Experimental quantification of pollen with DNA metabarcoding using ITS1 and trnL

Sandra Baksay1*, André Pornon1, Monique Burrus1, Jérôme Mariette2, Christophe Andalo1 & Nathalie Escaravage1

Although the use of metabarcoding to identify taxa in DNA mixtures is widely approved, its reliability in quantifying taxon abundance is still the subject of debate. In this study we investigated the relationships between the amount of pollen grains in mock solutions and the abundance of high-throughput sequence reads and how the relationship was affected by the pollen counting methodology, the number of PCR cycles, the type of markers and plant species whose pollen grains have different characteristics. We found a significant positive relationship between the number of DNA sequences and the number of pollen grains in the mock solutions. However, better relationships were obtained with light microscopy as a pollen grain counting method compared with flow cytometry, with the chloroplastic trnL marker compared with ribosomal ITS1 and with 30 when compared with 25 or 35 PCR cycles. We provide a list of recommendations to improve pollen quantification.

Environmental DNA metabarcoding is a molecular method that consists of investigating environmental DNA samples made of complex mixtures of genomes from numerous organisms1. Due to new sequencing technologies and bioinformatics tools, metabarcoding has been increasingly used to identify taxa in environmental samples1 to monitor biodiversity2–4, to investigate ecosystem functioning5 and interaction networks6–8, in both aquatic and terrestrial ecosystems. Nevertheless, its reliability in quantitative approaches, which depend on the match between counts of high-throughput sequence reads and the amount of sampled biological material2, is still the subject of debate4–8. While taxon identification can reveal individual diet breadth11, species richness, and the composition of habitats2, communities12 and ecological networks4, taxon quantification provides knowledge on species evenness in those habitats, communities and diets or on the level of individual or species specialization in networks, all of which is very useful in ecological studies. Research on pollination and knowledge of the quantities of pollen transported by pollinators allow for the estimation of plant-pollinator interaction strength and hence it gives a more realistic representations of networks than those made possible using traditional approaches such as observing visits to plants by pollinators6–8.

Metabarcoding has been used in pollen studies to identify pollen in honey14–16, insect loads6–8,17, insect nests18, airborne samples19, and to quantify pollen abundance across various sample types. Several studies found significant positive relationships between pollen abundance (estimated using light microscopy) or pollen DNA quantities, and the abundance or the frequencies of high-throughput sequencing reads in experimental samples10,16,17,20–22, airborne samples22,23, insect pollen loads21,24–27 or in brood cells of solitary bees28. Conversely, other studies found low or no significant pollen-sequence abundance relationships when using ITS2 markers applied to pollen provision in bee corbicula containing huge amounts of pollen21–23.

Many factors associated with handling samples, technical processes, or the biological material itself can affect the accuracy of metabarcoding quantification. Some factors, related to pollen (size, structure), species DNA (gene copy number) and the characteristics of DNA markers (nuclear vs plastid), cannot be corrected whereas it may be possible to reduce the impact of others. Establishing relationships between pollen quantity and DNA sequence abundance requires accurate estimates of the concentration of pollen in the samples used for DNA extraction. However, pollen concentration measurements made using light microscopy, and pollen DNA isolation are usually performed on different subsamples which possibly result in differences in pollen concentrations as a result of pipetting artefacts. In addition to the possible artefacts that occur during PCR and preparing sequencing,
these may cause significant variation among replicates and artificially affect the relationship between the concentration of pollen and sequence abundance. An alternative approach which, to our knowledge, has not yet been tested, is counting pollen grains in the DNA isolation pellets. Since at this stage of experimental processing there is no longer a risk of DNA contamination in the laboratory, pollen counting could be automated and applied to the whole pollen population, rather than to a small fraction, as is the case in traditional methods using microscopy. However, post-PCR counting would preclude crushing pollen grains for DNA extraction, which, in any case, is not required for efficient DNA isolation. Efforts are also needed to reduce the accumulation of spurious sequences, chimeras and Taq polymerase inhibitors, during PCR and sequencing processes, which are expected to fluctuate with the genetic markers, the PCR and sequencing conditions, and the reagents used. Many studies which investigated artificially or natural pollen mixtures used samples with high pollen concentrations (from 30,000 to potentially more than 1,000,000 pollen grains) stored in a single bee corbicula. Such large quantities of pollen could release large amounts of DNA polymerase inhibitors and cause PCR dysfunction. It would thus be useful to design experiments with lower pollen amounts, for instance, in the range of those involved in pollination, typically from a few to several thousand grains. All these potential biases can be exponentially amplified during PCR, increasing the variability of the data set, reducing the strength of statistical tests and affecting the relationship between pollen concentrations and sequence abundances.

We investigated relationships between the number of pollen grains and sequence counts abundance obtained by high-throughput sequencing, using two species (Chrysanthemum sp. - Asteraceae, and Hippeastrum sp. - Amaryllidaceae) of which the pollen has very different characteristics (size, exin structure, DNA content). We analysed two plant specific loci, the internal transcribed spacer region 1 (ITS1, ≈ 300 bp) of the nuclear ribosomal region, and the plastidial P6-loop of the trnL (UAA) intron (≈75 bp), both of which are expected to have high sequence diversity. The loci were amplified using three different PCR conditions (25, 30 and 35 cycles). The number of pollen grains in each subsample was estimated before DNA isolation using light microscopy. Moreover, as we demonstrated in previous work, the mechanical pollen disruption did not give higher DNA yield, we used flow cytometry to count uncrushed pollen grains in the extraction pellets after enzymatic DNA isolation. Our main objectives were to investigate, for pollen quantiles in the range of insect pollen loads involved in plant pollination: (1) whether using flow cytometry on extracted solutions provides more accurate estimates of pollen grain abundance than light microscopy, and consequently, a stronger relationship between the number of pollen grains and sequence abundance; (2) how the relationship is affected by the number of PCR cycles, the type of markers and plant species with different pollen characteristics.

Results

Read scores of trnL and ITS1 amplifications. We obtained 2,699,831 trnL reads (360 samples), of which 56.68% were assigned to Hippeastrum sp. (HIP) and 43.32% to Chrysanthemum sp. (CHR). The total number of ITS1 reads (360 samples) reached 302,934, most of which (78.59%) were assigned to CHR.

The 25 PCR cycle amplification conditions produced relatively few trnL (on average, 98 and 19 reads for HIP and CHR, respectively) and ITS1 reads (on average, 14 and 169 sequences for HIP and CHR, respectively). Compared to 25 PCR cycles, the numbers of trnL sequences at 30 and 35 PCR cycles were 23 and 235 multiplied in HIP and 65 and 954 multiplied in CHR, respectively. Compared to 25 PCR cycles, the number of ITS1 sequences at 30 and 35 PCR cycles respectively, was multiplied by 24 and 56 in HIP and by 11 and 12 in CHR, respectively. Therefore, ITS1 amplification was lower than that of trnL and increased slightly with an increase in the number of PCR cycles.

Microscopy and cytometry pollen counts. We found a highly significant positive relationship between the numbers of pollen grains estimated by microscopy and flow cytometry (Table 1). However, the relationship was clearly better for HIP (higher R²; line slope close to 1) than for CHR. Flow cytometry tended to detect fewer CHR pollen grains than microscopy, especially in samples with low pollen abundance (Table 1).

Relationships between DNA sequence abundance and pollen grain abundance estimated using light microscopy. For both markers and regardless of the number of PCR cycles, ANCOVA linear model (lm) revealed highly significant positive relationships between the number of sequences and the number of pollen grains (Table 1; Fig. 2). However, the relationships (Fig. 2) were generally stronger (R² always ≥ 0.70) and less variable for trnL (CHR: 0.89 ≤ R² ≤ 0.94; HIP: 0.70 ≤ R² ≤ 0.76) than for ITS1 (R² generally ≤ 0.61; CHR: 0.38 ≤ R² ≤ 0.87; HIP: 0.36 ≤ R² ≤ 0.59). For trnL, the slope of the relationship ranged from 0.61 to 0.85, that is a 1 log increase in the number of pollen grains, a 0.61–0.85 log increase in sequence quantity (Fig. 1). For ITS1, the values were generally both lower and more variable (0.31 ≤ slope ≤ 0.87) across PCR conditions and species. The two plant species had very similar (trnL) or quite similar slopes (ITS1), except HIP at PCR35 for ITS1, which had a steeper slope than CHR.

There was also a significant species effect in each PCR cycle for ITS1 but only at PCR25 for trnL (Table 1). Namely, CHR had slightly fewer trnL sequences at PCR25 but many more ITS1 sequences than HIP in all PCR cycles (Fig. 2). The two plant species had very similar (trnL) or quite similar pollen-sequence relationship slopes (ITS1), except HIP at PCR35 for ITS1 which had a steeper slope than CHR. Overall, there was no pollen quantity–plant species interaction, except for ITS1 at PCR35.

Finally, while the number of trnL sequences tended to steadily increase with increasing PCR cycles (Fig. 2), the number of ITS1 hardly increased between PCR30 and PCR35. Thus, the amplification efficiency of ITS1 decreased with the number of PCR cycles, however, without fundamentally altering the relationships between pollen and sequence abundances.
Relationships between DNA sequence abundance and pollen grain abundance estimated using flow cytometry. The ANCOVA linear model highlighted very similar trends in the relationship as microscopy estimates (Table 1; Fig. 3) i.e.: (i) a strong significant effect of the number of pollen grains on the number of ITS1 and trnL sequences; (ii) a species effect, mostly for ITS1; (iii) generally no pollen grain x plant species interactions, and (iv) less variability in R² and slope for trnL than for ITS1. On the other hand, flow cytometry pollen estimates gave weaker predictions (lower R² and line slopes) of trnL sequence quantities, thus showing that flow cytometry was generally less efficient than microscopy in counting pollen grains.

**Figure 1.** Relationships between the number of pollen grains in *Hippeastrum* sp. (red lines) and in *Chrysanthemum* sp. (black lines) estimated by microscopy and by flow cytometry.

| Source of variation | PCR25 Sum sq | PCR25 Df | PCR25 F value | PCR30 Sum sq | PCR30 Df | PCR30 F value | PCR35 Sum sq | PCR35 Df | PCR35 F value |
|---------------------|--------------|----------|---------------|--------------|----------|---------------|--------------|----------|---------------|
| **Counting using microscopy** |              |          |               |              |          |               |              |          |               |
| **ITS1**            |              |          |               |              |          |               |              |          |               |
| Plant species       | 15.51        | 1        | 102.30***     | 16.19        | 1        | 48.00***      | 7.10         | 1        | 31.20***      |
| Pollen quantity (log) | 5.82         | 1        | 38.39***      | 8.22         | 1        | 24.37***      | 8.41         | 1        | 36.99***      |
| Plant species x Pollen quantity | 0.04        | 1        | 0.29 ns     | 0.60         | 1        | 1.79 ns       | 1.88         | 1        | 8.25**        |
| Residuals           | 5.30         | 35       | 0.34          | 36           |          | 8.19          | 36           |          |               |
| **trnL**            |              |          |               |              |          |               |              |          |               |
| Plant species       | 2.60         | 1        | 24.12***     | 20.20        | 1        | 2.07 ns       | 20.21        | 1        | 2.96 ns       |
| Pollen quantity (log) | 17.31        | 1        | 160.95***    | 16.60        | 1        | 168.57***     | 9.84         | 1        | 138.26***     |
| Plant species x Pollen quantity | 0.03        | 1        | 0.29 ns     | 0.01         | 1        | 0.06 ns       | 0.03         | 1        | 0.36 ns       |
| Residuals           | 3.87         | 36       | 3.55          | 36           |          | 2.56          | 36           |          |               |
| **Counting using flow cytometry** |              |          |               |              |          |               |              |          |               |
| **ITS1**            |              |          |               |              |          |               |              |          |               |
| Plant species       | 15.51        | 1        | 90.47***     | 16.19        | 1        | 39.10***      | 7.10         | 1        | 25.61***      |
| Pollen quantity (log) | 4.79         | 1        | 27.94***     | 5.26         | 1        | 12.71**       | 6.38         | 1        | 23.00***      |
| Plant species x Pollen quantity | 0.38        | 1        | 2.19 ns     | 0.80         | 1        | 1.94 ns       | 2.12         | 1        | 7.67**        |
| Residuals           | 6.00         | 35       | 14.91         | 36           |          | 9.98          | 36           |          |               |
| **trnL**            |              |          |               |              |          |               |              |          |               |
| Plant species       | 2.29         | 1        | 8.83**      | 0.18         | 1        | 0.67 ns       | 0.23         | 1        | 1.47 ns       |
| Pollen quantity (log) | 8.76         | 1        | 33.81***    | 8.27         | 1        | 31.28***      | 5.33         | 1        | 33.78***      |
| Plant species x Pollen quantity | 0.86        | 1        | 3.33 ns     | 0.95         | 1        | 3.59 ns       | 0.42         | 1        | 2.68 ns       |
| Residuals           | 7.77         | 30       | 7.93          | 30           |          | 4.73          | 30           |          |               |

**Table 1.** ANCOVA model of the effects of plant species (HIP or CHR) and pollen quantity (log transformed) counted using light microscopy or flow cytometry on the quantity of ITS1 and trnL reads obtained with 25, 30 or 35 PCR amplification conditions (PCR25; PCR30, PCR35 respectively). ns: not significant; **P < 0.01; ***P < 0.001.
Discussion

In this study, we investigated whether it is possible to predict the abundance of DNA sequencing reads from the quantities of pollen grains and if so, how accurately, and how the prediction is affected by the methodology used to estimate pollen abundances in mock solutions, PCR conditions, plant species and the type of markers (nuclear versus plastid). In agreement with previous studies20–22,25,40 but in contrast to others17,26,30,31, we found a significant, and often strong positive relationship between the number of DNA sequences and the number of pollen grains in the mock solutions. However, the strength of the relationship was influenced by the pollen counting methodology, the marker, the species and the number of PCR cycles.

Efficiency of microscopy vs flow cytometry in counting pollen grains. Usually, in experiments which aim to evaluate the potential of metabarcoding for pollen quantification, the estimates of pollen abundance in suspension and the DNA extractions have been performed using different subsamples17,20–22,25–27,30,31,40, resulting in unavoidable variation in pollen concentrations. Moreover, since counting pollen grains under a light microscope is very time consuming and as mock solutions often have high concentrations of pollen, only small subsamples and a small proportion of microscope slides (and consequently of pollen population) are usually inspected23. Such methodological problems may then be further amplified by PCR, thus blurring the relationship between sequence and pollen abundance. In an attempt to reduce these biases, in every pollen sample based on microscopy counting, we counted pollen after DNA isolation using flow cytometry. While DNA isolation and flow cytometry were performed on the same mock solution and since more than 80% of the stock solution was counted, we expected flow cytometry to provide the best estimates of pollen grain quantity, and therefore, the best sequence-pollen grain abundance relationships. Unexpectedly, stronger relationships were found using light microscopy, showing that it provided better estimates of pollen grain quantity, and therefore, the best sequence-pollen grain abundance relationships. Unexplainedly, stronger relationships were found using light microscopy, showing that it provided better estimates of pollen abundance than flow cytometry. Moreover, the efficiency of flow cytometry appeared highly species-dependant. Indeed, both microscopy and flow cytometry provided similar estimates of HIP pollen grain abundance, but flow cytometry largely underestimated the abundance of the smaller CHR pollen grains, especially in the less concentrated mock solutions (50–250 grains). We observed that, despite the fact that pollen grains were not crushed before DNA extraction many CHR grains were in fact fragmented after DNA extraction. Fragments were then likely confused in spectograms with the uncounted myriad of tiny biological and mineral particles present in the mock solutions. On the other hand, even though partially destroyed, the bigger pollen grains of HIP could still have been detected and counted. Moreover, due to the absence of lacunae and only one aperture, HIP pollen grains were possibly less sensitive to fragmentation than CHR pollen grains.

Figure 2. Relationship between the number of \textit{trnL} and ITS1 reads and the number of pollen grains (log transformed) in \textit{Hippeastrum} sp. (red lines) and in \textit{Chrysanthemum} sp. (black lines) estimated by light microscopy in three different PCR cycles.
Effects of molecular markers and PCR conditions. We found that for trnL, very accurate prediction of DNA read abundance was obtained from pollen grain quantities estimated by light microscopy, with determination coefficients and slopes generally higher than those found in the literature with the same markers or with other markers. Moreover, except at PCR25, the good relationship was conserved across species and PCR cycles and trnL sequence abundance increased steadily with increasing PCR cycles. These positive results confirmed that: (i) light microscopy counting of pollen grains after DNA extraction was a suitable method to prepare mock solutions and replicates, (ii) trnL amplification was not subject to important PCR biases, and (iii) the degree of repeatability of high-throughput sequencing was high.

The relationship between pollen–DNA sequence counts and ITS1 was also highly and positively significant, in the same range as that obtained by other authors but with less accuracy (lower determination coefficients and slopes) and higher variability across species and PCR conditions than with trnL. Although HIP has higher pollen DNA content than CHR, it produced fewer ITS1 sequences than CHR, whereas both species produced similar quantities of trnL sequences regardless of the quantity of pollen and PCR conditions. Furthermore, in contrast to what we observed for trnL, ITS1 abundance did not increase steadily with the number of PCR cycles and remained almost stable between PCR30 and PCR35, although with a modification in the relationship between the pollen and sequence abundances. These results suggest that: (i) the amplification efficiency and copies of trnL in pollen from the two species is comparable despite the fact that the number of plastids may vary across species; (ii) HIP had fewer ITS1 copies in its nuclear genome, lower amplification or sequencing efficiency than CHR; (iii) the processes involved in these alternative causes would have been primarily influenced by DNA characteristics rather than by DNA quantities per se.

The observed differences between species, markers and PCR conditions could be related to the GC content in trnL and ITS1 sequences. Indeed, high GC contents can inhibit amplification due to primer mis-annealing, Taq DNA polymerase errors, gene mis-amplification and the synthesis of chimeras. High GC-richness can also alter sequence reading during sequencing. Analysing rDNA genes, Pinto & Raskin (2012) found that sequence counts were inversely proportional to the GC content of the samples. Consistently, we observed that CHR had 60% GC in ITS1 and only 30% in trnL and that HIP had 10% more GC in ITS1 but only 1% more GC in trnL than CHR. Furthermore, amplification and sequencing using Illumina technologies are more effective for short DNA fragments and cause less variability in amplicon length than for longer fragments. Together, the difference in GC-richness between markers and species and in sequence length probably explained the lower yield of ITS1 sequences compared to yields of trnL sequences in general and for HIP compared to CHR in particular, while both species had almost equal quantities of trnL sequences. Moreover, the accumulation of GC-induced inhibitors with PCR cycles is likely responsible for the fact that the number of ITS1 sequences did not increase beyond PCR30, a trend that was not observed in trnL.
Recommendations and methodological considerations when using metabarcoding for pollen quantification. Despite its high potential, we know only three other studies that (successfully) used trnL in pollen quantification\(^{(22,40)}\). Therefore, our findings call for further studies to determine whether trnL can be applied to many other plant species and routinely used to take greater account of interaction strength in polli

ation network studies. Kraaijveld et al., (2015) noted that trnL also had higher discriminatory power than other standard markers such as ITS, rbcL and matK. It is worth noting that, probably due to their relatively low GC content, plastid loci such as trnL\(^{(20,22,40)}\) (≈35% GC content), rbcL\(^{(20,40)}\) (≈42% GC content), matK (≈35% GC content) or trnT\(^{(23,40)}\) (≈25% GC content) may provide a better estimation of pollen number than nuclear markers (ITS1 and ITS2 ≈ 60% GC content). However, the drawbacks of ITS could to some extent be alleviated\(^{(49)}\) by using high-fidelity DNA polymerases such as Phusion High-Fidelity DNA polymerase\(^{(25,48)}\) or Herculase II Fusion DNA polymerase (the present study), 3% (our study) or 5% DMSO\(^{(50)}\) and by applying low primer annealing temperatures\(^{(45)}\). The detrimental impact of the accumulation of polymerase inhibitors could be reduced by increasing the number of PCR steps (3 to 5 successive PCRs) and by diluting a subsample of the previously obtained amplicons in fresh reagent mixtures at each step\(^{(51)}\). However, this method would be costly and time-consuming and not practical for high input studies.

Some experimental studies\(^{(20,31)}\) or studies whose goal was to understand patterns of floral resource use in bees analysed huge quantities of pollen\(^{(41,49)}\) sometimes corresponding to several pollen pellets. The resulting high DNA amounts could have led polymerase inhibitors to accumulate in solutions with increasingly negative consequences for PCR amplification and thus explain the low correlation or the lack of a correlation between sequence counts and pollen quantities reported in some studies\(^{(21,26,30,31)}\).

With 30 PC cycles, both trnL and ITS1 markers provided the highest relationship slope and the best sequence-to-pollen prediction (trnL) while the effects of plant species either alone (PCR25/trnL) or in interaction with pollen amounts (PCR35/ITS1) were minimised. Moreover, for the low-pollen amount samples, PCR25 had the disadvantage of producing few sequences, which could be confused with contamination and be removed from the data set by protocols designed to remove eDNA contamination (airborne pollen or non-pollen plant tissues deposited on insect bodies). A dual-indexing strategy (dual-tagging PCR amplification and single-run sequencing) has sometimes been used to identify pollen by metabarcoding\(^{(21,28,30,40)}\), but this approach is not recommended for pollen quantification\(^{(28)}\), especially if the single-run sequencing involves multi-locus analysis. In practice, better quantification results have often been obtained with single-indexing PCR amplification and single-locus sequencing\(^{(20,22,23,26,27,48)}\).

To sum up, to improve pollen quantification, we recommend that: (1) unless special care has been taken to prevent pollen break out, to avoid using flow cytometry, or trying to perform flow cytometric counting before DNA extraction; (2) in the case of very high pollen abundances, as typical in bee corbicula, to either dilute samples to obtain solutions with a thousand pollen grains or to use an alternative method such as PCR-free genome-skimming\(^{(31)}\); (3) to use a multi-locus approach including short plastid markers with low GC content, and (4) to use high-fidelity Taq DNA polymerases (such as Phusion High-Fidelity or Herculase II Fusion DNA polymerases), 3% or 5% DMSO, and a low primer annealing temperature to reduce some of ITS marker weaknesses; (5) to apply, at least for trnL and ITS1, a 30 cycle PCR; (6) to use single-locus sequence analysis rather than multi-locus sequencing; and finally; (7) to normalize the amplicon concentration before sequencing in order to reduce sequencing artefacts due to very different DNA amplicon concentrations\(^{(27,52)}\).

While we only investigated two plant species, additional studies are required to determine whether our findings can be generalised to many other species. Moreover, we did not consider species mixtures. Some studies did not detect any pollen-mixture effect\(^{(32)}\) but others did. For instance, Bell et al., (2018) and Richardson et al., (2015) found under- or over-representation of some species in sequencing products compared to pollen populations estimated using light microscopy. In studies using plant mixtures, an effect of species identity on the sequence-pollen DNA abundance correlation has been observed, but without deleting the sequence-pollen counts correlation\(^{(50)}\).

Furthermore, using ITS1 and trnL, Ponnor et al.\(^{(20)}\) found highly significant positive correlations between the number of insect visits to plant species (more than 23 species in mixtures) and the number of their sequences in sequencing products. This occurred even though visits to flowers do not usually accurately reflect pollen transport by insects\(^{(53)}\). So, we believe our findings are not an exception but that the question of pollen quantification in mixtures deserves further investigations that take all the recommendations we suggest here into account.

Material and methods

Plant models. We investigated variation in the number of reads produced by 50, 250, 1,200, 6,000 pollen grains belonging to the ornamental cultivars, *Hippeastrum* sp. (HIP, Amaryllidaceae) and *Chrysanthemum* sp. (CHR, Asteraceae). We were unable to identify the species to which the plants belonged, likely because they were the products of many species and cultivar crossings (*H. striatum*, *H. mandonii*, *H. papilio* and *D. dichromum*, *C. indicum*, *C. chamaeleontium*). The plants were chosen because they are widely sold in gardening stores, they produce large amounts of pollen (for more details, see a previous study\(^{(23)}\) in winter when native species are not in flower and because their pollen has strikingly different characteristics. CHR has small (20 μm to 35 μm diameter), echinate pollen grains with a tectate ectexine with large lacunae and three apertures\(^{(48)}\), and a 2C value of 11.87 pg DNA\(^{(46)}\). HIP pollen grains are bigger (50 μm to 100 μm in diameter), with a semitectate ectexine and one aperture\(^{(35)}\). Their DNA content ranges between 2C = 13.35 pg and 2C = 17.09 pg\(^{(42)}\). These differences between and within species in pollen grain structure and DNA content may potentially influence the amount of DNA extracted and the number of sequencing reads produced by each pollen grain.

Pre- and post-extraction pollen counts. Pollen stock solutions were obtained for each species by vigorously shaking fresh stamens in 10 mL sterilized tubes, kept sealed at 4 °C, then adding 3 mL of lysis buffer CF solution (Nucleospin Food Kit, Macherey–Nagel) to each tube and mixing thoroughly. The concentration of

---

**References**

1. Bell, E., Richardson, D. & Schrader, M. (2018). *PLOS ONE*. 13(2), e0193990. 10.1371/journal.pone.0193990

2. Ponnor, S., Pöyhönen, J., Karjalainen, J. & Vuorinen, J. (2015). *Journal of Ecology*. 103(2), 248-257. 10.1111/1365-2745.12301

3. Richardson, D. J., Bell, E. J., Kallio, J. M. & Schrader, M. (2013). *Journal of Ecology*. 101(3), 762-774. 10.1111/j.1365-2745.2013.01666.x

4. Schrader, M., Bell, E. J., Kallio, J. M., Ponnor, S. & Richardson, D. J. (2011). *Proceedings of the National Academy of Sciences of the United States of America*. 108(42), 17571-17576. 10.1073/pnas.1107637108
pollen grains in each stock suspension was estimated (pre-extraction estimations) by counting pollen grains in 10 µl (HIP) or 2 µl subsamples (CHR, due to the higher pollen abundance) under a light microscope, with seven replicates per stock suspension. Based on the known pollen concentration of the stock suspensions, we took subsamples of each species containing 50, 250, 1,200, 6,000 pollen grains with five replicates (giving a total of 40 mock solutions for the two species). After DNA extraction and isolation, we performed an additional pollen grain count (post-extraction estimations) of each mock solution using flow cytometry (CyFlowSpace Sysmex). As the samples with ITS1 and trnL are not the same, two batches of 40 pellets were counted for the two markers, containing respectively the two types and the four quantities of pollen. To prevent rapid pollen sedimentation, extraction pellets with empty pollen grains were re-suspended in 1 ml of 60% glycerol before counting using flow cytometry.

**DNA extraction, PCR amplification and sequencing.** Total DNA extraction of mock solutions was performed with the DNeasy Plant Mini kit (Qiagen) according to the protocol of Pornon et al.20. trnL and ITS1 markers were then amplified at 25, 30, 35 PCR cycles with three replicates per PCR cycle condition (giving a total of 720 PCR products, corresponding to five extraction samples for each quantity of pollen x 4 pollen quantities x 3 PCR conditions x 3 PCR replicates x 2 plant species x 2 markers). For both trnL and ITS1 amplification, PCR reactions were performed in a 25 µl reaction volume containing 5 µl 5x Herculase II reaction Buffer, 25 mM each dNTP, 0.4 µM of each tagged PCR primer, 0.25 µl Herculase II fusion DNA polymerase, 2 µl DNA and distilled water. After reviewing the PCR protocol from Pornon et al.20 and optimizing the annealing temperatures at 55 °C and 50 °C for trnL and ITS1 respectively, the PCR programs were: 2 min denaturation at 95 °C; followed by 25, 30 or 35 cycles (20 s denaturation at 95 °C, 20 s annealing at 55 °C (for ITS1), 30 s elongation at 72 °C and a final elongation at 72 °C for 3 min. For ITS1, 3% DMSO was added in the reaction solution to increase Taq polymerase specificity20. PCRs were performed in the Thermal Cycler GeneAmp PCR System 9700 (Applied Biosystems) and each PCR product was visualized on 1% agarose in TAE 0.5X buffer and quantified on the QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems). Two libraries (one for each marker) were generated using 2 µl of each PCR product following the manufacturer’s guidelines for the Illumina TruSeq Nano kit, except that no sonication was performed. Libraries were sequenced using the MiSeq Illumina technology, 2 × 250 paired-end run for ITS1 and 2 × 75 paired-end run for P6-loop of trnL, using the NGS core facility at the Génopole Toulouse Midi-Pyrénées (www.get.genotoul.fr).

**Sequence analysis and identification of plant taxa.** We built our own barcoding reference library using barcodes from Pornon et al.20 and completed with barcodes of all species belonging to the Chrysanthemum and Hippeastrum genera obtained from the EMBL database, using the ecoPCR function of the OBITOOL package54 and following the OBITOOL pipelines (git.metabarcoding.org/obitools/ecopcr/). Sequence treatment followed a step-by-step analysis procedure of the OBITOOL protocol. Paired-end reads were assembled using the Illumina paired-end utility that aligns the two reads and returns the reconstructed sequences. Sequences of low alignment quality (< 40%) were discarded. Each sequence was affiliated to its corresponding sample using the ngsfilter command and dereplicated into unique sequences using the obuniq command. As some of these sequences may contain PCR and/or sequencing errors, as well as chimeras, they have to be discarded using the obigrep command while keeping sequences more than 20 bp long and with a count equal to or greater than 10 sequences. As a final denoising step, we kept sequences with no variants, with a count greater than 5% of their own count (command obiclean). Then, a single taxon was assigned to each sequence using the ecoTag program that compared the sequences produced to our taxonomic reference library. When assigning our two target plant species, a best match score >95% was allowed for each marker.

**Statistical analysis.** To compare the two methods of estimating the number of pollen grains in HIP and CHR, we fitted a linear regression model to the number of pollen grains estimated by flow cytometry using the number of pollen grains estimated by microscopy as predictive variable. To analyse the relationship between the number of pollen grains (estimated by microscopy and by flow cytometry, respectively) and the number of sequence counts, we fitted an ANCOVA linear model with the lm function of the R base package55. Independent statistical analyses were performed for each combination of markers (ITS1 or trnL) and PCR conditions (25, 30 or 35 cycles). We tested the fixed effects of the logarithm of the pollen quantity, plant species (HIP and CHR) and their interaction on the mean number of sequence reads averaged over the three PCR replicates. Different analyses were performed for either pre-extraction or post-extraction estimates of pollen grain quantities. When the interaction between plant species and pollen amount was not significant with respect to sequence abundance, it was removed from the model and simple effects of either plant species or pollen quantity were tested (type II sum interaction between plant species and pollen amount was not significant with respect to sequence abundance, it was removed from the model and simple effects of either plant species or pollen quantity were tested (type II sum interaction between plant species and pollen amount). As the samples with ITS1 and trnL are not the same, two batches of 40 pellets were counted for the two markers, containing respectively the two types and the four quantities of pollen. To prevent rapid pollen sedimentation, extraction pellets with empty pollen grains were re-suspended in 1 ml of 60% glycerol before counting using flow cytometry.

**Data availability**

The data that support the findings of this study will be available in a public repository.

Received: 3 December 2019; Accepted: 18 February 2020;
Published online: 06 March 2020

**References**

1. Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L. H. Environmental DNA. *Mol. Ecol.* **21**, 1789–1793 (2012).
2. Bohmann, K. *et al.* Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* **29**, 358–367 (2014).
3. Deiner, K. *et al.* Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Mol. Ecol.* **26**, 5872–5895 (2017).
4. Evans, N. T. *et al.* Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Mol. Ecol. Resour.* **16**, 29–41 (2016).
32. Kainz, P. The PCR plateau phase – towards an understanding of its limitations. Environ. Toxicol. Chem. 33, 359–369 (2014).
33. Galliot, J.-N. et al. Investigating a flower-insect forager network in a mountain grassland community using pollen DNA barcoding. J. Insect Conserv. 21, 827–837 (2017).
34. Lucas, A. et al. Generalisation and specialisation in hoverfly (Syrphidae) grassland pollen transport networks revealed by DNA metabarcoding. J. Anim. Ecol. 87, 1008–1021 (2018).
35. Porzon, A., Andalo, C., Burrus, M. & Escaravage, N. DNA metabarcoding data unveils invisible pollination networks. Sci. Rep. 7, 1–11 (2017).
36. García, N., Meerow, A. W., Soltis, D. E. & Soltis, P. S. Testing Deep Reticulate Evolution in Amaryllidaceae Tribe Hippeastrae. Mol. Ecol. Resour. 18, 391–406 (2019).
37. Bruni, I. et al. A DNA barcoding approach to identify plant species in multiflower honey. Food Chem. 170, 308–315 (2015).
38. Meo, A. A. & Khan, M. A. Pollen morphology as an aid to the identification of Chrysanthemum species (Compositae-Anthemideae). Pak. J. Bot. 56, (2014).
39. Mamedov, T. G. A DNA Barcoding Approach to Characterize Pollen Collected by Honeybees. PLOS One 9, e109363 (2014).
40. Keller, A. et al. Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. Appl. Plant Sci. 3, 1–11 (2015).
41. Haas, B. J. et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 21, 494–504 (2011).
42. Bell, K. L. et al. Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. Mol. Ecol. Resour. 18, 431–455 (2018).
43. Bell, K. L., Loeffler, V. M. & Brosi, B. J. An rbcL reference library to aid in the identification of plant species mixtures by DNA metabarcoding. Appl. Plant Sci. 5, 1600110 (2017).
44. Kainz, P. The PCR plateau phase – towards an understanding of its limitations. Biochim. Biophys. Acta BBA - Gene Struct. Expr. 1494, 23–27 (2000).
45. Delmas, C. E. L., Fort, T. L. C., Escaravage, N. & Porzon, A. Pollen transfer in fragmented plant populations: insight from the pollen loads of pollinators and stigmas in a mass-flowering species. Envol. Ecol. 6, 5663–5673 (2016).
46. Gyan, K. Y. & Woodell, S. R. J. Analysis of Insect Pollen Loads and Pollination Efficiency of Some Common Insect Vectors of Four Species of Woody Rosaceae. Fuct. Ecol. 1, 269 (1987).
47. Sodmergen et al. Behavior of organelle nuclei (nucleoids) in generative and vegetative cells during maturation of pollen in Lilium longiflorum and Pelargonium zonale. Protoplasma 168, 73–81 (1992).
48. Mamedov, T. G. et al. A fundamental study of the PCR amplification of GC-rich DNA templates. Comput. Biol. Chem. 32, 452–457 (2008).
49. Acinas, S. G., Sarma-Rupavarm, R., Klepac-Ceraj, V. & Polz, M. F. PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. Appl. Environ. Microbiol. 71, 8966–8969 (2005).
50. Tilak, M.-K., Botero-Castro, F., Galtier, N. & Nabholz, B. Illumina Library Preparation for Sequencing the GC-Rich Fraction of Heterogeneous Genomic DNA. Genome Biol. Evol. 10, 616–622 (2018).
51. Head, S. R. et al. Library construction for next-generation sequencing: Overviews and challenges. BioTechniques 56, (2014).
48. Alan, Ş. et al. A new method to quantify atmospheric Poaceae pollen DNA based on the trnT-F cpDNA region. Turk. J. Biochem. 44, 248–253 (2019).
49. Hardjasa, A., Ling, M., Ma, K. & Yu, H. Investigating the Effects of DMSO on PCR Fidelity Using a Restriction Digest-Based Method. 14, 4 (2010).
50. Miranda, V. F. O., de Martins, V. G., Furlan, A. & Bacci, M. Jr. Plant or fungal sequences? An alternative optimized PCR protocol to avoid ITS (nrDNA) misamplification. Brz. Arch. Biol. Technol. 53, 141–152 (2010).
51. Lang, D., Tang, M., Hua, J. & Zhou, X. Genome-skimming provides accurate quantification for pollen mixtures. Mol. Ecol. Resour. 19, 1433–1446 (2019).
52. Saitoh, S. et al. A quantitative protocol for DNA metabarcoding of springtails (Collembola). Genome 59, 705–723 (2016).
53. Popic, T. J., Davila, Y. C. & Wardle, G. M. Evaluation of Common Methods for Sampling Invertebrate Pollinator Assemblages: Net Sampling Out-Perform Pan Traps. PLoS One 8, e66665 (2013).
54. Boyer, E. et al. obitools: a unix-inspired software package for DNA metabarcoding. Mol. Ecol. Resour. 16, 176–182 (2016).
55. R Core Team. R: a language and environment for statistical computing. https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing.

Acknowledgements
This work was supported by the French Laboratory of Excellence project “TULIP”, by the Centre National de la Recherche Scientifique-INEE (ERESA program: DJ/ST/IP/2013/D-112), the région Occitanie-Midi Pyrénées and Université Fédérale Toulouse. The authors thank Olivier Catrice (LIPM - INRA) for his assistance in the use of the flow cytometer, Dr. Julian Donald (EDB) and Sean Heighton (EDB) for their precious comments.

Author contributions
A.P. supervised the study. A.P., N.E., M.B. collected samples. S.B. and N.E. prepared the samples and performed DNA laboratory work. S.B. and J.M. treated genetic raw data. S.B. and C.A. analysed data. S.B. wrote the first draft of the manuscript, A.P. corrected the manuscript. All authors contributed to the revisions.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to S.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020