

DNA metabarcoding of airborne pollen: new protocols for improved taxonomic identification of environmental samples

Kleopatra Leontidou · Cristiano Vernesi · Johannes De Groeve ·
Fabiana Cristofolini · Despoina Vokou · Antonella Cristofori

Received: 21 December 2016 / Accepted: 23 August 2017
© Springer Science+Business Media B.V. 2017

Abstract Metabarcoding is a promising DNA-based method for identifying airborne pollen from environmental samples with advantages over microscopic methods. Sample preparation and DNA extraction are of fundamental importance for obtaining an optimal DNA yield. Currently, there is no standard procedure for these steps, especially for gravimetric pollen samplers. Therefore, the aim of this study was to develop protocols for processing environmental samples for pollen DNA extraction and for metabarcoding analysis and to assess the efficacy of these protocols for the taxonomic assignment of airborne pollen collected by gravimetric (Tauber trap) and volumetric (Hirst-type trap) samplers. Protocols were tested

across an increasing complexity of samples, from pure single-species pollen to environmental multi-species samples. A short fragment (about 150 base pairs) of the chloroplast *trnL* gene was amplified using universal primers for plants. After PCR amplification, amplicons were Sanger-sequenced and taxonomic assignment was accomplished by comparison with a custom-made reference database including chloroplast DNA sequences from most of the anemophilous taxa occurring in the study area (Trentino, northern Italy), representing 46 plant families. Using the classical morphological pollen analysis as a benchmark, we show that DNA metabarcoding is efficient and applicable even in complex samples, provided that protocols for sample preparation, DNA extraction, and metabarcoding analysis are carefully optimized.

Electronic supplementary material The online version of this article (doi:[10.1007/s10453-017-9497-z](https://doi.org/10.1007/s10453-017-9497-z)) contains supplementary material, which is available to authorized users.

K. Leontidou · C. Vernesi · J. De Groeve ·
F. Cristofolini · A. Cristofori (✉)

Department of Biodiversity and Molecular Ecology,
Research and Innovation Centre, Fondazione Edmund
Mach, Via E. Mach 1, 38010 San Michele all'Adige,
Trentino, Italy
e-mail: antonella.cristofori@iasma.it

K. Leontidou · D. Vokou
Department of Ecology, School of Biology, Aristotle
University of Thessaloniki, 54124 Thessaloniki, Greece

J. De Groeve
Department of Geography, Ghent University, Krijgslaan
281, 9000 Ghent, Belgium

Keywords Aerobiology · DNA extraction ·
Cloning · Sanger-sequencing · *trnL*

1 Introduction

Aerobiological surveys often reveal the airborne pollen spectrum of a catchment area. This information can be used to estimate plant biodiversity, to study the impact of climate change on pollen abundance and spatio-temporal distribution (Charalampopoulos et al. 2013; Jochnner et al. 2012) or to address public health

issues related to pollen (Cristofori et al. 2010; Damialis et al. 2007).

Airborne pollen surveys include volumetric or gravimetric methods, which collect pollen (and fungal spores) on different samplers (Levetin 2004). In the case of the commonly used Hirst-type volumetric sampler, pumped air directs pollen grains toward an adhesive surface, where they are trapped by impaction. Conversely, in the case of gravimetric sampling, pollen grains settle by gravity on a surface or into a sampling solution (e.g., Tauber traps). In the current aerobiological studies, the sampling medium is then analyzed by traditional light microscopy to identify pollen taxa and their abundance and to define a pollen spectrum for the catchment area of the sampler.

A well-known limitation of analysis by light microscopy is that only a small percentage of pollen grains are identifiable to species level, due to common morphological features shared within genera, families, and even orders. Analysis by light microscopy is also problematic in large-scale studies, since it requires considerable time and expertise for pollen identification and enumeration (Richardson et al. 2015).

Alternative methods for pollen identification have been investigated (e.g., spectroscopic techniques, optical-based detectors, and DNA-based methods (Dell'Anna et al. 2010; Oteros et al. 2015; Longhi et al. 2009). Among these, DNA-based methods appear particularly promising (Bell et al. 2016), thanks to the recent advent of Next Generation Sequencing (NGS) (Keller et al. 2015). DNA metabarcoding has been particularly promoted as a method for identifying multiple species from a single complex environmental sample (Taberlet et al. 2007). Metabarcoding targets a specific fragment of DNA (barcode), which is similar enough across taxa to be amplified using the same primers, yet enough variable to allow taxonomic identification. A metabarcoding approach commonly includes two important steps. First, a comprehensive reference database needs to be created from verified barcode sequences of the main taxa in the research area (Taberlet et al. 2007); second, sample processing and DNA extraction (i.e., cell lysis and DNA purification) need to be customized for the type of sample (Bell et al. 2016).

DNA-based airborne pollen analysis has already been applied to aerobiological research, using the classical volumetric Hirst-type samplers (Hirst 1952). Longhi et al. (2009) used real-time PCR techniques

and taxon-specific primers for pollen identification. They obtained promising results even with pollen mixtures and argued that a simultaneous identification of different taxa would reduce the time and effort of the analysis. More recently, the drawback of having to create a different pair of primers for each species was overcome by Kraaijeveld et al. (2015). These authors proposed metabarcoding in combination with NGS as a successful method not only to identify multiple pollen taxa simultaneously, but also to distinguish grass-allergen genera, and concluded that the method could easily be applied to other bioaerosols.

The aim of this study is to develop protocols for processing aerobiological samples for DNA extraction and metabarcoding analysis and to assess the efficacy of these protocols for the taxonomic assignment of airborne pollen. Using the classical morphological pollen analysis as a benchmark and building a custom reference database, we examined the conditions under which DNA metabarcoding (combined here with Sanger-sequencing) is applicable across a complexity of samples: from those containing only pure single-species pollen to complex aerobiological samples, collected by both gravimetric (Tauber) and volumetric (Hirst-type) air samplers (henceforth environmental samples).

2 Materials and methods

2.1 Study area

This study was conducted in Trentino (45°40'–46°30'N, 10°30'–12°00'E), a mountainous region of northern Italy, with a total surface area of 2607 km² and an elevation ranging from 65 to 3764 m above sea level. The region is characterized by diverse phyto-climatic types.

The Aerobiological Monitoring Centre of Fondazione Edmund Mach is situated in the valley of the river Adige, in San Michele all'Adige, 220 m above sea level. The aerobiological sampling site is in proximity to vineyards and apple orchards, which cover 44% of the surrounding area. A further 8% is covered by herbaceous weeds from the genera *Parietaria* (*P. diffusa*, *P. officinalis*), *Artemisia* (*A. absinthium*, *A. vulgaris*, *A. verlotorum*), and *Ambrosia* (*Ambrosia artemisiifolia*); at higher elevations, 48% of the territory is forested: 24% mixed (e.g., *Pinus*

sylvestris, *Ostrya carpinifolia*, *Fraxinus ornus*, *Quercus pubescens*), 11% coniferous (mainly *P. sylvestris*, *Pinus nigra* and *Picea excelsa*), and 8% broad-leaved (e.g., *Fagus sylvatica*, *F. ornus*, *O. carpinifolia*, *Q. pubescens*); the remaining 5% consists of semi-natural areas, transitional woodland, shrubs, and sparsely vegetated areas (Cristofori et al. 2010).

2.2 Collection, processing, and microscopic analysis of environmental samples

Environmental samples were collected at the Aerobiological Monitoring Centre of Fondazione Edmund Mach using volumetric and gravimetric air samplers.

Volumetric samples corresponding to 1 week (February 23 to March 2, 2015) were collected using two Hirst-type traps (Hirst 1952): Lanzoni (VPPS 2000, Lanzoni, Bologna, Italy) and Burkard (Sporewatch Electronic Spore & Pollen Sampler, Burkard Manufacturing, Rickmansworth, Hertfordshire, England), set 10 m off the ground, next to each other. Pollen was collected on a silicone-coated adhesive tape (Melinex; DuPont Teijin Films Luxembourg, SA, Luxembourg City, Luxembourg). The Burkard tape was cut in two longitudinal halves which were analyzed as replicates by molecular techniques. The Lanzoni tape was analyzed by light microscopy.

A comparison of preliminary results from the two volumetric samplers, Lanzoni and Burkard, was performed over a 5-month period (June 10 to November 3, 2013) and showed a good correlation of daily pollen data (Spearman's $\rho = 0.91$; $p < 0.001$), with no significant differences in terms of pollen load (Mann–Whitney U test). We subsequently considered morphological pollen data from the Lanzoni tape as a reference to compare with metabarcoding data derived by the Burkard tape.

Gravimetric samples corresponding to 2 months (August 7 to October 14, 2015) were collected using Tauber traps (Tauber 1974), set two meters away from the volumetric samplers, with the aperture 0.5 m from the ground. The aperture was equipped with a collar, from which air flows into the trap, covered by a 5 mm mesh to prevent insects and/or other large particles entering the trap. Samples were collected in 700 ml of preservative solution (1:1:1 water, alcohol, glycerol, plus 2 g l⁻¹ phenol). Three gravimetric traps were set up concurrently; two were analyzed as replicates by

molecular techniques and the third by microscopic techniques.

Pollen pellets for molecular and microscopic analysis were retrieved using the following protocol:

- (a) Filtration (for Tauber trap samples only): Samples were pre-filtered through a 200- μ m-metal mesh sieve (Retsch, Haan, Germany) to remove large particles (e.g., small insects, plant remains), and pollen was collected on a 5 μ m mixed cellulose ester filter (47-mm diameter, Merck Millipore Ltd, Cork, Ireland), using a vacuum pump filtration system (Sigma-Aldrich, Milan, Italy). When samples were prepared for microscopic analysis, a counting marker (*Lycopodium* spore tablets, Batch 3862, Lund University, Lund, Sweden) was added before filtration. Filters were dried for 2–3 h at 65 °C and either immediately processed or stored at –20 °C until further analysis.
- (b) Pellet preparation (for Tauber trap filters and longitudinal halves of Burkard tape corresponding to 1 week): Samples were placed into 15-ml-glass tubes (with the tape samples first cut into smaller pieces) and 5 ml of acetone was added. The samples were vortexed for 2–3 min to dissolve the filters and elute the tapes and centrifuged for 3 min at 2300 rpm. The same vortexing and centrifuging conditions were applied for the subsequent steps. After centrifuging, the supernatant was discarded and 1 ml of acetic acid was added, vortexing and centrifuging steps followed. After discarding the supernatant, two or four washing steps were performed to clean the pellet using 2 ml of distilled water, plus one drop of ethanol to reduce the surface tension, vortexing and centrifuging steps followed. The supernatant was discarded and the pellet was re-suspended in 1 ml of distilled water, transferred into 2 ml tubes, and centrifuged for 3 min at 13,500 rpm.

For the two Tauber filters and the two longitudinal halves of the Burkard tape, the supernatant was discarded and the pellets were stored at –20 °C until molecular analysis. For the pellet obtained from the third Tauber filter, 0.5 ml of the supernatant was kept, two drops of glycerol were added, and the samples were stored at 4 °C until analysis by light microscopy.

A classical morphological pollen analysis was performed using an optical microscope (Leitz Diaplan, Ernst Leitz Wetzlar, Wetzlar, Germany) at 400× magnification for both sample types. For the Tauber sample, at the end of the pelleting procedure, a small aliquot of the pellet was transferred onto a microscope slide and colored with basic fuchsin. A minimum of 400 pollen grains were counted and identified at the microscope, and the total pollen was then calculated by taking into account the number of counted markers (Faegri and Iversen 1989). For the volumetric sampler (Lanzoni), daily segments of the tape were mounted on slides, colored with basic fuchsin, and four horizontal continuous sweeps were analyzed (14% of the total area), following the European standard (UNI CEN/TS 16868:2015), and the weekly sum was calculated. This method allowed a quantitative estimation of the pollen present in the samples (Table 1).

2.3 DNA extraction experiments

Experiments were designed to define the DNA extraction conditions yielding the highest quantity of

DNA from pollen samples. Three replicates were performed for DNA extraction from single-species pollen samples and two for environmental samples. DNA yield was measured by a Qubit 2.0 Fluorometer (Life Technologies, Thermo Fisher Scientific Inc.). The final elution volume was 100 µl. Two different kits were tested, DNeasy Plant mini kit (Qiagen, GmbH, Hilden, Germany) (henceforth Qiagen kit), which is based on silica gel spin column technology, and Nucleomag kit (Macherey–Nagel, Düren, Germany), which is based on magnetic bead technology, using the automated DNA extraction systems Qiacube (Qiagen) and Kingfisher (Thermo Fisher Scientific, Waltham, MA, USA), respectively. Samples for DNA extraction were of increasing complexity (see Table 1): (a) pure single-species pollen, (b) single-species pollen added on sampling tapes or in Tauber solutions, and (c) environmental samples.

(a) Pure single-species pollen: The best conditions for pollen disruption and DNA extraction kit were evaluated. The samples consisted each of 2.5 mg of *Corylus avellana* L., *Juniperus communis* L., or *Artemisia vulgaris* L. pure pollen that was collected

Table 1 Features of pollen samples used for microscopic and molecular analyses

Sample type	Sample description	Contributing taxa	Estimated no. of pollen grains	Sampling period	Complexity factors
Pure, single-species pollen	Collected directly from flowers and used for molecular analysis	<i>Corylus avellana</i>	260,000		None
		<i>Juniperus communis</i>	400,000		
		<i>Artemisia vulgaris</i>	460,000		
Single-species pollen on tape and in Tauber solution	Simulations of volumetric and gravimetric sampling methods used for molecular analysis	<i>J. communis</i>	400,000		Sampling medium
Environmental sample on tape	Volumetric samples collected by Burkard trap, used for molecular analysis and by Lanzoni trap, used for microscopic analysis	Pollen from several species	11,500	23/02/2015 to 02/03/2015	Sampling medium, degradation processes and occurrence of other bioparticles
Environmental sample in Tauber solution	Gravimetric samples collected by Tauber traps; two traps for molecular analysis and one for microscopic analysis	Pollen from several species	79,000	07/08/2015 to 14/10/2015	Sampling medium, degradation processes and occurrence of other bioparticles

directly from flowers and stored at 4 °C before analysis. This quantity corresponds to approximately 260,000–460,000 pollen grains, depending on the species, as estimated by a Fuchs-Rosenthal counting chamber (HBG Henneberg-Sander GmbH, Giesen, Germany). The quantity was chosen to reflect the effective quantity of pollen grains found in seasonal Tauber traps located within various NATURA 2000 habitat sites, as estimated in a preliminary study (15,000–360,000 pollen grains per Tauber trap, in spring, higher to lower altitudes) using *Lycopodium* as a counting marker.

Prior to DNA extraction, to improve cell lysis by mechanical disruption of cell walls, high-energy agitation with beads was applied with a Retsch MM200 mixer mill. The beating power of glass beads was compared to that of steel beads. In an additional test, we froze samples in liquid nitrogen for 30 s prior to bead beating, as suggested by the Qiagen handbook. For each sample, beads were added as follows: one steel bead with a diameter of 5 mm (Qiagen) or, alternatively, a mixture of 0.5 g of glass beads (0.3 g of 212–300 µm in diameter and 0.2 g of 425–600 µm in diameter) (Sigma-Aldrich). All samples were ground for 1 min at 30 Hz, in two steps. All possible combinations of the disruption method and the extraction kit variables were tested.

Microscope analysis of pollen suspensions was carried out before and after lysis under the conditions found optimal to validate the lysis efficiency of the method across all three taxa. For this, we prepared the pollen suspensions as suggested by Kraaijeveld et al. (2015) and then used a Fuchs-Rosenthal chamber to count ten microscope fields before and after lysis for each pollen sample.

Finally, we tested the optimal protocol on a range of starting concentrations of *C. avellana* pollen to verify its efficacy on low quantities of pollen. For this experiment, we added 25 mg of *C. avellana* in 10 ml of water, we estimated the concentration with a Fuchs-Rosenthal chamber, and then diluted at three different concentrations (1:10, 1:100, and 1:1000), corresponding, respectively, to 26,000, 2600, and 260 total pollen grains, for subsequent DNA extraction.

(b) Single-species pollen on tape or in Tauber solution: Here we simulated the conditions applied in the field for volumetric and gravimetric sampling of airborne pollen, and we tested different extraction conditions to establish the optimal ones. 2.5 mg of *J.*

communis pure pollen was distributed (1) on a section of aerobiological tape corresponding to 1 day (48 mm) or (2) in a Tauber trap aqueous solution. For the preparation of the tape and the Tauber solution samples and for their processing, we followed the protocol described above for environmental samples (Sect. 2.2), and DNA was extracted from the obtained pellet.

We also examined if the DNA yield was negatively influenced by chemicals used in the sample processing or by filter remains (in the case of Tauber samples). For this purpose, we applied two or four washing steps during sample processing, before DNA extraction with any of the kits and under the optimal disruption conditions.

(c) Environmental samples: We examined the efficiency of the optimal conditions identified for single-species pollen when applied to the complex pollen spectrum of environmental samples. These are characterized by several pollen species and other bioparticles and possibly suffer from DNA degradation processes, due to their prolonged exposure in the field (1 week for the tape samples and 2 months for the Tauber ones). For this experiment, two Tauber traps and one Burkard tape corresponding to 1 week were collected and processed as described above (Sect. 2.2), and DNA was extracted from the obtained pellet.

2.4 Pollen taxonomic identification

2.4.1 Local reference database

The workflow for the taxonomic identification of pollen samples is summarized in Figure S1. First, a local reference database was constructed. A bioinformatics search was performed in Genbank database (National Center for Biotechnology Information, www.ncbi.nlm.nih.gov/genbank) for a short fragment of the chloroplast *trnL* intron (*trnL* c-h barcode) of species known to occur in the study area, as described in ‘La nostra Flora,’ a reference book for the flora in Trentino (Dalla Fior 1985). In total, 1470 species were searched, belonging to 46 families. Using the R package ‘rentrez’ (Winter 2016; <https://CRAN.R-project.org/package=rentrez>), *trnL* sequences were downloaded from the database ‘Nucleotide’ of Genbank using species names as search terms (e.g., ‘*Abies alba trnL*,’ Fig. S1.1a). For each of the sequences, the

portion between the primers was retrieved and stored in a database including information on the family, genus, species, sequence identifier number in Genbank (gi), and other metadata. The resulting database was then filtered for incorrectly labeled sequences. For this, an exhaustive manual validation was done for synonyms of all related species, and duplicates were removed. For each family, sequences were then aligned using the 'Muscle' algorithm with the R package 'Bioconductor muscle' (Edgar 2004). Sequences that could not be aligned due to low quality were removed (Fig. S1.1b).

Relevant taxa with low or no availability of *trnL* sequences were identified, and new sequences were generated. Young leaf samples were collected from different individuals of taxonomically verified plants in San Michele all'Adige (Trentino) and stored at -20°C until DNA extraction. Total DNA was extracted from 100 mg of leaf tissue, following the manufacturer's protocol of Qiagen kit, and the whole *trnL* intron was amplified using the primer pair c-A49325 and d-B49863 [5'-CGAAATCGGTA-GACGCTACG-3' and 5'-GGGGATAGAGG-GACTTGAAC-3'] (Taberlet et al. 2007) produced by Sigma-Aldrich. The PCR mixture contained 1.25 U Hotmaster Taq polymerase (5 Prime GmbH, Hilden, Germany), 1x of Hotmaster Buffer, 1 mM of dNTPs, and 0.5 μM of each primer. The program was set as follows: 94 $^{\circ}\text{C}$ for 2 min, 30 cycles at 94 $^{\circ}\text{C}$ for 30 s, 55 $^{\circ}\text{C}$ for 15 s and 65 $^{\circ}\text{C}$ for 2 min, followed by 10 min at 65 $^{\circ}\text{C}$. Amplicons were Sanger-sequenced in a Genetic Analyzer 3130 (Applied Biosystems, Foster City, CA, USA), and the newly generated sequences were added to already available ones (Fig. S1.1c), thus forming the final local reference database. All sequences stored in the database, together with the related information (i.e., gi, family, genus, and species), were exported in a FASTA file which was used for the final taxonomic assignment (Fig. S1.1d).

2.4.2 PCR amplification and sequencing

Pollen DNA was amplified with c-A49325 and h-B49466 *trnL* primers [5'-CGAAATCGGTA-GACGCTACG-3' and 5'-CCATTGAGTCTCTG-CACCTATC-3'] (Taberlet et al. 2007) produced by Sigma-Aldrich. DNA amplification was carried out in a final volume of 50 μL using 2 μL of DNA as a

template. The amplification mixture contained 1.25 U GoTaq Polymerase (Promega, Madison, WI, USA), 1x of GoTaq Flexi Buffer, 3 mM of MgCl_2 , 0.2 mM of dNTPs, and 0.2 μM of each primer. All PCR amplifications were carried out on a Veriti 96 well thermal cycler (Applied Biosystems), with the following program: 2 min at 95 $^{\circ}\text{C}$ and 40 cycles of 15 s at 95 $^{\circ}\text{C}$, 15 s at 52 $^{\circ}\text{C}$ and 30 s at 72 $^{\circ}\text{C}$, followed by 5 min at 72 $^{\circ}\text{C}$. Band sizes and concentrations were checked on a gel electrophoresis using QIAxcel (Qiagen) and analyzed by QIAxcel ScreenGel Software.

For samples containing pollen from multiple species (environmental samples), cloning prior to sequencing was performed. First, PCR products were run on a 2% agarose gel, and bands of the adequate size were excised and purified with the commercial QIAquick gel extraction kit (Qiagen). Cloning was then performed, using the TOPOTA Cloning Kit, Dual Promoter, with pCRII-TOPO Vector and One Shot Mach1 T1 Phage-Resistant Chemically Competent *E. coli* (Invitrogen, Carlsbad, CA, USA), following the manufacturer's protocol. For all positive colonies, DNA amplification was carried out with the M13 forward and reverse primers [5'-GTAAAC-GACGGCCAG-3' and 5'-CAGGAAACAGCTAT-GAC-3'] (Sigma-Aldrich) in a final volume of 20 μL , using 1 μL of DNA as template. The PCR mixture contained 1.25 U Hotmaster Taq polymerase (5 Prime GmbH, Hilden, Germany), 1x of Hotmaster Buffer, 1 mM of dNTPs, and 0.5 μM of each primer. The program was set as follows: 94 $^{\circ}\text{C}$ for 2 min, 35 cycles at 94 $^{\circ}\text{C}$ for 30 s, 55 $^{\circ}\text{C}$ for 15 s and 65 $^{\circ}\text{C}$ for 2 min, followed by 10 min at 65 $^{\circ}\text{C}$. The results of this amplification were verified on QIAxcel software for presence/absence of amplification, size, and concentration of bands. For sequencing, we selected about 30 amplicons of different sizes to ensure a good coverage of the species present in the samples.

For all samples, we purified the PCR products using ExoSap-IT (Amersham Pharmacia Biotech, Uppsala, Sweden). The purified DNA was Sanger-sequenced using 3.2 μM of the M13 reverse primer for the cloned samples and 3.2 μM of *trnL* primers for the rest of the samples and a BrightDye Terminator 2.5X Premix Cycle sequencing mix (Nimagen, Nijmegen, Netherlands). All samples were loaded in a Genetic Analyzer 3130 (Applied Biosystems).

2.4.3 Taxonomic assignment

The taxonomic assignment was done with BLAST 2.4.0+ software (Camacho et al. 2009) by comparison of the new unknown sequences to our local reference database, which was used as a BLASTN database (Fig. S1.2a). After obtaining the unknown sequences, we used Sequencher version 5.4 sequence analysis software (Gene Codes Corporation, Ann Arbor, MI, USA) in order to assess the quality of the sequences and further process them. All good quality sequences (i.e., those with unambiguous base peaks in the electropherogram of Sequencher) were stored in a FASTA file (Fig. S1.2b). Hence, each of the unknown sequences was aligned to the sequences in the local reference database (Fig. S1.2c), and information about the alignment was requested (e.g., identity percentage and query coverage) to taxonomically assign the sequences (Fig. S1.2d). Sequences with more than 98% identity score and query coverage were assigned to the presumed taxa.

2.5 Statistic analysis

The DNA yield data were analyzed using R software (Ihaka and Gentleman 1996). A three-way ANOVA for all effects and interactions was applied on the DNA extraction data from pure single-species pollen, with independent variables being the kit, the freezing step, and the bead material, while the dependent variable the DNA yield. To compare results from pure single-species pollen samples and single-species tape and Tauber samples, we applied a one-way ANOVA, using the type of sample as the independent variable and the DNA yield as the dependent variable. A three-way ANOVA for factor interactions was also applied on data from single-species tape or Tauber samples with the independent variables being the kit, the type of sample and the number of washing steps, and the dependent variable the DNA yield.

3 Results

3.1 Selection of ‘optimal’ DNA extraction protocols

(a) Pure single-species pollen: The ‘optimal’ conditions were selected by averaging results of the three

species (*C. avellana*, *J. communis*, and *A. vulgaris*), as our aim was to work with complex mixtures of multiple species (environmental samples). Results of the statistical tests and of the obtained DNA yields are shown in Tables 2 and 3, respectively. The DNA yields were significantly different between extraction kits (ANOVA; $p < 0.05$) and bead materials (ANOVA; $p < 0.001$). In particular, the highest DNA yield was obtained with a steel bead disruption step and the ‘Nucleomag’ extraction kit. Adding a freezing step did not influence the yield. Finally, the ANOVA showed no significant interactions between the variables examined.

Pollen lysis took place effectively for all three species. Before lysis, the concentration was found to be 408 pollen grains μl^{-1} for *C. avellana*, 480 pollen grains μl^{-1} for *J. communis*, and 576 pollen grains μl^{-1} for *A. vulgaris*. After lysis, the concentrations were 24 pollen grains μl^{-1} (95% lysis), 0 pollen grains μl^{-1} (100% lysis), 56 pollen grains μl^{-1} (90% lysis), respectively, resulting in a >90% lysis for all taxa.

Finally, we found that different dilutions of pure *C. avellana* pollen (starting quantity: 256,000 pollen grains) did not affect the DNA yield significantly, even when pollen was very diluted. DNA yields are reported in Table 3.

(b) Single-species pollen on tape or in Tauber solution: When applying the same extraction protocol to pollen of *J. communis* from tape (as in volumetric sampling) or Tauber solution (as in gravimetric sampling), the DNA yield decreased significantly compared to pure pollen samples in both cases (ANOVA; $p < 0.05$) (Fig. 1a).

For both sampling methods (tape and Tauber trap), there was a significant increase in DNA yield when the Nucleomag kit was used and, in case of Tauber trap samples, DNA yield increased when the washing steps were doubled (ANOVA; $p < 0.001$); no significant difference of this type was found for the tape samples (Fig. 1b).

The ‘optimal’ conditions were defined as those including disruption of the pollen cell wall with steel beads and DNA extraction with the Nucleomag kit for both Tauber and tape samples; four washing steps during the sample processing are suggested for Tauber samples, while for tape samples two washing steps are enough.

(c) Environmental samples: The applied ‘optimal’ conditions resulted into lower DNA yield than that

Table 2 Variables examined to identify the ‘optimal’ extraction conditions

Protocols	Tested variables	Selected option (<i>p</i> value)	Sample type
DNA extraction	Freezing step: with or without	ns	Pure pollen of <i>Juniperus communis</i> , <i>Corylus avellana</i> , <i>Artemisia vulgaris</i>
	Bead material: glass or steel	Steel beads (<i>p</i> < 0.001)	
	Extraction kit: Qiagen or Nucleomag	Nucleomag kit (<i>p</i> < 0.05)	
Sample processing	Washing steps: two or four	four washing steps in Tauber solution (<i>p</i> < 0.001); ns for tape	Pure pollen of <i>J. communis</i> on tape and in Tauber solution

The selection criterion was a significant increase in DNA yield (*p* < 0.05); ns = no significant effect

Table 3 DNA yields with the ‘optimal’ extraction conditions (i.e., steel beads, Nucleomag kit, four washing steps for Tauber solution, two washing steps for tape) for each type of sample:

pure single-species pollen of three different taxa, pollen of *J. communis* on tape and in Tauber solution, environmental samples

Species	Sample type	Average DNA yield ($\pm$ sd) (ng μl^{-1})
<i>Artemisia vulgaris</i>	Pure pollen	2.66 $\pm$ 1.84
<i>Corylus avellana</i>	Pure pollen	0.31 $\pm$ 0.14
<i>C. avellana</i>	Pure pollen (dilution 1:10)	0.24 $\pm$ 0.2
<i>C. avellana</i>	Pure pollen (dilution 1:100)	0.14 $\pm$ 0.01
<i>C. avellana</i>	Pure pollen (dilution 1:1000)	0.14 $\pm$ 0.01
<i>Juniperus communis</i>	Pure pollen	18.56 $\pm$ 11.01
<i>J. communis</i>	Pure pollen in Tauber solution	8.8 $\pm$ 4.39
<i>J. communis</i>	Pure pollen on tape	3.1 $\pm$ 3.16
various	Environmental sample in Tauber solution	1.36 $\pm$ 0.21
various	Environmental sample on tape	0.2 $\pm$ 0.02

obtained from single-species pollen on tape or in Tauber solution (Table 3).

3.2 Taxonomic identification

After a bioinformatics search for the *trnL* c-h barcode, a total of 1188 sequences were retrieved, corresponding to 403 species of 198 genera and 46 families; 44 additional sequences, corresponding to 26 species, were also generated by us (Table S1).

PCR results observed in the electropherogram of QIAxcel always showed well-defined fragments, without primer dimers, indicating good quality of extracted and amplified DNA. No PCR inhibition was detected. For the environmental samples, 124 positive colonies were recovered from the tape and 117

colonies from the Tauber solution; 31 and 28 amplicons, respectively, were selected for sequencing. The amplified fragments were successfully sequenced by Sanger-sequencing for 99% of all samples (*n* = 155), with only one bad quality sequence, in terms of ambiguities in the base peaks, which was excluded from the analysis.

For pure single-species pollen, the extraction test resulted in a correct taxonomic assignment for 100% of the samples (*n* = 72, three species, three replicates, eight tests). *J. communis* tape or Tauber samples were correctly assigned taxonomically for 100% of the samples (*n* = 24, one species, two sample types, three replicates, four tests).

The taxonomic assignment of unknown sequences from environmental samples resulted in seven

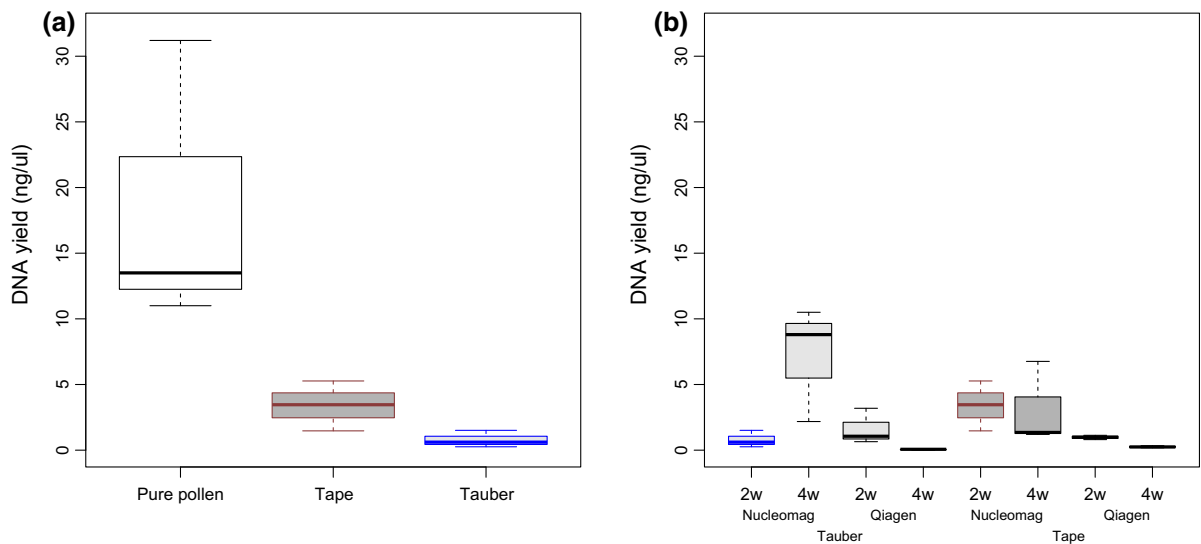


Fig. 1 DNA yield for the different types of *J. communis* samples. The boxplots give the 25th and 75th percentile (bottom and top of the box, respectively) and the median (black, brown, and blue thick line). The upper and lower whiskers correspond to the maximum and minimum obtained DNA yield values, respectively. Boxplot **a** provides DNA yields of the extraction for pure single-species pollen (white box), single-species pollen

from tape (gray box), and Tauber solution (light gray box) after two washing steps, Nucleomag kit and steel beads. Boxplot **b** provides DNA yield of single-species pollen from tape (gray box) and Tauber (light gray box) samples after two (2w) and four (4w) washing steps for two extraction kits (Nucleomag and Qiagen) and steel beads

different taxa for tape and three different taxa for Tauber samples. All sequences were identified with a 100% identity match on the whole sequence length (100% query coverage), except for *Urtica dioica* whose taxonomic identity was 99.2%, with a coverage of 99%. When compared to light microscope results (Table 4), 75 and 37% of the taxa identified by microscope were revealed by metabarcoding for the volumetric and the gravimetric sampling, respectively. *Cedrus* was dominant in the sequencing results of the Tauber environmental sample.

4 Discussion

In this study, we optimized a DNA extraction protocol that maximizes the DNA yield from pollen collected by gravimetric and volumetric samplers. With the exception of the cell lysis, the steps of this procedure can be easily automated (automatic DNA extraction instrument), making the procedure fast and cost-effective. The method worked properly for all tested pollen types, although the DNA yield was different across taxa in single-species samples. This is probably related to the species-specific starting quantity of

DNA. It is reported, for example, that conifers have a large, highly repetitive genome (De La Torre et al. 2014) and they are especially rich in cpDNA because of its exceptional paternal inheritance (Parducci et al. 2005). This could explain the higher DNA yield of the coniferous *J. communis* (compared to *A. vulgaris* and *C. avellana*). Cell wall structure also influences the DNA extraction (Kraaijeveld et al. 2015), as seen by the 100% lysis efficiency of *J. communis* pollen grains, characterized by a very thin exine layer, which leads to higher DNA yields than in case of *C. avellana* and *A. vulgaris* pollen grains, characterized by higher resistance of the cell wall.

We found that the total amount of DNA recovered in the Tauber solution became significantly higher with extra washing steps, while it does not seem to add any advantage on tape samples (test performed on *J. communis* pollen). This suggests that residues related to the medium of the gravimetric trap negatively affected the DNA extraction, probably due to their interactions with kit chemicals.

The DNA extraction conditions that were considered optimal after the exploratory experiments, when applied on environmental samples, resulted in lower DNA yield. This may be due to at least three reasons:

Table 4 Pollen taxa identified with the molecular and microscopic methods from samples subject to same environmental conditions

Environmental sample	Molecular analysis	Microscopic analysis
On tape	<i>Juniperus communis</i>	Cupressaceae-Taxaceae
	<i>Thuja orientalis</i>	Cupressaceae-Taxaceae
	<i>Ulmus</i> sp.	<i>Ulmus</i> sp.
	<i>Alnus</i> sp.	<i>Alnus</i> sp.
	Corylaceae	<i>Corylus</i> sp.
	Poaceae: <i>Lolium</i> sp., <i>Festuca</i> spp.	Poaceae
	Asteraceae: <i>Artemisia</i> spp., <i>Achillea</i> spp.	
	<i>Leucanthemum vulgare</i>	—
	<i>Tanacetum vulgare</i>	
	—	<i>Ostrya</i> sp.
	—	Urticaceae
In Tauber solution	—	<i>Fraxinus excelsior</i>
	<i>Urtica dioica</i>	Urticaceae
	<i>J. communis</i>	Cupressaceae-Taxaceae
	<i>Cedrus</i> sp.	Pinaceae
	—	Chenopodiaceae-Amaranthaceae
	—	Poaceae
	—	<i>Artemisia</i>
	—	<i>Plantago</i>
	—	Cannabaceae

(1) The concentration of pollen grains in the environmental samples was lower than in single-species samples, (2) environmental DNA degradation may have happened, and (3) other biological particles in the sample (e.g., fungi, bacteria) may have negatively affected the DNA extraction steps. For the sample preparation, we tried to minimize material loss selecting a filter that dissolves in acetone, which we used also to elute the tape. Finally, by applying the same conditions in highly diluted samples, we verified that the starting quantity is not a limiting factor for the efficiency of the protocol.

Sanger-sequencing on DNA extracted from single-species samples simulating field samples collected volumetrically and gravimetrically resulted in a correct taxonomic assignment on all samples of this type. For environmental samples, we adopted a semi-metabarcoding approach by using (1) universal primers for plants, which allowed to amplify multiple species at the same time without the need of taxon-specific amplification and (2) the traditional

molecular-based approach of cloning and sequencing of several amplicons, partially overcoming the limit of Sanger-sequencing, which can only sequence single-species, individually. We used the results of classical morphological identification for comparison of pollen grains from replicate environmental samples. In the case of the aerobiological tape samples, the *trnL* c-h barcode identified the majority of the taxa that were determined by morphological analysis, whereas for Tauber solution samples, it was less efficient.

For both Tauber and tape samples, our molecular method provided on average a higher taxonomic resolution than microscopy. For example, pollen identified under the microscope as Cupressaceae-Taxaceae type was assigned to *J. communis* and *Thuja orientalis* by molecular analysis. For big families, such as Poaceae and Asteraceae, the molecular method gave two and five assigned taxa, respectively. For instance, after morphological analysis, the Poaceae pollen was assigned to the genera *Lolium* and *Festuca*, by molecular analysis. Our results, even though based

on Sanger-sequencing, confirmed a resolution at the genus level for grasses, as showed by Kraaijeveld et al. (2015). In contrast, within the Corylaceae family, the *trnL* could not discriminate among *Ostrya carpinifolia*, *Carpinus betulus*, and *C. avellana*, whereas microscopy could assign pollen to the species level. For all taxa, the molecular method was able to assign sequences at least to the family level; this was not the case for the light microscope method which could not distinguish Chenopodiaceae from Amaranthaceae or Cupressaceae from Taxaceae.

The most effective approach to take advantage of the efficiency of DNA metabarcoding is by determining a large number of different sequences, by means of NGS. However, the successful outcome of any NGS approach is highly dependent on DNA quality and quantity, i.e., sufficient amounts of high-quality nucleic acids, which will be representative of the pollen cells in the sample, for subsequent library preparation (Bell et al. 2016). Therefore, we developed protocols for obtaining high-quality pollen DNA from environmental samples, without PCR inhibitors or sequence ambiguities after Sanger-sequencing. Although we used a suboptimal methodology for taxonomic assignment, i.e., Sanger-sequencing of cloned PCR products from each environmental sample, nonetheless, the final outcome is encouraging: Good DNA quality can be efficiently retrieved from a complex mixture of pollen grains and other bioparticles. Moreover, as expected, the molecular analysis allowed higher taxonomic resolution compared to analysis by light microscope.

The most critical issue is the low efficiency in the identification of pollen from gravimetric samples: Only 37% of the morphologically identified taxa were detected. The volumetric sampling has advantages over the gravimetric one regarding the DNA quality as the tape is exposed in the field only for a short period, 1 week, whereas the Tauber traps normally remain for longer, 2 months in our experiments. Moreover, the aqueous medium of the Tauber trap can lead to increased DNA degradation compared to the tape of the volumetric sampler, which is an inert medium where particles attach. The medium of the gravimetric sampler is expected to be more prone to oxidative reactions and microorganism proliferation resulting in higher rates of DNA degradation and accumulation of molecules different than those of pollen. Secondary metabolites, produced by degradation of biological

particles, may interact negatively with chemicals used in DNA extraction and amplification. Also, since the Tauber traps are closer to the ground level, it cannot be excluded that more particulate matter is collected in the medium. This is another factor potentially affecting DNA amplification. The non-proportional representation of taxa in the results from molecular analysis of Tauber samples could be attributed, at least to a certain degree, to the high presence of conifers in the area, *Cedrus* in particular, that are reported to have high amounts of cpDNA (Parducci et al. 2005), thus affecting the sequencing results.

Several lines of evidence suggest that a NGS-based metabarcoding approach could significantly increase the information from environmental samples in terms of number of identified taxa (Bell et al. 2016). Also, for a better taxonomic assignment in families, the use of an additional DNA barcode or a multi-locus metabarcoding approach may be very efficient (Keller et al. 2015; Richardson et al. 2015). Parducci et al. (2013) argued that metabarcoding and microscope-based methods can be considered as complementary in revealing different pollen taxa. Our results show that metabarcoding is a very useful method for taxonomic assignment of pollen grains retrieved from aerobiological samples, but also that currently it cannot entirely substitute classical microscopy techniques, particularly in the case of gravimetric samples. This is where more research is needed for attaining higher resolution of the molecular method. DNA metabarcoding of airborne pollen can therefore be efficiently used for practical applications, such as assessment of air biodiversity. The metabarcoding approach provides an added value to classical analysis, by validating the taxonomic assignment and improving its resolution. In case of large-scale biodiversity studies, it could be used as a rapid tool for the taxonomic identification of pollen. Provided that protocols are customized and well-defined, the use of more automated methods and the simultaneous handling of a large number of samples can significantly increase, both quantitatively and qualitatively our ability to analyze airborne pollen.

Acknowledgements This project is supported by FIRS >T (FEM International Research School). We thank Maria Cristina Viola for microscopic analyses and for helping with the taxonomic identification of plants. We thank Heidi Christine Hauße for comments that greatly improved the manuscript and other members of the Conservation Genetics group of FEM

Research and Innovation Centre (Matteo Girardi, Barbara Crestanello, Andrea Gandolfi, Alice Fietta) for providing support in the molecular techniques. Finally, we thank members of the Zooplant Lab group of Milano Bicocca University (Maurizio Casiraghi, Antonella Bruno, and Anna Sandionigi) and other colleagues (Camilla Capelli, Athanasios Charalampopoulos, and Duccio Rocchini) for their useful suggestions.

References

- Bell, K. L., Vere, N. De, Keller, A., Richardson, R. T., Gous, A., Burgess, K. S., et al. (2016). Pollen DNA barcoding: Current applications and future. *Genome*, 59(9), 1–12. doi:10.1139/gen-2015-0200.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10, 421. doi:10.1186/1471-2105-10-421.
- Charalampopoulos, A., Damialis, A., Tsiropidis, I., Mavromatis, T., Halley, J., & Vokou, D. (2013). Pollen production and circulation patterns along an elevation gradient in Mt Olympus (Greece) National Park. *Aerobiologia*, 29(4), 455–472. doi:10.1007/s10453-013-9296-0.
- Cristofori, A., Cristofolini, F., & Gottardini, E. (2010). Twenty years of aerobiological monitoring in Trentino (Italy): Assessment and evaluation of airborne pollen variability. *Aerobiologia*, 26(3), 253–261. doi:10.1007/s10453-010-9161-3.
- Dalla Fior, G. (1985). *La Nostra Flora, Guida alla conoscenza della flora della regione Trentino-Alto Adige* (3rd ed.). Trento: Casa Editrice G.D. Monauni.
- Damialis, A., Halley, J. M., Gioulekas, D., & Vokou, D. (2007). Long-term trends in atmospheric pollen levels in the city of Thessaloniki, Greece. *Atmospheric Environment*, 41(33), 7011–7021. doi:10.1016/j.atmosenv.2007.05.009.
- De La Torre, A. R., Birol, I., Bousquet, J., Ingvarsson, P. K., Jansson, S., Jones, S. J. M., et al. (2014). Insights into conifer giga-genomes. *Plant Physiology*, 166, 1724–1732. doi:10.1104/pp.114.248708.
- Dell'Anna, R., Cristofori, A., Gottardini, E., & Monti, F. (2010). A critical presentation of innovative techniques for automated pollen identification in aerobiological monitoring networks. In B. J. Kaiser (Ed.), *Pollen: Structure, types and effects* (pp. 273–288). New York: Nova Science.
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. doi:10.1093/nar/gkh340.
- Faegri, K., & Iversen, J. (1989). *Textbook of pollen analysis*. London: Wiley.
- Hirst, J. M. (1952). An automated volumetric spore trap. *Annals of Applied Biology*, 39(2), 257–265.
- Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314. doi:10.2307/1390807.
- Jochner, S., Ziello, C., Bock, A., Estrella, N., Buters, J., Weichenmeier, I., et al. (2012). Spatio-temporal investigation of flowering dates and pollen counts in the topographically complex Zugspitze area on the German–Austrian border. *Aerobiologia*, 28(4), 541–556. doi:10.1007/s10453-012-9255-1.
- Keller, A., Danner, N., Grimmer, G., Ankenbrand, M., Von Der Ohe, K., & Von Der Ohe, W. (2015). Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. *Plant Biology*, 17(2), 558–566. doi:10.1111/plb.12251.
- Kraaijeveld, K., de Weger, L. A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P. S., et al. (2015). Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Molecular Ecology Resources*, 15(1), 8–16. doi:10.1111/1755-0998.12288.
- Levetin, E. (2004). Methods for aeroallergen sampling. *Current Allergy and Asthma Reports*, 4(5), 376–383. doi:10.1007/s11882-004-0088-z.
- Longhi, S., Cristofori, A., Gatto, P., Cristofolini, F., Grando, M. S., & Gottardini, E. (2009). Biomolecular identification of allergenic pollen: A new perspective for aerobiological monitoring? *Annals of Allergy, Asthma & Immunology*, 103(6), 508–514. doi:10.1016/S1081-1206(10)60268-2.
- Oteros, J., Pulsch, G., Weichenmeier, I., Heimann, U., Möller, R., Röseler, S., et al. (2015). Automatic and online pollen monitoring. *International Archives of Allergy and Immunology*, 167(3), 158–166. doi:10.1159/000436968.
- 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. *Molecular Ecology*, 22(13), 3511–3524. doi:10.1111/mec.12298.
- Parducci, L., Suyama, Y., Lascoux, M., & Bennett, D. (2005). Ancient DNA from pollen: A genetic record of population history in Scots pine. *Molecular Ecology*, 14(9), 2873–2882. doi:10.1111/j.1365-294X.2005.02644.x.
- Richardson, R. T., Lin, C., Quijia, J. O., Riusech, N. S., Goodell, K., & Johnson, R. M. (2015). Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. *Applications in Plant Sciences*, 3(11), 1500043. doi:10.3732/apps.1500043.
- 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 Research*, 35(3), e14–e14.
- Tauber, H. (1974). A static non-overload pollen collector. *New Phytologist*, 73(2), 359–369.
- Winter, D. (2016). Rentrez: Entrez in R. R package version 1.0.2. <https://CRAN.R-project.org/package=rentrez>.