DNA metabarcoding provided much greater levels of repeatability compared to melissopalynology.

Airborne allergen monitoring

Plant pollen is one of the major allergens contributing to respiratory disease, and causes a substantial economic burden in terms of the number of drugs purchased, clinics visited, and loss of productivity due to employees being absent from work (D'Amato et al. 2005; Davies et al. 2015). Disease symptom severity differs based on the taxonomic origin of the pollen allergen (Hrabina et al. 2008) and the aeroallergen concentration (Hrabina et al. 2008; Erbas et al. 2012). Both patients and health services have more control over health problems when they have access to this information (Davies et al. 2015; Kraaijeveld et al. 2015). Many pollen-monitoring programs have been implemented to date, but the current

methodology used makes them unsuitable to implement in continent-wide networks (Kraaijeveld et al. 2015).

Pollen monitoring is mainly performed by volumetric pollen samplers, whirling arm samplers, or passive samplers. National pollen monitoring networks often use Hirst-type volumetric samplers that inspire ambient air and immobilize particulates on sticky tape (Scheifinger et al. 2013). Microscopic counting and taxonomic identification of pollen are then performed on the tape, often after staining with appropriate dyes, with the same limitations of microscopic identification we have previously described.

DNA metabarcoding has been successfully used to make species-level identifications of airborne pollen for the monitoring of allergens in the Netherlands (Kraaijeveld et al. 2015). Pollen samples were obtained by volumetric samplers, and a comparison drawn between microscopic palynology and DNA analyses. Increased taxonomic resolution among classifications was achieved using DNA metabarcoding, while microscopy could only identify pollen to family level in many cases. One possible extension specific to aerobiological monitoring is taxonomic expansion from simply identifying pollen, to monitoring for occurrences of certain bacteria or fungi of health interest.

Forensic palynology

Forensic palynology is the use of pollen to link persons or objects with particular places and times (Horrocks and Walsh 1998; Taylor and Skene 2003; Bryant and Jones 2006; Mathewes 2006). This technique is of great utility to forensics because (i) pollen is a nearly ubiquitous feature of the environment; (ii) different geographic locations have different pollen signatures, allowing for inference related to spatial tracking; (iii) plants flower at different times, allowing for temporal inference; and (iv) pollen is extremely durable (hence its widespread use in paleontological studies) and thus can be utilized for forensic studies for decades or longer after sample collection (Horrocks and Walsh 1998; Bryant and Jones 2006; Mathewes 2006; Mildenhall 2006; Mildenhall et al. 2006; Wiltshire 2006; Walsh and Horrocks 2008).

Like many other palynological applications, forensic palynology as currently practiced is reliant on visual microscopic identification of pollen grains by an expert palynologist (Bryant and Jones 2006; Mildenhall et al. 2006; Walsh and Horrocks 2008), and DNA metabarcoding could very likely increase its applicability to a broader range of situations (Bell et al. 2016). For example, forensic palynology could take greater advantage of DNA barcoding when combined with a universal database of geographic and temporal knowledge of plants (Goodman et al. 2015). Such a database could also be of potential utility to fields outside of forensics (airborne pollen monitoring for allergens, pollination biology, biodiversity inventories, and potentially even monitoring of plant populations).

As with any genetic analysis, DNA barcoding requires destructive sampling of pollen grains, which means the pollen can no longer be analyzed with morphological methods, including examination of other particles such as biosilicates, ashes, etc. One possible work-around to this issue is to split pollen samples into partitions for DNA extraction, morphological examination, and permanent storage. Comparing the results of the DNA barcoding and morphological identifications for consistency could validate the accuracy of this approach. Although work is needed on these technical issues, and others including in some cases adaptations for very small samples, the method is very close to being feasible for routine analysis in forensics, and improvements are occurring rapidly. We envision that DNA metabarcoding will greatly expand the use of pollen as a biomarker, giving forensic scientists new leads and evidence toward enhancing global security and justice.

The future of pollen DNA metabarcoding

Applications of pollen DNA barcoding and metabarcoding are by no means limited to the above-mentioned examples. Other applications may include assessments of pollination efficiency of wind-pollinated plant species, monitoring gene flow between populations and between hybridizing species, monitoring climate change via changes in plant phenology, and many more. Pollen DNA barcoding and metabarcoding methods are likely to become applicable to an increasingly broad range of research questions as technical issues are resolved and laboratory techniques and bioinformatics pipelines become more standardized and user friendly. Clearly, the case studies outlined above demonstrate that there is great potential for the use of pollen DNA barcoding in a wide variety of ecological and socio-economic applications. In particular, the development of pollen barcoding technology could be of immediate benefit to address questions in pollination biology, climate change, invasive species, plant conservation, and agriculture. Moreover, making data available is of high importance for issues related to environmental protection, human health, and food security. Technical developments to make methods more precise, consistent, and quantitative will go a long way to facilitating the DNA barcoding of pollen and future applications.

Acknowledgements

K.L.B., K.S.B., and B.J.B. were supported by the US Army Research Office (grants W911NF-13-1-0247 and W911NF-13-1-0100) for funding. A.K. has been supported by the German Research Foundation (DFG, KE1743/4-1). A.G. has been supported by the South African Department of Science and Technology/National Research Foundation Professional Development Programme (99781) and the University of KwaZulu-Natal. We thank K. Kraaijeveld and two anonymous reviewers for providing comments on the manuscript.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**(3): 403–410. doi:10.1016/S0022-2836(05)80360-2. PMID:2231712.
- Ankenbrand, M.J., Keller, A., Wolf, M., Schultz, J., and Förster, F. 2015. ITS2 Database V: Twice as Much: Table 1. *Mol. Biol. Evol.* **32**(11): 3030–3032. doi:10.1093/molbev/msv174. PMID:26248563.
- Anklam, E. 1998. A review of the analytical methods to determine the geographical and botanical origin of honey. *Food Chem.* **63**(4): 549–562. doi:10.1016/S0308-8146(98)00057-0.
- Aziz, A.N., and Sauve, R.J. 2008. Genetic mapping of *Echinacea purpurea* via individual pollen DNA fingerprinting. *Mol. Breed.* **21**(2): 227–232. doi:10.1007/s11032-007-9123-9.
- Bell, K.L., Burgess, K.S., Okamoto, K.C., Aranda, R., and Brosi, B.J. 2016. Review and future prospects for DNA barcoding methods in forensic palynology. *Foren. Sci. Int. Genet.* **21**: 110–116. doi:10.1016/j.fsigen.2015.12.010.
- Benson, D.A., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. 2015. GenBank. *Nucleic Acids Res.* **43**(Database issue): D30–D35. doi:10.1093/nar/gku1216.
- Biesmeijer, J.C., Roberts, S.P.M., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T., et al. 2006. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science*, **313**(5785): 351–354. doi:10.1126/science.1127863. PMID:16857940.
- Borisenko, A.V., Sones, J.E., and Hebert, P.D.N. 2009. The front-end logistics of DNA barcoding: challenges and prospects. *Mol. Ecol. Resour.* **9**(s1): 27–34. doi:10.1111/j.1755-0998.2009.02629.x. PMID:21564961.
- Brooks, J.P., Edwards, D.J., Harwich, M.D., Rivera, M.C., Fettweis, J.M., Serrano, M.G., et al. 2015. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol.* **15**: 66. doi:10.1186/s12866-015-0351-6. PMID:25880246.
- Brosi, B.J., Daily, G.C., Shih, T.M., Oviedo, F., and Durán, G. 2008. The effects of forest fragmentation on bee communities in tropical countryside. *J. Appl. Ecol.* **45**(3): 773–783. doi:10.1111/j.1365-2664.2007.01412.x.
- Bruni, I., Galimberti, A., Caridi, L., Scaccabarozzi, D., De Mattia, F., Casiraghi, M., and Labra, M. 2015. A DNA barcoding approach to identify plant species in multiflower honey. *Food Chem.* **170**: 308–315. doi:10.1016/j.foodchem.2014.08.060. PMID:25306350.
- Bryant, V.M., and Jones, G.D. 2006. Forensic palynology: current status of a rarely used technique in the United States of America. *Foren. Sci. Int.* **163**(3): 183–197. doi:10.1016/j.forsciint.2005.11.021. PMID:16504436.
- Burgess, K.S., Fazekas, A.J., Kesanakurti, P.R., Graham, S.W., Husband, B.C., Newmaster, S.G., et al. 2011. Discriminating plant species in a local temperate flora using the *rbcl+matK* DNA barcode. *Methods Ecol. Evol.* **2**: 333–340. doi:10.1111/j.2041-210X.2011.00092.x.
- CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U.S.A.* **106**(31): 12794–12797. doi:10.1073/pnas.0905845106. PMID:19666622.
- Champlot, S., Berthelot, C., Pruvost, M., Bennett, E.A., Grange, T., and Geigl, E.M. 2010. An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS ONE*, **5**(9): e13042. doi:10.1371/journal.pone.0013042. PMID:20927390.
- 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**(1): e8613. doi:10.1371/journal.pone.0008613. PMID:20062805.
- Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., and Zhou, S. 2016. Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. *Mol. Ecol. Resour.* **16**(1): 138–149. doi:10.1111/1755-0998.12438. PMID:26084789.
- Cornman, R.S., Otto, C.R.V., Iwanowicz, D., and Pettis, J.S. 2015. Taxonomic characterization of honey bee (*Apis mellifera*) pollen foraging based on non-overlapping paired-end sequencing of nuclear ribosomal loci. *PLoS ONE*, **10**(12): e0145365. doi:10.1371/journal.pone.0145365. PMID:26700168.
- Corriveau, J.L., and Coleman, A.W. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Am. J. Bot.* **75**(10): 1443–1458. doi:10.2307/2444695.
- Crane, E. 1975. Honey. A comprehensive survey. William Heinemann Ltd., London.
- D'Amato, G., Liccardi, G., D'Amato, M., and Holgate, S. 2005. Environmental risk factors and allergic bronchial asthma. *Clin. Exp. Allergy*, **35**(9): 1113–1124. doi:10.1111/j.1365-2222.2005.02328.x. PMID:16164436.
- Davies, J.M., Beggs, P.J., Medek, D.E., Newnham, R.M., Erbas, B., Thibaudon, M., et al. 2015. Trans-disciplinary research in synthesis of grass pollen aerobiology and its importance for respiratory health in Australasia. *Sci. Total Environ.* **534**: 85–96. doi:10.1016/j.scitotenv.2015.04.001. PMID:25891684.
- de Vere, N., Rich, T.C.G., Ford, C.R., Trinder, S.A., Long, C., Moore, C.W., et al. 2012. DNA barcoding the native flowering plants and conifers of Wales. *PLoS ONE*, **7**(6): e37945. doi:10.1371/journal.pone.0037945. PMID:22701588.
- Dong, W., Cheng, T., Changhao, L., Xu, C., Long, P., Chen, C., and Zhou, S. 2014. Discriminating plants using the DNA barcode *rbclb*: an appraisal based on a large data set. *Mol. Ecol. Res.* **14**(2): 336–343. doi:10.1111/1755-0998.12185.
- Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**(19): 2460–2461. doi:10.1093/bioinformatics/btq461. PMID:20709691.
- Edgar, R.C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods*, **10**: 996–998. doi:10.1038/nmeth.2604. PMID:23955772.
- Edgar, J.A., Roeder, E., and Molyneux, R.J. 2002. Honey from plants containing pyrrolizidine alkaloids: a potential threat to health. *J. Agric. Food Chem.* **50**(10): 2719–2730. doi:10.1021/jf0114482.
- Engel, P., Martinson, V.G., and Moran, N.A. 2012. Functional diversity within the simple gut microbiota of the honey bee. *Proc. Natl. Acad. Sci. U.S.A.* **109**(27): 11002–11007. doi:10.1073/pnas.1202970109. PMID:22711827.
- Erbaş, B., Akram, M., Dharmage, S.C., Tham, R., Dennekamp, M., Newbigin, E., et al. 2012. The role of seasonal grass pollen on childhood asthma emergency department presentations. *Clin. Exp. Allergy*, **42**(5): 799–805. doi:10.1111/j.1365-2222.2012.03995.x. PMID:22515396.
- Fazekas, A.J., and Kesanakurti, P.R. 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Mol. Ecol. Resour.* **9**(S1): 130–139. doi:10.1111/j.1755-0998.2009.02652.x. PMID:21564972.
- Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., et al. 2008. Multiple multi-locus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE*, **3**(7): e2802. doi:10.1371/journal.pone.0002802. PMID:18665273.
- Ford, C.S., Ayres, K.L., Toomey, N., Haider, N., Stahl, J., van A., Kelly, L.J., et al. 2009. Selection of candidate coding DNA barcoding regions for use on land plants. *Bot. J. Linn. Soc.* **159**: 1–11. doi:10.1111/j.1095-8339.2008.00938.x.
- Fujiwara, M.T., Hashimoto, H., Kazama, Y., Hirano, T., Yoshioka, Y., Aoki, S., et al. 2010. Dynamic morphologies of pollen plastids visualised by vegetative-specific FtsZ1-GFP in *Arabidopsis thaliana*. *Protoplasma*, **242**(1): 19–33. doi:10.1007/s00709-010-0119-7. PMID:20195657.

- 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**(10): e109363. doi:10.1371/journal.pone.0109363. PMID: 25296114.
- Gibson, J., Shokralla, S., Porter, T.M., King, I., van Konynenburg, S., Janzen, D.H., et al. 2014. Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metabarcoding. *Proc. Natl. Acad. Sci. U.S.A.* **111**(22): 8007–8012. doi:10.1073/pnas.1406468111. PMID: 24808136.
- Goodman, F.J., Doughty, J.W., Gary, C., Christou, C.T., Hu, B.B., Hultman, E.A., et al. 2015. PIGLT: a pollen identification and geolocation system for forensic applications. In 2015 IEEE International Symposium on Technologies for Homeland Security (HST). pp. 1–7. doi:10.1109/THS.2015.7225271.
- Hajibabaei, M., Spall, J.L., Shokralla, S., and van Konynenburg, S. 2012. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecol.* **12**(1): 28. doi:10.1186/1472-6785-12-28. PMID:23259585.
- Hawkins, J., de Vere, N., Griffith, A., Ford, C.R., Allainguillaume, J., Hegarty, M.J., et al. 2015. Using DNA metabarcoding to identify the floral composition of honey: a new tool for investigating honey bee foraging preferences. *PLoS ONE*, **10**(8): e0134735. doi:10.1371/journal.pone.0134735. PMID:26308362.
- Hebert, P.D., Cywinska, A., Ball, S.L., and deWaard, J.R. 2003a. Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* **270**: 313–321. doi:10.1098/rspb.2002.2218.
- Hebert, P.D.N., Ratnasingham, S., and deWaard, J.R. 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. B Biol. Sci.* **270**(Suppl.): S96–S99. doi:10.1098/rsbl.2003.0025.
- Hegland, S.J., Nielsen, A., Lázaro, A., Bjerknes, A.-L., and Totland, Ø. 2009. How does climate warming affect plant–pollinator interactions? *Ecol. Lett.* **12**(2): 184–195. doi:10.1111/j.1461-0248.2008.01269.x. PMID:19049509.
- Hollingsworth, M.L., Clark, A.A., Forrest, L.L., Richardson, J., Pennington, R.T., Long, D.G., et al. 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol. Ecol. Resour.* **9**: 439–457. doi:10.1111/j.1755-0998.2008.02439.x. PMID:21564673.
- Hollingsworth, P.M., Graham, S.W., and Little, D.P. 2011. Choosing and using a plant DNA barcode. *PLoS ONE*, **6**(5): e19254. doi:10.1371/journal.pone.0019254. PMID:21637336.
- Holt, K.A., and Bennett, K.D. 2014. Principles and methods for automated palynology. *New Phytol.* **203**(3): 735–742. doi:10.1111/nph.12848. PMID:25180326.
- Horrocks, M., and Walsh, K.A.J. 1998. Forensic palynology: assessing the value of the evidence. *Rev. Palaeobot. Palynol.* **103**: 69–74. doi:10.1016/S0034-6667(98)00027-X.
- Hrabina, M., Peltre, G., Van, Ree, R., and Moingeon, P. 2008. Grass pollen allergens. *Clin. Exp. Allergy Rev.* **8**: 7–11. doi:10.1111/j.1472-9733.2008.00126.x.
- Inouye, D.W. 2008. Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. *Ecology*, **89**(2): 353–362. doi:10.1890/06-2128.1. PMID:18409425.
- Janzen, D.H., Hallwachs, W., Burns, J., Solis, M.A., and Woodley, N.E. 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Mol. Ecol. Resour.* **9**: 1–26. doi:10.1111/j.1755-0998.2009.02628.x. PMID:21564960.
- Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., et al. 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecol. Lett.* **16**(10): 1245–1257. doi:10.1111/ele.12162. PMID:23910579.
- Jordano, P. 2010. Pollen, seeds and genes: the movement ecology of plants. *Heredity*, **105**(4): 329–330. doi:10.1038/hdy.2010.28. PMID:20332803.
- Jørgensen, T., Haile, J., Möller, P., Andreev, A., Boessenkool, S., Rasmussen, M., et al. 2012a. A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Mol. Ecol.* **21**(8): 1989–2003. doi:10.1111/j.1365-294X.2011.05287.x.
- Jørgensen, T., Kjær, K.H., Haile, J., Rasmussen, M., Boessenkool, S., Andersen, K., et al. 2012b. Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA Meta-barcoding. *Mol. Ecol.* **21**(8): 1980–1988. doi:10.1111/j.1365-294X.2011.05278.x.
- Kane, N., Sveinsson, S., Dempewolf, H., Yang, J.Y., Zhang, D., Engels, J.M.M., and Cronk, Q. 2012. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *Am. J. Bot.* **99**(2): 320–9. doi:10.3732/ajb.1100570.
- Kearns, C.A., and Inouye, D.W. 1993. Techniques for pollination biologists. University Press of Colorado, Niwot, Colo.
- 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**(2): 558–566. doi:10.1111/plb.12251. PMID:25270225.
- Kesanakurti, P.R., Fazekas, A.J., Burgess, K.S., Percy, D.M., Newmaster, S.G., Graham, S.W., et al. 2011. Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. *Mol. Ecol.* **20**: 1289–1302. doi:10.1111/j.1365-294X.2010.04989.x.
- Khansari, E., Zarre, S., Alizadeh, K., Attar, F., Aghabeigi, F., and Salmaki, Y. 2012. Pollen morphology of *Campanula* (Campanulaceae) and allied genera in Iran with special focus on its systematic implication. *Flora Morphol. Distrib. Funct. Ecol. Plants*, **207**(3): 203–211. doi:10.1016/j.flora.2012.01.006.
- Klein, A.-M., Vaissiere, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., and Tscharntke, T. 2007. Importance of pollinators in changing landscapes for world crops. *Proc. R. Soc. B Biol. Sci.* **274**: 303–313. doi:10.1098/rspb.2006.3721. PMID:17164193.
- Koca, I., and Koca, A.F. 2007. Poisoning by mad honey: a brief review. *Food Chem. Toxicol.* **45**(8): 1315–1318. doi:10.1016/j.fct.2007.04.006. PMID:17540490.
- Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. 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**(17): 5112–5120. doi:10.1128/AEM.01043-13. PMID:23793624.
- Kraaijeveld, K., de Weger, L.A., Ventayol, García, M., Buermans, H., Frank, J., Hiemstra, P.S., and den Dunnen, J.T. 2015. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Mol. Ecol. Resour.* **15**(1): 8–16. doi:10.1111/1755-0998.12288. PMID:24893805.
- Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., and Janzen, D.H. 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* **102**(23): 8369–8374. doi:10.1073/pnas.0503123102. PMID:15928076.
- Kress, W.J., Erickson, D.L., Jones, F.A., Swenson, N.G., Perez, R., Sanjurjo, O., and Bermingham, E. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proc. Natl. Acad. Sci.* **106**(44): 18621–18626. doi:10.1073/pnas.0909820106.
- Laube, I., Hird, H., Brodmann, P., Ullmann, S., Schöne-Michling, M., Chisholm, J., and Broll, H. 2010. Development of primer and probe sets for the detection of plant species in honey. *Food Chem.* **118**(4): 979–986. doi:10.1016/j.foodchem.2008.09.063.
- Little, D.P. 2014. A DNA mini-barcode for land plants. *Mol.*

- Ecol. Resour. **14**(3): 437–446. doi:10.1111/1755-0998.12194. PMID:24286499.
- Longhi, S., Cristofori, A., Gatto, P., Cristofolini, F., Grando, M.S., and Gottardini, E. 2009. Biomolecular identification of allergenic pollen: a new perspective for aerobiological monitoring? *Ann. Allergy Asthma Immunol.* **103**(6): 508–514. doi:10.1016/S1081-1206(10)60268-2.
- Louveaux, J., Maurizio, A., and Vorwohl, G. 1978. Methods of melissopalynology. *Bee World*, **59**(4): 139–157. doi:10.1080/0005772X.1978.11097714.
- Magyari, E.K., Major, A., Balint, M., Nedli, J., Braun, M., Racz, I., and Parducci, L. 2011. Population dynamics and genetic changes of *Picea abies* in the South Carpathians revealed by pollen and ancient DNA analyses. *BMC Evol. Biol.* **11**(1): 66. doi:10.1186/1471-2148-11-66. PMID:21392386.
- Mashima, J., Kodama, Y., Kosuge, T., Fujisawa, T., Katayama, T., Nagasaki, H., et al. 2015. DNA data bank of Japan (DDBJ) progress report. *Nucleic Acids Res.* **44**(D1): D51–D57. doi:10.1093/nar/gkv1105.
- Mathewes, R.W. 2006. Forensic palynology in Canada: an overview with emphasis on archaeology and anthropology. *Foren. Sci. Int.* **163**: 198–203. doi:10.1016/j.forsciint.2006.06.069. PMID:16901669.
- Matsuki, Y., Isagi, Y., and Suyama, Y. 2007. The determination of multiple microsatellite genotypes and DNA sequences from a single pollen grain. *Mol. Ecol. Notes*, **7**(2): 194–198. doi:10.1111/j.1471-8286.2006.01588.x.
- Matsunaga, S., Schutze, K., Donnison, I.S., Grant, S.R., Kuroiwa, T., and Kawano, S. 1999. Single pollen typing combined with laser-mediated manipulation. *Plant J.* **20**(3): 371–378. doi:10.1046/j.1365-3113.1999.00612.x. PMID:10571898.
- Matsushima, R., Tang, L.Y., Zhang, L., Yamada, H., Twell, D., and Sakamoto, W. 2011. A conserved, Mg²⁺-dependent exonuclease degrades organelle DNA during *Arabidopsis* pollen development. *Plant Cell*, **23**: 1608–1624. doi:10.1105/tpc.111.084012. PMID:21521697.
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., et al. 2008. The metagenomics RAST server — a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform.* **9**(1): 386. doi:10.1186/1471-2105-9-386. PMID:18803844.
- Mildenhall, D.C. 2006. Hypericum pollen determines the presence of burglars at the scene of a crime: an example of forensic palynology. *Foren. Sci. Int.* **163**: 231–235. doi:10.1016/j.forsciint.2005.11.028. PMID:16406430.
- Mildenhall, D.C., Wiltshire, P.E.J., and Bryant, V.M. 2006. Forensic palynology. *Foren. Sci. Int.* **163**(3): 161–162. doi:10.1016/j.forsciint.2006.07.013.
- Nagata, N., Saito, C., Sakai, A., Kuroiwa, H., and Kuroiwa, T. 1999. The selective increase or decrease of organellar DNA in generative cells just after pollen mitosis one controls cytoplasmic inheritance. *Planta*, **209**(1): 53–65. doi:10.1007/s004250050606. PMID:10467031.
- Olivieri, C., Marota, I., Rollo, F., and Luciani, S. 2012. Tracking plant, fungal, and bacterial DNA in honey specimens. *J. Foren. Sci.* **57**(1): 222–227. doi:10.1111/j.1556-4029.2011.01964.x. PMID:22074557.
- Paabo, S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 1939–1943. doi:10.1073/pnas.86.6.1939. PMID:2928314.
- Parducci, L., Suyama, Y., Lascoux, M., and Bennett, K. 2005. Ancient DNA from pollen: a genetic record of population history in Scots pine. *Mol. Ecol.* **14**(9): 2873–2882. doi:10.1111/j.1365-294X.2005.02644.x.
- Parducci, L., Matetovici, I., Fontana, S.L., Bennett, K.D., Suyama, Y., Haile, J., et al. 2013. Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Mol. Ecol.* **22**(13): 3511–3524. doi:10.1111/mec.12298. PMID:23587049.
- Pedersen, M.W., Ginolhac, A., Orlando, L., Olsen, J., Andersen, K., Holm, J., et al. 2013. A comparative study of ancient environmental DNA to pollen and microfossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quat. Sci. Rev.* **75**: 161–168. doi:10.1016/j.quascirev.2013.06.006.
- Pennisi, E. 2000. Taxonomic revival. *Science*, **289**(5488): 2306–2308. doi:10.1126/science.289.5488.2306. PMID:11041799.
- Persano Oddo, L., and Bogdanov, S. 2004. Determination of honey botanical origin: problems and issues. *Apidologie*, **35**(S1): S2–S3. doi:10.1051/apido:2004044.
- Petersen, G., Johansen, B., and Seberg, O. 1996. PCR and sequencing from a single pollen grain. *Plant Mol. Biol.* **31**(1): 189–191. doi:10.1007/BF00020620. PMID:8704154.
- Pompanon, F., Deagle, B.E., Symondson, W.O.C., Brown, D.S., Jarman, S.N., and Taberlet, P. 2012. Who is eating what: diet assessment using next generation sequencing. *Mol. Ecol.* **21**(8): 1931–1950. doi:10.1111/j.1365-294X.2011.05403.x.
- Rahl, M. 2008. Microscopic identification and purity determination of pollen grains. *Methods Mol. Med.* **138**: 263–9. doi:10.1007/978-1-59745-366-0_22. PMID:18612615.
- Richardson, R.T., Lin, C.-H., Quijia, J.O., Riusech, N.S., Goodell, K., and Johnson, R.M. 2015a. Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. *Appl. Plant Sci.* **3**(11): 1500043. doi:10.3732/apps.1500043.
- Richardson, R.T., Lin, C.-H., Sponsler, D.B., Quijia, J.O., Goodell, K., and Johnson, R.M. 2015b. Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. *Appl. Plant Sci.* **3**(1): 1400066. doi:10.3732/apps.1400066.
- Ricketts, T.H., Regetz, J., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Bogdanski, A., et al. 2008. Landscape effects on crop pollination services: are there general patterns? *Ecol. Lett.* **11**(5): 499–515. doi:10.1111/j.1461-0248.2008.01157.x. PMID:18294214.
- Roberts, R.J., Carneiro, M.O., and Schatz, M.C. 2013. The advantages of SMRT sequencing. *Genome Biol.* **14**(7): 405. doi:10.1186/gb-2013-14-6-405. PMID:23822731.
- Rogers, S.O., and Bendich, A.J. 1987. Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Mol. Biol.* **9**(5): 509–520. doi:10.1007/BF00015882. PMID:24277137.
- Sakamoto, W., Miyagishima, S., and Jarvis, P. 2008. Chloroplast biogenesis: control of plastid development, protein import, division and inheritance. *Arabidopsis Book*, **6**: e0110. doi:10.1199/tab.0110. PMID:22303235.
- Salmaki, Y., Jamzad, Z., Zarre, S., and Bräuchler, C. 2008. Pollen morphology of *Stachys* (Lamiaceae) in Iran and its systematic implication. *Flora Morphol. Distrib. Funct. Ecol. Plants*, **203**(8): 627–639. doi:10.1016/j.flora.2007.10.005.
- Sangwan, R.S., and Sangwan-Norreel, B.S. 1987. Ultrastructural cytology of plastids in pollen grains of certain androgenic and nonandrogenic plants. *Protoplasma*, **138**: 11–22. doi:10.1007/BF01281180.
- Scheifinger, H., Belmonte, J., Buters, J., Celenk, S., Damialis, A., Dechamp, C., et al. 2013. Monitoring, modelling and forecasting of the pollen season. In *Allergenic pollen*. Edited by M. Sofiev and K.-C. Bergmann. Springer, Dordrecht, the Netherlands. pp. 71–126. doi:10.1007/978-94-007-4881-1_4.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., et al. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**(23): 7537–7541. doi:10.1128/AEM.01541-09. PMID:19801464.
- Shokralla, S., Spall, J.L., Gibson, J.F., and Hajibabaei, M. 2012.

- Next-generation sequencing technologies for environmental DNA research. *Mol. Ecol.* **21**(8): 1794–1805. doi:[10.1111/j.1365-294X.2012.05538.x](https://doi.org/10.1111/j.1365-294X.2012.05538.x).
- Sickel, W., Ankenbrand, M.J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., et al. 2015. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecol.* **15**(1): 20. doi:[10.1186/s12898-015-0051-y](https://doi.org/10.1186/s12898-015-0051-y). PMID:[26194794](https://pubmed.ncbi.nlm.nih.gov/26194794/).
- Southworth, D. 1974. Solubility of pollen exines. *Am. J. Bot.* **61**(1): 36–44. doi:[10.2307/2441242](https://doi.org/10.2307/2441242).
- Squizzato, S., Park, Y.M., Buso, N., Gur, T., Cowley, A., Li, W., et al. 2015. The EBI Search engine: providing search and retrieval functionality for biological data from EMBL-EBI. *Nucleic Acids Res.* **43**(W1): W585–W588. doi:[10.1093/nar/gkv316](https://doi.org/10.1093/nar/gkv316). PMID:[25855807](https://pubmed.ncbi.nlm.nih.gov/25855807/).
- Steffan-Dewenter, I. 2003. Importance of habitat area and landscape context for species richness of bees and wasps in fragmented orchard meadows. *Conserv. Biol.* **17**(4): 1036–1044. doi:[10.1046/j.1523-1739.2003.01575.x](https://doi.org/10.1046/j.1523-1739.2003.01575.x).
- Suyama, Y., Kawamuro, K., Kinoshita, I., Yoshimura, K., Tsumura, Y., and Takahara, H. 1996. DNA sequence from a fossil pollen of *Abies* spp. from Pleistocene peat. *Genes Genet. Syst.* **71**(3): 145–149. doi:[10.1266/ggs.71.145](https://doi.org/10.1266/ggs.71.145). PMID:[8828176](https://pubmed.ncbi.nlm.nih.gov/8828176/).
- Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., et al. 2007. Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* **35**(3): e14–e14. doi:[10.1093/nar/gkl938](https://doi.org/10.1093/nar/gkl938). PMID:[17169982](https://pubmed.ncbi.nlm.nih.gov/17169982/).
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., and Willerslev, E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**(8): 2045–2050. doi:[10.1111/j.1365-294X.2012.05470.x](https://doi.org/10.1111/j.1365-294X.2012.05470.x).
- Tang, M., Tan, M., Meng, G., Yang, S., Su, X., Liu, S., et al. 2014. Multiplex sequencing of pooled mitochondrial genomes—a crucial step toward biodiversity analysis using metagenomics. *Nucleic Acids Res.* **42**(22): e166. doi:[10.1093/nar/gku917](https://doi.org/10.1093/nar/gku917). PMID:[25294837](https://pubmed.ncbi.nlm.nih.gov/25294837/).
- Taylor, B., and Skene, K.R. 2003. Forensic palynology: spatial and temporal considerations of spore deposition in forensic investigations. *Aust. J. Foren. Sci.* **35**: 193–204. doi:[10.1080/00450610309410582](https://doi.org/10.1080/00450610309410582).
- Valentini, A., Miquel, C., and Taberlet, P. 2010. DNA barcoding for honey biodiversity. *Diversity*, **2**(4): 610–617. doi:[10.3390/d2040610](https://doi.org/10.3390/d2040610).
- Walsh, K.A.J., and Horrocks, M. 2008. Palynology: Its position in the field of forensic science. *J. Foren. Sci.* **53**(5): 1053–1060. doi:[10.1111/j.1556-4029.2008.00802.x](https://doi.org/10.1111/j.1556-4029.2008.00802.x). PMID:[18636981](https://pubmed.ncbi.nlm.nih.gov/18636981/).
- Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebe, T.J.C., et al. 2002. Ecological responses to recent climate change. *Nature*, **416**: 389–395. doi:[10.1038/416389a](https://doi.org/10.1038/416389a). PMID:[11919621](https://pubmed.ncbi.nlm.nih.gov/11919621/).
- Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**(16): 5261–5267. doi:[10.1128/AEM.00062-07](https://doi.org/10.1128/AEM.00062-07). PMID:[17586664](https://pubmed.ncbi.nlm.nih.gov/17586664/).
- Willerslev, E., and Cooper, A. 2005. Review Paper. Ancient DNA. *Proc. R. Soc. B Biol. Sci.* **272**(1558): 3–16. doi:[10.1098/rspb.2004.2813](https://doi.org/10.1098/rspb.2004.2813). PMID:[15875564](https://pubmed.ncbi.nlm.nih.gov/15875564/).
- Willerslev, E., Hansen, A.J., Binladen, J., Brand, T.B., Gilbert, M.T.P., Shapiro, B., et al. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**(5620): 791–795. doi:[10.1126/science.1084114](https://doi.org/10.1126/science.1084114). PMID:[12702808](https://pubmed.ncbi.nlm.nih.gov/12702808/).
- Wilmshurst, J.M., Moar, N.T., Wood, J.R., Bellingham, P.J., Findlater, A.M., Robinson, J.J., and Stone, C. 2014. Use of pollen and ancient DNA as conservation baselines for offshore islands in New Zealand. *Conserv. Biol.* **28**(1): 202–212. doi:[10.1111/cobi.12150](https://doi.org/10.1111/cobi.12150). PMID:[24024911](https://pubmed.ncbi.nlm.nih.gov/24024911/).
- Wilson, E.E., Sidhu, C.S., LeVan, K.E., and Holway, D.A. 2010. Pollen foraging behaviour of solitary Hawaiian bees revealed through molecular pollen analysis. *Mol. Ecol.* **19**(21): 4823–4829. doi:[10.1111/j.1365-294X.2010.04849.x](https://doi.org/10.1111/j.1365-294X.2010.04849.x).
- Wiltshire, P.E.J. 2006. Consideration of some taphonomic variables of relevance to forensic palynological investigation in the United Kingdom. *Foren. Sci. Int.* **163**: 173–182. doi:[10.1016/j.forsciint.2006.07.011](https://doi.org/10.1016/j.forsciint.2006.07.011). PMID:[16920306](https://pubmed.ncbi.nlm.nih.gov/16920306/).
- Zhang, Q., Liu, Y., and Sodmergen. 2003. Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species. *Plant Cell Physiol.* **44**(9): 941–951. doi:[10.1093/pcp/pgc121](https://doi.org/10.1093/pcp/pgc121). PMID:[14519776](https://pubmed.ncbi.nlm.nih.gov/14519776/).