Received: 1 May 2018 | Revised: 11 July 2018 | Accepted: 19 July 2018

DOI: 10.1111/1755-0998.12948

RESOURCE ARTICLE

WILEY MOLECULAR ECOLOGY RESOURCES

Pollen metabarcoding as a tool for tracking long-distance insect migrations

Tomasz Suchan[1*] | Gerard Talavera[2,3*] | Llorenç Sáez[4] | Michał Ronikier[1] | Roger Vila[2]

[1]W. Szafer Institute of Botany, Polish Academy of Sciences, Kraków, Poland

[2]Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Spain

[3]Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts

[4]Systematics and Evolution of Vascular Plants, Associated Unit to CSIC, Unitat de Botànica, Facultat de Biociències, Universitat Autònoma de Barcelona, Bellaterra, Spain

Correspondence
Tomasz Suchan, W. Szafer Institute of Botany, Polish Academy of Sciences, Kraków, Poland.
Email: t.suchan@botany.pl

Funding information
AEI/FEDER, UE, Grant/Award Number: CGL2016-76322; MINECO programme Juan de la Cierva Incorporación, Grant/Award Number: IJCI-2016-29083; British Ecological Society, Grant/Award Number: LRB16/1015; Polish Academy of Sciences statutory fund; FP7, Grant/Award Number: 622716; National Geographic Society, Grant/Award Number: WW1-300R-18

Abstract

Insects account for a large portion of Earth's biodiversity and are key players for ecosystems, notably as pollinators. While insect migration is suspected to represent a natural phenomenon of major importance, remarkably little is known about it, except for a few flagship species. The reason for this situation is mainly due to technical limitations in the study of insect movement. Here, we propose using metabarcoding of pollen carried by insects as a method for tracking their migrations. We developed a flexible and simple protocol allowing efficient multiplexing and not requiring DNA extraction, one of the most time-consuming part of metabarcoding protocols, and apply this method to the study of the long-distance migration of the butterfly *Vanessa cardui*, an emerging model for insect migration. We collected 47 butterfly samples along the Mediterranean coast of Spain in spring and performed metabarcoding of pollen collected from their bodies to test for potential arrivals from the African continent. In total, we detected 157 plant species from 23 orders, most of which (82.8%) were insect-pollinated. Taxa present in Africa–Arabia represented 73.2% of our data set, and 19.1% were endemic to this region, strongly supporting the hypothesis that migratory butterflies colonize southern Europe from Africa in spring. Moreover, our data suggest that a northwards trans-Saharan migration in spring is plausible for early arrivals (February) into Europe, as shown by the presence of Saharan floristic elements. Our results demonstrate the possibility of regular insect-mediated transcontinental pollination, with potential implications for ecosystem functioning, agriculture and plant phylogeography.

KEYWORDS
insect migration, pollen metabarcoding, pollination, Sahara, *Vanessa cardui*

1 | INTRODUCTION

Insects undergo aerial long-distance migrations (Chapman, Reynolds, & Wilson, 2015; Holland, Wikelski, & Wilcove, 2006) that outnumber migrations of larger organisms, such as birds, both in abundance and biomass (Hu et al., 2016). These long-range movements have important—albeit still largely unknown—implications for ecosystems and human economy (Bauer & Hoye, 2014; Chapman et al., 2015; Hu et al., 2016). Nevertheless, mostly due to the technical challenges associated with tracking small organisms (Chapman et al., 2015), our knowledge on insect migrations is extremely limited (Holland et al., 2006), especially in comparison with that on vertebrate migrations.

Tracking long-distance insect migrations involves assessing the actual path of individuals, either by mark–recapture studies, using variety of markers (reviewed in Hagler & Jackson, 2001), or by

[†]Tomasz Suchan and Gerard Talavera should be considered joint first authors.

telemetry (Kissling, Pattemore, & Hagen, 2014; Osborne, Loxdale, & Woiwod, 2002). However, mark–recapture studies of migrating insects have low recapture success rate and are feasible only for the most emblematic species, such as monarch butterflies (Garland & Davis, 2002; Knight, Brower, & Williams, 1999). On the other hand, radio telemetry is only suitable for tracking the largest insects at short distances and is technically limited by relatively high weight of the transmitters and short battery life (Kissling et al., 2014). Because of that, so far only one telemetric study was performed to assess long-distance migrations in insects—of the dragonfly *Anax junius*. It however involved considerable logistic challenges, such as using small planes and ground teams for tracking insect movements (Wikelski et al., 2006).

Due to the above limitations, insect migration studies are traditionally observation-based, linking the place, time and, sometimes, direction of the observations with hypothetical routes (Howard & Davis, 2009; Talavera & Vila, 2016). More recently, other observational methods, such as vertical-looking radars (Chapman, Drake, & Reynolds, 2011) or Doppler weather radars (Chilson et al., 2012; Westbrook & Eyster, 2017), have been used in insect migration studies, often supplemented by aerial nets or other type of traps for accurate species identification (Chapman et al., 2002; Chapman, Reynolds, Smith, Smith, & Woiwod, 2004). The new developments in this technology allow insect body mass, shape, wing-beat frequency, flight altitude and heading directions to be measured, often allowing species determination (Chapman et al., 2011; Dean & Drake, 2005; Drake et al., 2017), but observations are usually constrained to particular areas because radars have no or limited mobility.

Other methods of studying migrations use intrinsic markers, such as genetic markers or isotope composition. By using genetic markers, populations connected by regular gene flow can be identified (Lowe & Allendorf, 2010), which may suggest migratory routes and natural barriers (Nagoshi, Meagher, & Hay-Roe, 2012). The utility of genetic markers is dependent on spatial genetic structuring. For migratory insects, structuring is expected between independent migratory circuits, but might not be maintained in the case of regular gene flow between migrating lineages (global panmixia) (Lukhtanov, Pazhenkova, & Novikova, 2016; Lyons et al., 2012). Because the stable isotope ratios, such as $^{2}H/^{1}H$ or $^{13}C/^{12}C$, of organic tissues are related to the site where insects developed, these can also be used, along with modelled geographic isotope patterns (isoscapes), to infer probabilistic natal origins of migrating individuals (Brattström, Bensch, Wassenaar, Hobson, & Åkesson, 2010; Hobson, Wassenaar, & Taylor, 1999; Stefanescu, Soto, Talavera, Vila, & Hobson, 2016; Talavera, Bataille, Benyamini, Gascoigne-Pees, & Vila, 2018). This technique does not rely on marking–recapturing specimens, and it is thus also suitable for small species (Hobson, 2008). However, geospatial assignments depend on limited isoscape resolution and are usually only helpful at inferring large-scale geographic patterns.

As insects visit and feed on flowers, the pollen is deposited on their bodies and can be transported across large distances (Ahmed, Compton, Butlin, & Gilmartin, 2009). Therefore, pollen of plants endemic to certain areas could also be exploited for tracking long-distance insect migrations (Hagler & Jackson, 2001; Jones & Jones, 2001), and, indeed, it has been used as a marker in a handful of studies (Gregg, 1993; Hendrix & Showers, 1992; Hendrix, Mueller, Phillips, & Davis, 1987; Lingren et al., 1993, 1994; Mikkola, 1971; Westbrook et al., 1997). However, conventional pollen identification by light or electron microscopy is time-consuming and requires specialized taxonomic knowledge. It is therefore difficult to apply as a widely accessible tool for large-scale studies (Galimberti et al., 2014; Keller et al., 2015; Richardson, Lin, Sponsler, et al., 2015; Sickel et al., 2015). Moreover, taxonomical classification of pollen grains is often unachievable to the species or even genus level (Hawkins et al., 2015; Kraaijeveld et al., 2015; Richardson, Lin, Sponsler, et al., 2015).

The development of the next-generation sequencing (NGS) technologies allowed straightforward sequencing of DNA barcodes from mixed environmental samples, termed "metabarcoding" (Deiner et al., 2017; Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012). In this study, we use a DNA metabarcoding approach to identify pollen grains carried by a long-distance migratory insect species—the painted lady butterfly *Vanessa cardui*. This is a virtually cosmopolitan species adapted to seasonally exploit a wide range of habitats and sometimes observed even at extreme latitudes or in the open ocean (Shields, 1992). The Palearctic-African migratory system involves populations that undergo yearly long-distance latitudinal migrations in a circuit between Tropical Africa (September to February) and the temperate zone (February to September) (Talavera & Vila, 2016). Such annual circuit might involve 10 generations or more, some of them performing long-distance movements of thousands of kilometres. The distances crossed by individual butterflies within a single generation are unclear. It has been recently shown, both by stable isotope and observational evidence, that the species massively immigrate and breed in autumn in sub-Saharan Africa and that these populations are of European origin (Stefanescu et al., 2016; Talavera & Vila, 2016). It has been also proved that these populations return during the next spring to the Mediterranean, thus colonizing Europe again and performing an annual multigenerational roundtrip (Talavera et al., 2018). These results depict a new spatiotemporal model for the migration of *V. cardui* in this part of the world, which involves migratory movements across differentiated floristic regions: central and northern Europe, the Mediterranean, the Sahara and Tropical Africa. Although the phenology of migratory movements within the western Palearctic is well known, both from observations and entomological radars (Stefanescu et al., 2013), the African locales and routes of the species from October to March are unclear (Talavera & Vila, 2016).

Using pollen metabarcoding from captured butterflies apparently migrating northwards into Europe, we test (a) for the presence of DNA from pollen grains deposited on insects' bodies after a long-distance migration from Africa and, if detected, (b) whether the sequences obtained are of African endemic plant species that could be informative on the migration pathways.

Despite great potential of pollen DNA metabarcoding for pollination biology or palynological studies, it has been used only in a

handful of cases, mostly to investigate honey composition (Bruni et al., 2015; Hawkins et al., 2015; Prosser & Hebert, 2017; Valentini, Miquel, & Taberlet, 2010), honeybee foraging (de Vere et al., 2017; Galimberti et al., 2014; Richardson, Lin, Quijia, et al., 2015; Richardson, Lin, Sponsler, et al., 2015) and, recently, plant–pollinator interactions (Bell et al., 2017; Keller et al., 2015; Lucas et al., 2018a, 2018b; Pornon et al., 2016; Sickel et al., 2015). Several DNA markers have been proposed for identifying mixed pollen loads from insects (Bell et al., 2016), the internal transcribed spacer 2 (ITS2) nuclear ribosomal fragment being one of the most frequently employed (Keller et al., 2015; Richardson, Lin, Sponsler, et al., 2015; Sickel et al., 2015). We chose ITS2 as it has several apparent advantages for pollen metabarcoding: (a) It is a nuclear marker and is thus present in pollen grains for all the plant taxa; (b) the sequence length is short enough for amplicon sequencing using Illumina MiSeq technology; (c) it is sufficiently informative to discriminate most plant species (Chen et al., 2010); and (d) the number of available reference sequences in databases is the highest among the available plant DNA barcodes (Bell et al., 2016). Nevertheless, plastid markers have also been successfully amplified from pollen grains and used as metabarcoding markers (Bell et al., 2017; Kraaijeveld et al., 2015; Lucas et al., 2018a, 2018b ; Richardson, Lin, Quijia, et al., 2015) but also show some family-specific biases (Richardson, Lin, Quijia, et al., 2015).

Several library preparation protocols have been proposed for pollen metabarcoding. Richardson, Lin, Sponsler, et al. (2015) performed ITS2 amplification using standard primers followed by purification and NGS library preparation using commercial kits. Keller et al. (2015) and Sickel et al. (2015) proposed an approach based on Kozich, Westcott, Baxter, Highlander, and Schloss (2013) protocol, where an ITS2 amplicon Illumina library is prepared within a single PCR step. These authors used ITS2 primers tailed with appropriate technical sequences and custom sequencing primers, to avoid losing sequencing cycles for ITS2 primer sequences and avoid problems related to low sequence diversity of these regions. The downside of such protocol is the need of replacing all the oligonucleotides when targeting other markers, template-specific biases when using indexed primers (O'Donnell, Kelly, Lowell, & Port, 2016), and the necessity to use custom sequencing primers.

We applied a two-step laboratory protocol: First, the ITS2 fragment was amplified using standard primers (Chen et al., 2010; White, Bruns, Lee, & Taylor, 1990), tailed with partial Illumina sequences; in the second step, the fragments were double-indexed and the final library produced—an approach similar to the Illumina 16S sequencing protocol (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf). Compared to the protocol by Sickel et al. (2015), our approach uses one additional PCR, but it is more flexible and cost-effective because the sequenced DNA barcode can be changed just by modifying the two primers used in the first PCR. Moreover, in the second PCR, we use primers compatible with standard Illumina sequencing protocol, which simplifies the sequencing step and allows sequencing the metabarcoding libraries along with other standard libraries. We also contribute a bioinformatic pipeline to analyse and classify the obtained reads.

2 | MATERIALS AND METHODS

2.1 | Sampling

We monitored seven sites along the Mediterranean coast of Spain to test for potential arrivals of *V. cardui* from the African continent (Figure 1, Supporting Information Table S2). Sampling was designed to capture specimens with high probability to be in migratory phase. To do that, we sampled sites where the species was unlikely to be found while nectaring or breeding. In particular, we sampled points in the Mediterranean shorelines, usually in the sandy beaches, cliffs

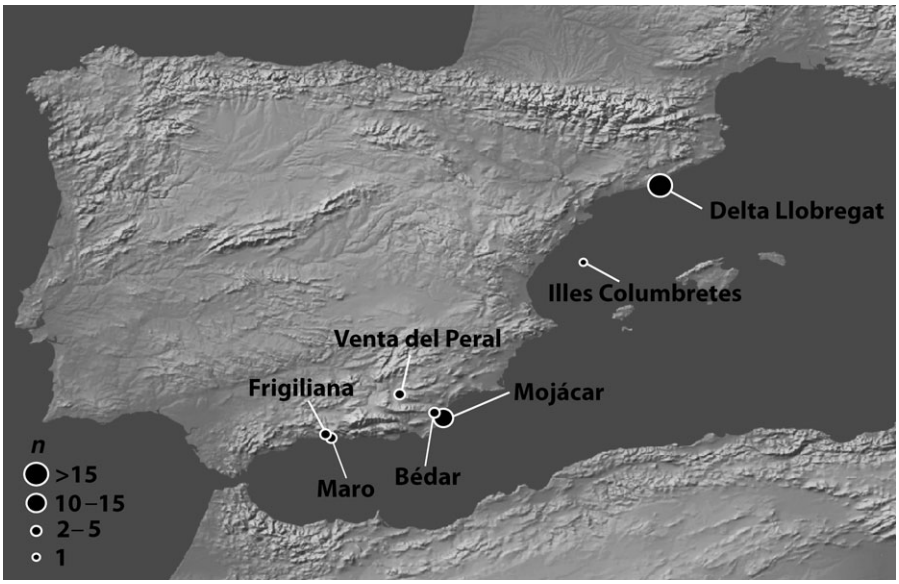


FIGURE 1 Map of the sampling site. The size of the points indicates the number of samples collected in each locality

or inspecting the vegetation nearby the coast. Timing was also chosen within the time frame where *V. cardui* arrivals are expected to colonize the Iberian Peninsula (February–April), and when consistent wind patterns or storms from Africa occurred, that could aid insect migrations. All the samples were immediately bagged in glassine envelopes that were sealed and stored at −20°C until pollen isolation and library preparation.

2.2 | NGS library construction

Pollen isolation and library construction were performed in four batches (see Supporting Information Table S2). We amplified the internal transcribed region 2 (ITS2) of the ribosomal DNA by using a combination of ITS-S2F (Chen et al., 2010) and ITS-4R (White et al., 1990) primers, tested in other pollen metabarcoding studies (Keller et al., 2015; Sickel et al., 2015). In the first step, the ITS2 fragment was amplified using the above primers tailed with technical sequences: six random nucleotides to increase sequence diversity during the first sequencing cycles and a part of the Illumina adapter. In the second reaction, we used modified Illumina TruSeq primers (Table 1) in order to index the samples and produce the final library. We used index sequences from TruSeq Amplicon series, that allow to uniquely tag up to 96 samples. These indexes can be supplemented by TruSeq CD index sequences (see https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/experiment-design/illumina-adapter-sequences-1000000002694-08.pdf) as in Sickel et al. (2015), for a total of 384 combinations. All the steps of the protocol, from pollen isolation up to the first PCR, were conducted under a laminar flow cabinet in a pre-PCR area in order to avoid external contamination. All the working spaces and equipment were cleaned with 10% bleach solution and 70% ethanol before and after work. We used only ultrapure nuclease-free water and a separate stock of reagents and plastics dedicated solely for metabarcoding work. Moreover, library preparation steps for all the samples and pollen isolation from most of the samples (batches 1–3) were conducted in a laboratory located far away from the sampling sites (at W. Szafer Institute of Botany, Polish Academy of Sciences, Kraków, Poland). Pollen isolation from batch 4 was performed at Institut de Biologia Evolutiva, Barcelona, Spain.

TABLE 1 Sequences of the primers used: ITS2-S2F and ITS2-4R primers were used in the first PCR; PCR_F_D50x and PCR_R_D7xx are the indexing primers used in the second PCR; XXX XXX XX are 8 nt-long index sequences

ITS2-S2F: ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT NNN NNN ATG CGA TAC TTG GTG TGA AT
ITS2-4R: GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TNN NNN NTC CTC CGC TTA TTG ATA TGC
PCR_F_D50x: AAT GAT ACG GCG ACC ACC GAG ATC TAC ACX XXX XXX XAC ACT CTT TCC CTA CAC GAC GC
PCR_R_D7xx: CAA GCA GAA GAC GGC ATA CGA GAT XXX XXX XXG TGA CTG GAG TTC AGA CGT GTG C

After preliminary trials (see Supporting Information Appendix S1), we decided not to extract DNA but to use Phire Plant Direct Polymerase (Thermo Fisher Scientific, Waltham, MT, USA) in the first PCR step, which successfully amplified ITS2 from pollen mix without DNA extraction. Pollen was collected by vortexing butterfly bodies (with wings removed) in a 2-ml tube with 50 µl of sterile water with 0.1% SDS, centrifugation of the obtained solution and drying it under vacuum. The obtained pellet was diluted in 15 µl of Phire Plant Direct sample buffer and homogenized with five zirconium beads on the TissueLyser II machine (Qiagen, Hilden, Germany) at 30 Hz for 1 min before proceeding to the PCR. We added blank sample to each of the four extraction batches ("extraction blank samples")—a 1.5-ml tube filled with 50 µl of water, left open during the whole extraction procedure and processed like a normal sample. We processed each sample in three independent PCRs to avoid reaction-specific biases (Fierer, Hamady, Lauber, & Knight, 2008; Sickel et al., 2015) with another four blank samples added at the PCR step ("PCR blank samples"). Each of the replicate PCRs consisted of 1 µl of the disrupted pollen sample, 25 µl of Phire Plant Direct Polymerase Mix, and 0.5 µM of each primer in 50 µl reaction volume with the following PCR programme: 98°C for 5 min, 20 cycles of denaturation at 98°C for 40 s, annealing at 49°C for 40 s and elongation at 72°C for 40 s, and final extension step at 72°C for 5 min. After the reaction, we combined the PCR triplicates and purified the product using 1× ratio of AMPure XP (Beckman Coulter, Indianapolis, IN, USA) and eluted in 10 µl of water. In the second reaction, we indexed each sample using a unique combination of primers. The second reaction, also performed in triplicates for each sample, consisted of 1 µl of the purified PCR product, 1× Q5 buffer, 0.2 U of Q5 Hot Start Polymerase, 0.5 µM of forward and 0.5 µM of reverse indexed primer in 10 µl reaction volume. We amplified the reaction using a PCR programme as follows: 30-s initial denaturation at 98°C; 12 cycles of denaturation at 98°C for 10 s and combined annealing and extension at 72°C for 30 s (shuttle PCR); and final extension at 72°C for 5 min. After the reaction, we combined the triplicates and verified the reaction success on TapeStation 4200 (Agilent, Santa Clara, CA, USA). We then pooled 10 µl of the PCR product from each sample and purified the pool using 1× ratio of AMPure XP. After that, the pooled library was quantified using the Qubit instrument (Thermo Fisher Scientific, Waltham, MA, USA) and sequenced with 15% PhiX spike-in on half of the lane of Illumina MiSeq (San Diego, CA, USA) using 600-cycle MiSeq Reagent Kit v3, according to manufacturer's instructions.

2.3 | Data analysis

We merged the raw paired-end reads using PEAR v0.9.8 (Zhang, Kobert, Flouri, & Stamatakis, 2013) and retained only the successfully merged reads. We then trimmed the primer sequences with the *remove_primers* script (www.biopieces.org). The obtained sequences were processed using VSEARCH v2.4.3 (Rognes, Flouri, Nichols, Quince, & Mahé, 2016): filtered by an expected error rate (*maxe* parameter) of 0.5, minimum length of 250 nt and maximum length of 450 nt,

and removing the reads with ambiguous nucleotides. Next, the reads were dereplicated and singletons removed. We classified the filtered reads with SINTAX (from the USEARCH package v10.0.240; Edgar, 2016) using the ITS2 database of Sickel et al. (2015), modified for SINTAX by parsing with a custom script. Only reads classified to the species level with a probability of 95% or above were further considered. The classified reads were then summarized by species with a minimum threshold of 100 reads per plant species per sample. We also summarized SINTAX classification probabilities by providing maximum classification probability per plant species (Supporting Information Table S3) and for each butterfly sample and plant species combination (Supporting Information Table S4). The raw sequences, the scripts used to process the data and ITS2 database are available online (see Data Availability).

2.4 | Plant distributions

Geographical distributions of all detected plants were compiled by checking specific literature and occurrences available in the online databases: GBIF (2018), the African Plant Database (2018) and Euro+Med (2006) (Supporting Information Table S1). The plants with the geographical range not including the butterfly sampling sites were further examined as potential marker candidates to unravel migratory paths. These mainly included Saharan-North African endemics and African species with ranges reaching southern Mediterranean (Supporting Information Figure S2). We further considered some of the classifications as assignment errors, that is with geographical distributions outside the possible migratory routes of *V. cardui*. Consequently, we regarded the area of occurrence of these taxa as undetermined in our further analyses. The extent of occurrence for each informative plant species was delimited on the maps by the minimum convex polygon containing all known sites of occurrence, according to IUCN criteria (IUCN, 2012).

2.5 | Wind track

Backward wind trajectories were reconstructed for sampling sites and dates, using the Hybrid Single-Particle Lagrangian Integrated Trajectory dispersion model from NOAA's Air Resources Laboratory (Rolph, Stein, & Stunder, 2017; Stein et al., 2015). Analyses were based on the Reanalysis database and computed on 48-hr back trajectories arriving at sites at 12:00 hr UTC and for three different altitudinal layers (500, 1,000 and 1,500 m asl).

3 | RESULTS

3.1 | Sampling

A total of 47 *V. cardui* specimens were collected in the seven sites (Figure 1; Supporting Information Table S1) and processed for pollen metabarcoding analysis. The sampling consisted mostly of two migratory groups, according to the regions (Andalusia and Catalonia) and timing (February and April). First, a group of samples was collected in several localities of Andalusia between February 16th and 25th of 2016, a year with exceptionally early sightings of the species in the region. The individuals from this group were generally collected close to the beach, but not exactly at the moment of arrival, although a noticeable increase in individuals was observed on the 21st of February. A second group of samples was collected at the beach of Delta del Llobregat, near Barcelona, along two consecutive days (27th and 28th) in April of 2012, after a storm carrying Saharan dust and strong winds from the south. The individuals from this group were collected when landing on the beach, coming from the sea in a northwest direction. Finally, one sample was collected in June 2012 from the Columbretes Islands, a small archipelago 50 km from Castelló coast.

3.2 | Sequencing results and reads processing

We obtained 9,654,286 raw paired-end reads, of which only 1,420 (0.0147%) were present in the blank samples. The average number of reads per sample was 205,380 (excluding blank samples; range: 7–608,640), of which 98.7% were successfully merged (Supporting Information Table S2). After quality filtering, we retained a total of 7,315,458 reads, of which 1,236,996 (16.9%) were discarded as singletons. The rest of the reads were dereplicated (separately within each sample) into 229,295 unique sequences of which 35,393 (15.4%; that is 2,908,006 reads or 47.8% of all the filtered reads) were classified to the species level with probability ≥95%. The proportion of the filtered reads classified to higher taxonomic ranks was much higher: 99.9% to the division, 78.3% to the order and 65.7% to the family level (Supporting Information Figure S2). After filtering out the plant species represented by <100 reads in a butterfly sample, 2,880,443 reads were retained (47.4% of the filtered reads; Supporting Information Table S2).

Sequencing success was not even among the butterfly samples but rather followed a bimodal distribution (Supporting Information Figure S2a)—17 samples yielded <10,000 raw sequences (mean = 1,041; range: 7–7,702), but the remaining 30 samples had high number of reads (mean = 321,172; range: 37,420–608,640; Supporting Information Table S2). We also detected variation among the four sample batches: Samples from batches number 1 and 3 had a low coverage, as did five out of 11 samples from batch 2. In contrast, most of the samples of batch 4 had a high coverage (Supporting Information Figure S2b; Table S2). Sequences assigned to the species level with high probability and coverage higher than 100 reads were found in 30 out of 47 (63.8%) butterfly samples and were only present in the samples that produced at least 10,000 raw reads. No reads were retained after the filtering steps for the sample from Columbretes Islands (the only one not attributable to Andalusia or Catalonia).

The amount of PCR product in blank samples was too low to be visible in TapeStation profiles. During the sequencing, PCR blank samples yielded from 0 to 5 reads. Extractions from blank samples produced between 3 and 26 raw reads, except EB4 from which we obtained 1,366 reads. After processing the latter, we detected 42

unique sequences, of which only three were classified to the species level with probability ≥95%, belonging to *Moricandia moricandioides* (11 reads in total) and *Prunus dulcis* (27 reads). These two plant species were detected in some of the studied butterflies and are not present in the area where the libraries were prepared (Kraków, Poland), pointing rather to a low level of cross-contamination than to external contamination in one of the batches. As these sequences were well below the threshold of 100 reads per species, they were not retained in the final data set.

3.3 | Sequence classification and plant diversity

In total, the filtered reads were classified into 157 species (Supporting Information Table S3). Most of the detected plant species were classified with very high probabilities: 97 species with 100% probability, 25 with 99%, 19 with 98%, 8 with 97%, 3 with 96% and 6 species with 95% probability (Supporting Information Table S3). The highest classification probabilities for each butterfly sample and plant species combination are given in Supporting Information Table S4. The sequence most frequently represented in the samples (present in 21 samples) was classified as *Alternanthera* sp. This appears to be an erroneous GenBank entry (JX136744.1), most likely of fungal origin, and it was removed from our data set. Excluding this, the most represented species (in 12 out of 30 butterfly samples for which plant sequences were detected) was *Reseda lutea* a widespread species in Europe and the Mediterranean region, including both sampling areas (Andalusia and Catalonia). Approximately half of the plant species (76) were present only in a single butterfly sample each and another 42 species in just two butterfly samples each. On average, there were 12.2 species detected per butterfly sample (*SD*: 6.9; range: 2–26).

The diversity of detected pollen included species from 23 orders of plants in total, where 13 orders were represented in samples from Catalonia (April) and 23 orders in Andalusia (February). Asterales were the dominant order at both sites, with 31 species (14 and 21 species in Catalonia and Andalusia, respectively). Most of the plants detected were typically insect-pollinated (82.8%), while 15.3% were species not pollinated by insects. The proportion of insect-pollinated plant species was high at both sites: 78.7% Catalonia and 85.6% in Andalusia (Figure 2).

Sequences from all green plants (Viridiplantae) were present in the reference database, and some sequences from our data matched green algae (Chlorophyta). Sequences from two samples were assigned to the algae genus of *Trebouxia*. Another species of green algae, *Pseudostichococcus monallantoides*, was also detected in one sample.

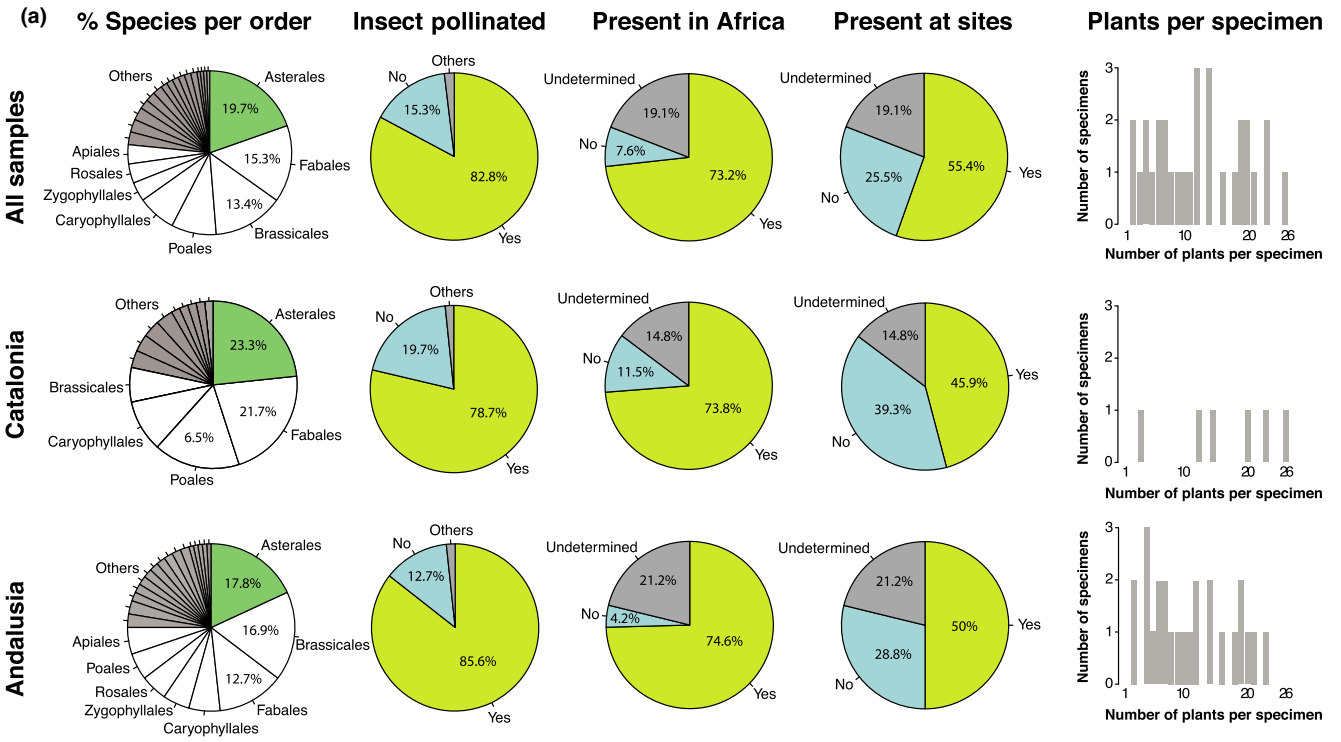
Of the plant species detected from sequencing data, 40 (25.5%) were alien to the sampling sites (24 for Catalonia and 34 for Andalusia) and are thus potentially informative in estimating migratory paths (listed and illustrated in Supporting Information Table S3 and Figure S5). A total of 115 plant species detected (73.2%) are present in Africa–Arabia and 85 in both Europe and Africa–Arabia (54.1%). Reads classified as these also constituted the majority of the reads (77.9%). From these, 30 species (19.1% of the total) are present in Africa and absent in Europe, and were detected in 20 of our samples (9.0% of the classified reads; Supporting Information Table S3). Another four species (2.5%) are endemic to the Canary Islands and were detected in seven butterfly samples. Six plant species were present in more than two butterfly samples: *Calendula stellata* (nine butterfly samples)—occurring in Morocco, Algeria, Tunisia and Sicily, introduced in the Canary Islands; *Launaea mucronata* (six samples)—a species of Saharo-Arabian distribution, also present in the Canary Islands; *Lotus weilleri* (four samples)—a species endemic to northern Atlantic Morocco; *Oxalis compressa* (four samples)—native to Southern Africa but introduced in the Mediterranean Region of Morocco and Algeria; *Raffenaldia primuloides* (four samples)—present in Morocco and Algeria; and *Farsetia aegyptia* (three samples)—a desert species distributed throughout Northern Africa (Morocco, Algeria and Tunisia but absent from Libya and Egypt) and Asia (Supporting Information Table S3, Figure S5).

4 | DISCUSSION

4.1 | Pollen metabarcoding as a tool

Morphological identifications of pollen grains carried by insects have already been used to infer long-distance migratory patterns in insects (Gregg, 1993; Hendrix & Showers, 1992; Hendrix et al., 1987; Lingren et al., 1993, 1994; Mikkola, 1971; Westbrook et al., 1997). The use of this method, however, is limited because morphological identification by light microscopy is a time-consuming task, it requires specialized taxonomic expertise and can hardly provide species-level determinations (Galimberti et al., 2014; Hawkins et al., 2015; Keller et al., 2015; Kraaijeveld et al., 2015; Richardson, Lin, Sponsler, et al., 2015; Sickel et al., 2015). DNA metabarcoding is a fast, high-throughput method that greatly simplifies the identification process (Taberlet et al., 2012) and captures high diversity of pollen that is transported by insects. Our results show that identifying pollen grains carried by migrating insects through DNA metabarcoding is feasible to the species level and that high pollen diversity per specimen can be detected. Thanks to species-level identifications for multiple taxa, migratory paths of the insects can be traced and narrowed by additive geographic distributions of the plants. The method presented here has a wide application to all major insect orders that visit flowers (Coleoptera, Diptera, Hymenoptera and Lepidoptera; Kevan & Baker, 1983), as well as some vertebrates such as birds (Cronk & Ojeda, 2008) or bats (Fleming, Geiselman, & Kress, 2009). Moreover, it could also be used to track historical locations or origins of products and potentially even provide forensic evidence.

We show that the *V. cardui* individuals analysed here were migrating and originated in the African continent. Among the samples from which pollen was detected, all but three were carrying pollen attributed to plant species alien to their collecting site, most of it corresponding to African endemic plants. *Vanessa cardui* is typically colonizing the western Mediterranean Europe in spring, with most



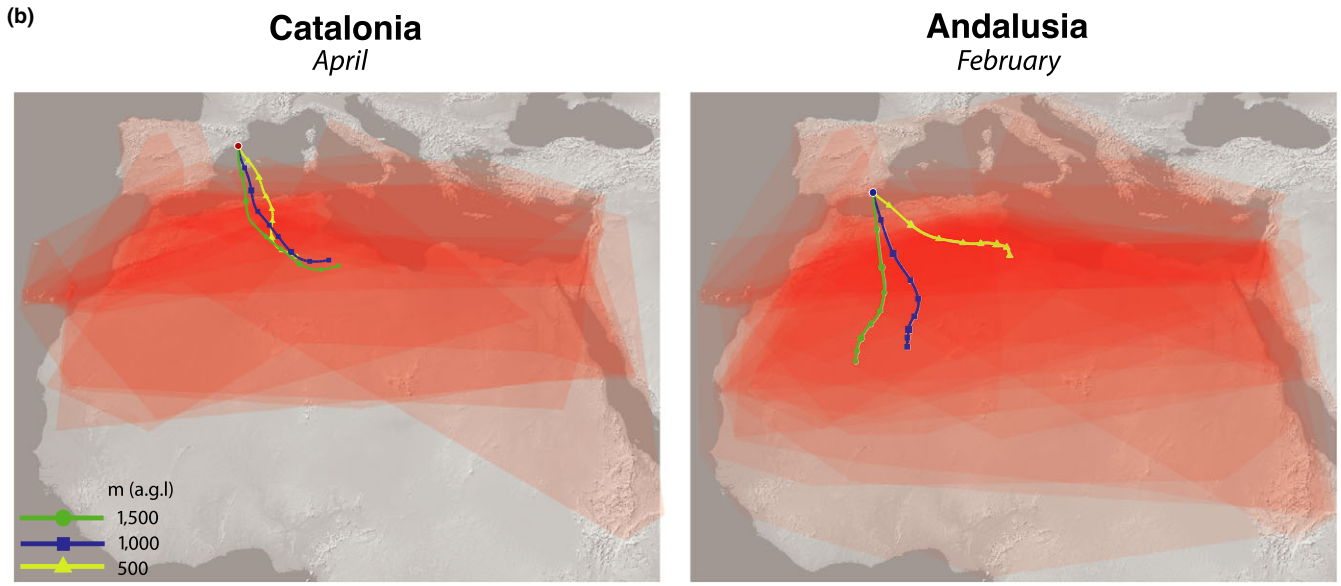


FIGURE 2 Detected plant species that are informative of migration: (a) Pies showing classified percentages of plants per order, insect-pollinated and present in Africa and at sampling sites. Barplots show number of informative plant species detected per specimen. Pies and barplots are shown for all samples together and for two migratory waves (Catalonia and Andalusia) independently. (b) Additive extent of occurrences for informative plants detected in Catalonia and Andalusia migrants. Higher red intensity indicates higher probability of plant overlap. Forty-eight hours backward wind trajectories at three altitudes (500, 1,000 and 1,500 m agl) are shown (colour lines) for the specific dates of observed peaks of migration (April 27th—Catalonia, February 21—Andalusia)

arrivals usually observed in April and May (Stefanescu et al., 2013). The origin of spring arrivals into the Iberian Peninsula has typically been associated to large breeding sources found in Morocco in March–April (Stefanescu, Alarcón, Izquierdo; Páramo, & Ávila, 2011; Stefanescu et al., 2013). Our results partially agree with this timing and path. On the one hand, we show that arrivals to Europe can occur as early as February in Andalusia. On the other hand, the pollen associated with butterflies exhibits a much larger geographical range where butterflies might originate and transit (Figure 2). Unlike isotopic analyses, our approach does not test for natal origins of migrations, but for the most likely paths used during their migrations. During migration, butterflies generally stop in the evening, feed and

rest at night. In the morning, they feed until it is warm enough and winds are suitable for continuing the migration (Shields, 1974). Migratory paths would be thus defined by the position of stepping-stone locations where they fed. In fact, it is expected that pollen from most recently visited flowers is better represented on butterfly bodies than that from plants visited right after emergence. Thus, we may assume a "dilution effect" of the signal to a certain degree.

4.2 | Sequencing protocol

The flexible two-step laboratory protocol applied here can be easily adapted to other markers, for instance standard plant plastid barcodes such as rbcL and matK (CBOL Plant Working Group et al., 2009), just by replacing the primers used in the first PCR. Importantly, the proposed method allows for high sample multiplexing: Up to 384 samples can be analysed simultaneously when combined with Illumina TruSeq CD index sequences, and even more when longer index sequences are used (Fadrosh et al., 2014). Previous pollen metabarcoding protocols used pollen pulverization with bullet blender and DNA isolation with commercial kits (Kraaijeveld et al., 2015; Simel, Saidak, & Tuskan, 1997). Assessing the best pollen isolation method was beyond the scope of this study, but we found that skipping the pollen isolation step and using homogenized pollen directly in the PCR with "direct" polymerase mix (Wong et al., 2014) is an efficient method of amplifying DNA markers from pollen loads (see Supporting Information Appendix S1). This way, DNA isolation—the most time-consuming step of pollen metabarcoding projects (Bell et al., 2017)—is avoided altogether. Nevertheless, further assessment of the effectiveness of such procedure in amplifying markers from all the plant species present in the pollen loads is necessary. Moreover, careful contamination and cross-contamination control by using blank samples and following best practices to avoid contamination are necessary in metabarcoding studies (Deiner et al., 2017; Goldberg et al., 2016). In our case, we used blank samples both at the pollen isolation and the library preparation steps. Both isolation and pre-PCR steps were also conducted under laminar flow cabinet. These remedial steps ensured no external contamination in our samples, as shown by blank samples and the virtual absence of plants native to Central Europe in our data set.

Another source of bias is the errors occurring at the PCR and sequencing step (Coissac, Riaz, & Puillandre, 2012). Many metabarcoding pipelines perform clustering of similar sequences in order to reduce the number of low-copy reads that are usually erroneous artefacts and cluster them together with the centroid sequence (Edgar, 2010; Rognes et al., 2016) or "denoising" the reads in order to remove the putatively erroneous sequences (Edgar, 2016, 2016). Studies with mock pollen samples are still needed to assess the relative performance of these methods for pollen metabarcoding.

Many factors can bias the PCR amplification of DNA templates and skew the quantitative representation of the sequenced species in the obtained reads. These factors involve PCR biases, either caused by sequence polymorphisms in priming sites (Elbrecht & Leese, 2015; Piñol, Mir, Gomez-Polo, & Agust, 2015; Sipos et al., 2007), formation of chimeric reads (Bjørnsgaard Aas, Davey, & Kauserud, 2017), sequence length or GC content (Krehenwinkel et al., 2017). Despite the variation in the number of plastid genome copies in pollen is poorly understood (Bell et al., 2016), some studies show some relationship between pollen abundance and the number of reads for plastid markers (Kraaijeveld et al., 2015; Richardson, Lin, Quijia, et al., 2015) and ITS1 (Pornon et al., 2016). The studies on ITS2 marker did not find such relationship (Hawkins et al., 2015; Richardson, Lin, Sponsler, et al., 2015), and the factors affecting rDNA copy variation are unclear (Prokopowich, Gregory, & Crease, 2003). The number of reads obtained per species should be therefore treated with caution and only as a semi-quantitative method of estimating pollen abundance. Thus, following Yu et al. (2012), we used only presence–absence information when interpreting our results.

4.3 | Limitations to track migrations: taxonomic assignments and species distributions

Several key factors determine the accuracy and resolution of our method for studying insect migrations. First, reference sequences from correctly determined plant species are necessary to properly classify the obtained reads. Uneven geographical coverage of sequences present in reference databases, with a bias towards better studied areas such as Europe or Northern America (Ankenbrand, Keller, Wolf, Schultz, & Förster, 2015; Bell et al., 2016), is of a particular concern. Long-distance migration studies would benefit largely from global plant species coverage, which still remains a remote prospect. In our case, although we were able to classify many reads with high probability, a high proportion (52.6%) of unclassified reads remained. This is most likely caused by a relatively high number of African species/populations missing in our reference database that were therefore classified with low probability to the species level. Indeed, a large proportion of reads were only classified with high probability into higher taxonomic ranks (Supporting Information Figure S3). Representation of species in databases and taxonomic errors are especially problematic with best-hit approaches (e.g., BLAST; Boratyn et al., 2013), and nevertheless, such methods are still used in metabarcoding studies (Hawkins et al., 2015; Kraaijeveld et al., 2015; Richardson, Lin, Sponsler, et al., 2015). Sequence classifiers, such as SINTAX used here, are more robust in such cases. Nevertheless, all current methods display high over-classification rates in cases when taxa are missing from reference databases (Edgar, 2018, 2016) and a trade-off between sensitivity and error rate (Richardson, Bengtsson-Palme, & Johnson, 2017).

In this study, we did not prioritize a full assessment of plant pollen present on the migrating butterflies, and, in order to reduce false positives, we used a conservative 95% sequence assignment threshold for the classified reads to be retained in a final data set. Despite that, we could still detect a small number of species that are probably assignment errors, that is with geographical distributions outside the possible migratory routes of *V. cardui* even for the species classified with 100% probability. This is probably because the

representation of plant species belonging to taxonomically complex and diverse genera is far from complete in reference databases used for taxonomic assignments and could affect the positive identification of some pollen grains. Some examples of such taxa in our data set include *Astragalus*, with about 2,500–3,000 species, the largest genus of flowering plants (Bagheri, Maassoumi, Rahiminejad, Brassac, & Blattner, 2017; Podlech & Zarre, 2013); *Artemisia* that covers approximately 600 species (Richardson, Page, Bajgain, Sanderson, & Udall, 2012); and genus *Thymus* that includes around 400 species (Karaca, Ince, Aydin, Elmasulu, & Turgut, 2015). Most of the species of these genera are native to the Mediterranean region, Northern Africa and Western Asia. Molecular approaches have limitations to identify and define species in some of these complex genera due to various biological phenomena, such as interspecific hybridization and polyploidy, which are often correlated (Soltis & Soltis, 2009) and can contribute significantly to the taxonomic complexity of *Thymus* (Morales, 2010), *Artemisia* (Richardson et al., 2012) and *Astragalus* (Bagheri et al., 2017; Doyle, 2012). In *Thymus*, genetic polymorphism at the intraspecific level can hinder the positive identification of some species (Karaca et al., 2015). Conversely, DNA sequence diversity is generally very low in some *Artemisia* species (Koloren, Koloren, & Eker, 2016) and also among species included within several sections of *Astragalus* (Bagheri et al., 2017), which has been attributed to a rapid radiation (Sanderson & Wojciechowski, 1996).

In order to infer the migration routes from metabarcoding sequences, detailed plant distribution data are required. Evident gaps in the distribution of the plant species detected in North Africa exist. In particular, presence data available for Algeria and Libya are extremely poor when compared to Morocco or Tunisia. Such biases preclude more detailed analyses based on actual presence records or geographical grids (see Supporting Information Figure S4 for results based on a presence grid). To avoid the influence of important gaps in presence data, distribution ranges delimited by peripheral presence records may be used (Figure 2). The above-mentioned limitations point to the importance of basic taxonomic, barcoding and floristic research, which is the cornerstone for myriad of studies.

More research is needed on pollen retention on insects. For instance, Del Socorro and Gregg (2001) found that the sunflower pollen is a transient marker that is only informative of plant visits that occurred during the previous 2 days. On the other hand, some studies, including the one presented here, show support for long-distance pollen transport (Ahmed et al., 2009; Hendrix & Showers, 1992). In this line, it is worth noting that no informative data were retrieved from a percentage of butterfly specimens (36%). In any case, pollen grains are probably lost along time, and a dilution effect of the pollen load signal is to be expected, with a higher representation of recently visited flowers.

Despite the current limited resolution due to the completeness of reference genetic libraries and plant species presence record databases, pollen metabarcoding is an effective and informative method for tracking insect movements. The accumulation of this kind of information grows rapidly thanks to the new sequencing techniques and citizen science initiatives, for example. As these two factors will most likely improve in the near future, the resolution and usefulness of pollen metabarcoding as a tool for tracking insect migrations can only increase.

4.4 | Pollen detected and migrations of *Vanessa cardui*

Using our metabarcoding approach, we were able to amplify a wide range of plant DNA sequences from migrating *V. cardui*. We generally collected the butterflies immediately or soon after they landed on the beaches of the Mediterranean (note that in several instances, we cannot discard that they fed on local flowers). Our study was therefore designed to test the feasibility of pollen detection after a long-distance migration. In this particular case, we expected to detect pollen of plant species present in Africa, including the Maghreb, the Sahara and the sub-Sahara. According to our hypothesis, a large proportion of classified reads belonged to plants present in Africa–Arabia (73.2% of the detected species and 77.9% of the classified reads) and a notable amount were African or African-Arabian endemics (19.1%).

Pollen composition may explain individual migratory histories, but it can also reflect collective migratory histories given a time and site. Butterflies collected from the same spatiotemporal migratory waves (Andalusia in February and Catalonia in April) show parallelisms, but also differences, in their carrying pollen composition. Such differences among specimens could be explained by the diversity of the visited flowers and in the retention of pollen grains. In some cases, though, the plants found in different specimens are geographically exclusive, which suggest that, if taxonomic attributions are correct, the butterflies may have had different breeding origins and confluence during their migratory paths or at destination.

The two waves of migrants studied (Andalusia and Catalonia) could either correspond to populations originated during the winter in the Maghreb or to populations originated in tropical Africa that may replenish the temperate zone in early spring (Talavera & Vila, 2016). The latter hypothesis cannot be excluded based on our data because, in addition to a marked influence from Maghreb flora, several floristic elements from the Sahara and Sub-Saharan Africa were detected (Figure 2). Generally, the butterflies collected in February (Andalusia) showed a higher number of plants of predominantly Saharan distribution (*Farsetia stylosa*, *Launaea capitata*, *L. mucronata*, *Moltkioptis ciliata*, *Reseda villosa*, *Gymnocarpus decandrus*, *Euphorbia guyoniana*) or with an important sub-Saharan representation (*Pergularia tormentosa*, *Musa acuminata*). A sub-Saharan origin for many of these early *V. cardui* migrants arriving to the Iberian Peninsula has been unambiguously demonstrated by isotope analysis (Talavera et al., 2018). Pollen of *Musa* (banana), only detected in specimen 16C413, is extensively cultivated in tropical Africa and is abundant in the Canary Islands, but rarer in the Maghreb and Europe (where is generally cultivated in greenhouses). In addition, this plant is apomictic and although some varieties preserve male flowers, these are very rare. As the probability to find male banana flowers is higher where the plant is common, a sub-Saharan origin for this sample is likely.

Butterflies from April (Catalonia) had few strictly Saharan floristic elements, such as *L. capitata* and *L. mucronata*, and mostly had representation of flora from the Maghreb (Figure 2). Pollen endemic to the Canary Islands was detected in some instances, both from February and April. Thus, an origin in these islands cannot be discarded for some specimens.

Plant phenology (specifically flowering time) could be another potential source of information for inferring migratory paths, in addition to plant distribution ranges. The plant species for which sequences have been obtained must have been in flower somewhere along the route of the insects in the few hours/days previous to capture. In fact, all African plants detected in our study flower somewhere in N. Africa at the time that the butterflies were collected. If spatial variability on flowering times exists, it is theoretically possible to geographically narrow the potential origin of the pollen and, thus, of the insect route. This additional dimension may be very useful in areas where plant phenology is well studied, including spatial and inter-annual variation. This is clearly not the case for North Africa at present, but we urge the study of plant phenology in addition to ranges and genetics so that pollen metabarcoding techniques benefit and increase resolution.

Backtrack wind models agree with migratory paths coming from the African continent (Figure 2), considering that *V. cardui* migrations can greatly be aided by winds (Stefanescu, Alarcón, & Àvila, 2007). Winds consistently came from the south-east through central Algeria in both waves, at least during the previous 48 hr. Precise migratory sources or paths cannot be inferred based on pollen data or winds alone, but both can be combined to narrow predictions. Actually, the Algerian Maghreb (for the Catalonia April migratory wave) and the Algerian Sahara (for the Andalusia February wave) are the areas where higher accumulative probabilities of identified plants overlap. The combined evidence points to a highly probable origin or pass of these two migratory waves across Algerian grounds, although some individuals may have followed different paths.

Insect-pollinated plants prevailed in our results (82.8%), as expected in pollen obtained from butterfly bodies. Many of the detected plant species belonged to the Asterales, an order that includes many of *V. cardui* typical host plants and nectar sources (Nylin, Slove, & Janz, 2014). In fact, the spectrum of plants visited by *V. cardui* is very wide, including most plant orders and life forms from small plants to trees. For example, numerous *V. cardui* specimens migrating northwards in February 2017 in south Morocco were observed to stop briefly to feed on the flowers of *Prunus dulcis* orchards (R. Vila, pers. obs.), a tree for which we detected pollen in seven butterfly specimens.

Most non-insect-pollinated plants detected were wind-pollinated plants that belonged to the Poales. This points to the accidental transport of such pollen due to physical contact, as *V. cardui* tends to rest on the ground and grass, and only rarely on trees. Sequences from two samples were assigned to the algae genus *Trebouxia*, a common and widespread photobiont in lichens (Dal Grande et al., 2014). Lichens often reproduce asexually by soredia—small powdery propagules containing both fungus and algae. Such small structures could possibly stick to the body of butterflies accidentally and be transported on larger distances.

4.5 | Implications for plant pollination

Here, we show that *V. cardui* could potentially mediate transcontinental pollination, and thus gene flow, for plants species that occur in both Europe and Africa (these accounted for 54% of our data set). Our approach does not allow a precise measure of pollen load, although it is obvious that pollen loads after a transcontinental trip must be relatively low. However, given the millions of individuals of *V. cardui* in particular, and of insects in general, that migrate every year across continents (Hu et al., 2016), the effects of this phenomenon on ecosystems and crops may not be negligible.

Regular long-distance gene flow is a phenomenon that has rarely been acknowledged, but that could potentially explain particular phylogeographic patterns in plants. For example, the Strait of Gibraltar has been shown as not effective at interrupting gene flow in *Androcymbium gramineum* (Caujapé-Castells & Jansen, 2003), several *Cistus* spp. (Fernández-Mazuecos & Vargas, 2010), *Hypochaeris salzmanniana* (Ortiz, Tremetsberger, Talavera, Stuessy, & García-Castaño, 2007) and *Rosmarinus officinalis* (Mateu-Andrés et al., 2013). It is also worth noting that we detected sequences belonging to crops in our data set—*Allium sativum*, *Cucumis sativus* and *Prunus dulcis*—which points to the possibility of intercontinental pollination for these economically important species. We suggest that knowledge on the insect species that perform long-distance migration and the routes and temporal patterns they follow may be of high importance for better understanding intercontinental plant pollination, with implications extending from plant phylogeography to ecosystem services.

ACKNOWLEDGEMENTS

We thank Martin Gascoigne-Pees for important sample contribution. GT is supported by the Marie Curie Actions FP7-People-2013 IOF (project 622716) and by the MINECO programme Juan de la Cierva Incorporación (IJCI-2016-29083). Funding was provided by the statutory funds of W. Szafer Institute of Botany, Polish Academy of Sciences to TS and MR, by the British Ecological Society (grant LRB16/1015) to GT, by project CGL2016-76322 (AEI/FEDER, UE) and by the Committee for Research and Exploration of the National Geographic Society (grant WW1-300R-18) to GT and RV.

AUTHOR CONTRIBUTIONS

T.S., G.T. and R.V. conceived the study and wrote the manuscript. G.T. and R.V. collected samples. T.S. and M.R. designed the laboratory part of the study and T.S. carried out laboratory work. T.S. and G.T. analysed data. L.S. gathered plant distribution data and assessed plant determination and taxonomy. All authors edited and approved the final version of the manuscript.

DATA ACCESSIBILITY

Scripts used to process the data: https://github.com/TomaszSuchan/pollen-metabarcoding

ITS2 database for SINTAX: https://github.com/molbiodiv/meta-barcoding-dual-indexing/blob/master/precomputed/viridiplantae_all_2014.sintax.fa

Raw sequences: European Nucleotide Archive, https://www.ebi.ac.uk/ena/data/view/PRJEB26439

ORCID

Tomasz Suchan http://orcid.org/0000-0002-0811-8754
Gerard Talavera http://orcid.org/0000-0003-1112-1345
Llorenç Sáez http://orcid.org/0000-0003-4551-2432
Michał Ronikier http://orcid.org/0000-0001-7652-6787
Roger Vila http://orcid.org/0000-0002-2447-4388

REFERENCES

African Plant Database (2018). *African Plant Database (version 3.4.0)*. Conservatoire et Jardin botaniques de la Ville de Genève and South African National Biodiversity Institute, Pretoria. Retrieved from https://www.ville-ge.ch/musinfo/bd/cjb/africa/

Ahmed, S., Compton, S. G., Butlin, R. K., & Gilmartin, P. M. (2009). Wind-borne insects mediate directional pollen transfer between desert fig trees 160 kilometers apart. *Proceedings of the National Academy of Sciences*, *106*(48), 20342–20347. https://doi.org/10.1073/pnas.0902213106

Ankenbrand, M. J., Keller, A., Wolf, M., Schultz, J., & Förster, F. (2015). ITS2 database V: Twice as much. *Molecular Biology and Evolution*, *32*(11), 3030–3032.

Bagheri, A., Maassoumi, A. A., Rahiminejad, M. R., Brassac, J., & Blattner, F. R. (2017). Molecular phylogeny and divergence times of *Astragalus* section *Hymenostegis*: An analysis of a rapidly diversifying species group in Fabaceae. *Scientific Reports*, *7*, 14033. https://doi.org/10.1038/s41598-017-14614-3

Bauer, S., & Hoye, B. J. (2014). Migratory animals couple biodiversity and ecosystem functioning worldwide. *Science*, *344*, 1242552.

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

Bell, K. L., Fowler, J., Burgess, K. S., Dobbs, E. K., Gruenewald, D., Lawley, B., … Brosi, B. J. (2017). Applying pollen DNA metabarcoding to the study of plant–pollinator interactions. *Applications in Plant Sciences*, *5*(6), 1600124. https://doi.org/10.3732/apps.1600124

Bjørnsgaard Aas, A., Davey, M. L., & Kauserud, H. (2017). ITS all right mama: Investigating the formation of chimeric sequences in the ITS2 region by DNA metabarcoding analyses of fungal mock communities of different complexities. *Molecular Ecology Resources*, *17*(4), 730–741. https://doi.org/10.1111/1755-0998.12622

Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., … Raytselis, Y. (2013). BLAST: A more efficient report with usability improvements. *Nucleic Acids Research*, *41*(W1), W29–W33. https://doi.org/10.1093/nar/gkt282

Brattström, O., Bensch, S., Wassenaar, L. I., Hobson, K. A., & Åkesson, S. (2010). Understanding the migration ecology of European red admirals *Vanessa atalanta* using stable hydrogen isotopes. *Ecography*, *33*, 720–729. https://doi.org/10.1111/j.1600-0587.2009.05748.x

Bruni, I., Galimberti, A., Caridi, L., Scaccabarozzi, D., De Mattia, F., Casiraghi, M., & Labra, M. (2015). A DNA barcoding approach to identify plant species in multiflower honey. *Food Chemistry*, *170*, 308–315. https://doi.org/10.1016/j.foodchem.2014.08.060

Caujapé-Castells, J., & Jansen, R. K. (2003). The influence of the Miocene Mediterranean desiccation on the geographical expansion and genetic variation of *Androcymbium gramineum* (Cav.) McBride (Colchicaceae). *Molecular Ecology*, *12*, 1515–1525. https://doi.org/10.1046/j.1365-294X.2003.01839.x

CBOL Plant Working Group, Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., … Fazekas, A. J. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, USA*, *106*(31), 12794–12797. https://doi.org/10.1073/pnas.0905845106

Chapman, J. W., Drake, V. A., & Reynolds, D. R. (2011). Recent insights from radar studies of insect flight. *Annual Review of Entomology*, *56*, 337–356. https://doi.org/10.1146/annurev-ento-120709-144820

Chapman, J. W., Reynolds, D. R., Smith, A. D., Riley, J. R., Pedgley, D. E., & Woiwod, I. P. (2002). High-altitude migration of the diamondback moth *Plutella xylostella* to the UK: A study using radar, aerial netting, and ground trapping. *Ecological Entomology*, *27*(6), 641–650. https://doi.org/10.1046/j.1365-2311.2002.00472.x

Chapman, J. W., Reynolds, D. R., Smith, A. D., Smith, E. T., & Woiwod, I. P. (2004). An aerial netting study of insects migrating at high altitude over England. *Bulletin of Entomological Research*, *94*, 123–136. https://doi.org/10.1079/BER2004287

Chapman, J. W., Reynolds, D. R., & Wilson, K. (2015). Long-range seasonal migration in insects: Mechanisms, evolutionary drivers and ecological consequences. *Ecology Letters*, *18*, 287–302. https://doi.org/10.1111/ele.12407

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., … Luo, K. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE*, *5*(1), e8613. https://doi.org/10.1371/journal.pone.0008613

Chilson, P. B., Frick, W. F., Kelly, J. F., Howard, K. W., Larkin, R. P., Diehl, R. H., … Kunz, T. H. (2012). Partly cloudy with a chance of migration: Weather, radars, and aeroecology. *Bulletin of the American Meteorological Society*, *93*(5), 669–686. https://doi.org/10.1175/BAMS-D-11-00099.1

Coissac, E., Riaz, T., & Puillandre, N. (2012). Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology*, *21*(8), 1834–1847. https://doi.org/10.1111/j.1365-294X.2012.05550.x

Cronk, Q., & Ojeda, I. (2008). Bird-pollinated flowers in an evolutionary and molecular context. *Journal of Experimental Botany*, *59*(4), 715–727. https://doi.org/10.1093/jxb/ern009

Dal Grande, F., Alors, D., Divakar, P. K., Bálint, M., Crespo, A., & Schmitt, I. (2014). Insights into intrathalline genetic diversity of the cosmopolitan lichen symbiotic green alga *Trebouxia decolorans* Ahmadjian using microsatellite markers. *Molecular Phylogenetics and Evolution*, *72*, 54–60. https://doi.org/10.1016/j.ympev.2013.12.010

de Vere, N., Jones, L. E., Gilmore, T., Moscrop, J., Lowe, A., Smith, D., … Ford, C. R. (2017). Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. *Scientific Reports*, *7*, 42838.

Dean, T. J., & Drake, V. A. (2005). Monitoring insect migration with radar: The ventral-aspect polarization pattern and its potential for target identification. *International Journal of Remote Sensing*, *26*(18), 3957–3974. https://doi.org/10.1080/01431160500165955

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., … Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895. https://doi.org/10.1111/mec.14350

Del Socorro, A. P., & Gregg, P. C. (2001). Sunflower (*Helianthus annuus* L.) pollen as a marker for studies of local movement in *Helicoverpa*

armigera (Hübner) (Lepidoptera: Noctuidae). *Austral Entomology*, *40*(3), 257–263.

Doyle, J. J. (2012). Polyploidy in legumes. In P. S. Soltis & D. E. Soltis (Eds.), *Polyploidy and genome evolution* (pp. 147–180). Berlin, Germany: Springer.

Drake, V. A., Chapman, J. W., Lim, K. S., Reynolds, D. R., Riley, J. R., & Smith, A. D. (2017). Ventral-aspect radar cross sections and polarization patterns of insects at X band and their relation to size and form. *International Journal of Remote Sensing*, *38*(18), 5022–5044. https://doi.org/10.1080/01431161.2017.1320453

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Edgar, R. C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv*, 081257.

Edgar, R. C. (2016). SINTAX: A simple non-Bayesian taxonomy classifier for 16S and ITS sequences. *BioRxiv*, 074161.

Edgar, R. C. (2018). Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. *PeerJ*, *6*, e4652. https://doi.org/10.7717/peerj.4652

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLoS ONE*, *10*(7), e0130324. https://doi.org/10.1371/journal.pone.0130324

Euro+Med (2006). *Euro+Med PlantBase. The information resource for Euro-Mediterranean plant diversity*. Retrieved form https://ww2.bgbm.org/EuroPlusMed/

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, *2*(1), 6. https://doi.org/10.1186/2049-2618-2-6

Fernández-Mazuecos, M., & Vargas, P. (2010). Ecological rather than geographical isolation dominates Quaternary formation of Mediterranean *Cistus* species. *Molecular Ecology*, *19*, 1381–1395.

Fierer, N., Hamady, M., Lauber, C. L., & Knight, R. (2008). The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proceedings of the National Academy of Sciences, USA*, *105*(46), 17994–17999. https://doi.org/10.1073/pnas.0807920105

Fleming, T. H., Geiselman, C., & Kress, W. J. (2009). The evolution of bat pollination: A phylogenetic perspective. *Annals of Botany*, *104*(6), 1017–1043. https://doi.org/10.1093/aob/mcp197

Galimberti, A., De Mattia, F., Bruni, I., Scaccabarozzi, D., Sandionigi, A., Barbuto, M., … Labra, M. (2014). A DNA barcoding approach to characterize pollen collected by honeybees. *PLoS ONE*, *9*, 109363. https://doi.org/10.1371/journal.pone.0109363

Garland, M. S., & Davis, A. K. (2002). An examination of monarch butterfly (*Danaus plexippus*) autumn migration in coastal Virginia. *The American Midland Naturalist*, *147*(1), 170–174. https://doi.org/10.1674/0003-0031(2002)147[0170:AEOMBD]2.0.CO;2

GBIF (2018). *GBIF home page*. Retrieved from https://www.gbif.org

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., … Cornman, R. S. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, *7*, 1299–1307. https://doi.org/10.1111/2041-210X.12595

Gregg, P. (1993). Pollen as a marker for migration of *Helicoverpa armigera* and *H. punctigera* (Lepidoptera: Noctuidae) from western Queensland. *Australian Journal of Ecology*, *18*, 209–219. https://doi.org/10.1111/j.1442-9993.1993.tb00445.x

Hagler, J. R., & Jackson, C. G. (2001). Methods for marking insects: Current techniques and future prospects. *Annual Review of Entomology*, *46*(1), 511–543.

Hawkins, J., de Vere, N., Griffith, A., Ford, C. R., Allainguillaume, J., Hegarty, M. J., … Adams-Groom, B. (2015). Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. *PLoS ONE*, *10*(8), e0134735. https://doi.org/10.1371/journal.pone.0134735

Hendrix, W. H., Mueller, T. F., Phillips, J. R., & Davis, O. K. (1987). Pollen as an indicator of long-range movement of *Heliothis zea* (Lepidoptera: Noctuidae). *Environmental Entomology*, *16*, 1148–1151.

Hendrix, W. H., & Showers, W. B. (1992). Tracing black cutworm and armyworm (Lepidoptera: Noctuidae) northward migration using *Pithocellobium* and *Calliandra* pollen. *Environmental Entomology*, *21*, 1092–1096.

Hobson, K. A. (2008). Applying isotopic methods to tracking animal movements. *Terrestrial Ecology*, *2*, 45–78.

Hobson, K. A., Wassenaar, L. I., & Taylor, O. R. (1999). Stable isotopes (δD and δ13C) are geographic indicators of natal origins of monarch butterflies in eastern North America. *Oecologia*, *120*(3), 397–404. https://doi.org/10.1007/s004420050872

Holland, R. A., Wikelski, M., & Wilcove, D. S. (2006). How and why do insects migrate? *Science*, *313*, 794–796. https://doi.org/10.1126/science.1127272

Howard, E., & Davis, A. K. (2009). The fall migration flyways of monarch butterflies in eastern North America revealed by citizen scientists. *Journal of Insect Conservation*, *13*(3), 279–286. https://doi.org/10.1007/s10841-008-9169-y

Hu, G., Lim, K. A., Horvitz, N., Clark, S. J., Reynolds, D. R., Sapir, N., & Chapman, J. W. (2016). Mass seasonal bioflows of high-flying insect migrants. *Science*, *354*, 1584–1587. https://doi.org/10.1126/science.aah4379

IUCN (2012). *IUCN red list categories and criteria: Version 3.1*. Gland, Switzerland.

Jones, G. D., & Jones, S. D. (2001). The uses of pollen and its implication for entomology. *Neotropical Entomology*, *30*(3), 314–349. https://doi.org/10.1590/S1519-566X2001000300001

Karaca, M., Ince, A. G., Aydin, A., Elmasulu, S. Y., & Turgut, K. (2015). Microsatellites for genetic and taxonomic research on thyme (*Thymus* L.). *Turkish Journal of Biology*, *39*, 147–159. https://doi.org/10.3906/biy-1406-20

Keller, A., Danner, N., Grimmer, G., Ankenbrand, M., von der Ohe, K., von der Ohe, W., … Steffan-Dewenter, I. (2015). Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. *Plant Biology*, *17*, 558–566. https://doi.org/10.1111/plb.12251

Kevan, P. G., & Baker, H. G. (1983). Insects as flower visitors and pollinators. *Annual Review of Entomology*, *28*(1), 407–453. https://doi.org/10.1146/annurev.en.28.010183.002203

Kissling, D. W., Pattemore, D. E., & Hagen, M. (2014). Challenges and prospects in the telemetry of insects. *Biological Reviews*, *89*(3), 511–530. https://doi.org/10.1111/brv.12065

Knight, A., Brower, L. P., & Williams, E. H. (1999). Spring remigration of the monarch butterfly, *Danaus plexippus* (Lepidoptera: Nymphalidae) in north-central Florida: Estimating population parameters using mark-recapture. *Biological Journal of the Linnean Society*, *68*(4), 531–556. https://doi.org/10.1111/j.1095-8312.1999.tb01187.x

Koloren, O., Koloren, Z., & Eker, S. (2016). Molecular phylogeny of *Artemisia* species based on the internal transcribed spacer (ITS) of 18S–26S rDNA in Ordu Province of Turkey. *Biotechnology & Biotechnological Equipment*, *30*(5), 929–934.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, *79*, 5112–5120. https://doi.org/10.1128/AEM.01043-13

Kraaijeveld, K., de Weger, L. A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P. S., & den Dunnen, J. T. (2015). Efficient and sensitive identification and quantification of airborne pollen using

next-generation DNA sequencing. *Molecular Ecology Resources*, *15*, 8–16. https://doi.org/10.1111/1755-0998.12288

Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W. B., & Gillespie, R. G. (2017). Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports*, *7*(1), 17668. https://doi.org/10.1038/s41598-017-17333-x

Lingren, P. D., Bryant, V. M., Raulston, J. R., Pendelton, M., Westbrook, J., & Jones, G. D. (1993). Adult feeding host range and migratory activities of corn earworm, cabbage looper and celery looper (Lepidoptera: Noctuidae) moths as evidenced by attached pollen. *Journal of Economic Entomology*, *86*, 1429–1433.

Lingren, P. D., Westbrook, J. K., Bryant, V. M., Raulston, J. R., Esquivel, J. F., & Jones, G. D. (1994). Origin of corn earworm (Lepidoptera: Noctuidae) migrants as determined by citrus pollen markers and weather systems. *Environmental Entomology*, *23*, 562–570.

Lowe, W. H., & Allendorf, F. W. (2010). What can genetics tell us about population connectivity? *Molecular Ecology*, *19*, 3038–3051. https://doi.org/10.1111/j.1365-294X.2010.04688.x

Lucas, A., Bodger, O., Brosi, B. J., Ford, C. R., Forman, D. W., Greig, C., … de Vere, N. (2018a). Floral resource partitioning by individuals within generalised hoverfly pollination networks revealed by DNA metabarcoding. *Scientific Reports*, *8*, 5133.

Lucas, A., Bodger, O., Brosi, B. J., Ford, C. R., Forman, D. W., Greig, C., … de Vere, N. (2018b). Generalisation and specialisation in hoverfly (Syrphidae) grassland pollen transport networks revealed by DNA metabarcoding. *Journal of Animal Ecology*, *87*(4), 1008–1021.

Lukhtanov, V. A., Pazhenkova, E. A., & Novikova, A. V. (2016). Mitochondrial chromosome as a marker of animal migratory routes: DNA barcoding revealed Asian (non-African) origin of a tropical migrant butterfly *Junonia orithya* in south Israel. *Comparative Cytogenetics*, *10*(4), 671. https://doi.org/10.3897/CompCytogen.v10i4.11085

Lyons, J. I., Pierce, A. A., Barribeau, S. M., Sternberg, E. D., Mongue, A. J., Roode, D., & Jacobus, C. (2012). Lack of genetic differentiation between monarch butterflies with divergent migration destinations. *Molecular Ecology*, *21*(14), 3433–3444. https://doi.org/10.1111/j.1365-294X.2012.05613.x

Mateu-Andrés, I., Aguilella, A., Boisset, F., Currás, R., Guara, M., Laguna, E., … Pedrola, J. (2013). Geographical patterns of genetic variation in rosemary (*Rosmarinus officinalis*) in the Mediterranean basin. *Botanical Journal of the Linnean Society*, *171*, 700–712.

Mikkola, K. (1971). Pollen analysis as a means of studying the migrations of Lepidoptera. *Annales Entomologici Fennici*, *37*, 136–139.

Morales, R. (2010). Thymus L. In R. Morales, A. Quintanar, F. Cabezas, A. J. Pujadas, & S. Cirujano (Eds.), *Flora iberica. Vol. XII. Verbenaceae-Labiatae-Callitrichaceae* (pp. 349–409). Madrid, Spain: Real Jardín Botánico de Madrid (C.S.I.C.).

Nagoshi, R. N., Meagher, R. L., & Hay-Roe, M. (2012). Inferring the annual migration patterns of fall armyworm (Lepidoptera: Noctuidae) in the United States from mitochondrial haplotypes. *Ecology and Evolution*, *2*, 1458–1467. https://doi.org/10.1002/ece3.268

Nylin, S., Slove, J., & Janz, N. (2014). Host plant utilization, host range oscillations and diversification in nymphalid butterflies: A phylogenetic investigation. *Evolution*, *68*(1), 105–124. https://doi.org/10.1111/evo.12227

O'Donnell, J. L., Kelly, R. P., Lowell, N. C., & Port, J. A. (2016). Indexed PCR primers induce template-specific bias in large-scale DNA sequencing studies. *PLoS ONE*, *11*(3), e0148698. https://doi.org/10.1371/journal.pone.0148698

Ortiz, M. A., Tremetsberger, K., Talavera, S., Stuessy, T., & García-Castaño, J. L. (2007). Population structure of *Hypochaeris salzmanniana* DC (Asteraceae), an endemic species to the Atlantic coast on both sides of the Strait of Gibraltar, in relation to Quaternary sea level changes. *Molecular Ecology*, *16*, 541–552. https://doi.org/10.1111/j.1365-294X.2006.03157.x

Osborne, J. L., Loxdale, H. D., & Woiwod, I. P. (2002). Monitoring insect dispersal: Methods and approaches. In J. M. Bullock, R. E. Kenward, & R. S. Hails (Eds.), *Dispersal ecology: The 42nd symposium of the British Ecological Society* (pp. 24–49). Oxford, UK: Blackwell Science.

Piñol, J., Mir, G., Gomez-Polo, P., & Agust, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, *15*, 819–830. https://doi.org/10.1111/1755-0998.12355

Podlech, D., & Zarre, S. (2013). *A taxonomic revision of the genus Astragalus* L. (Leguminosae) in the old world, vol. 1–3. Wien, Austria: Naturhistorisches Museum.

Pornon, A., Escaravage, N., Burrus, M., Holota, H., Khimoun, A., Mariette, J., … Vidal, M. (2016). Using metabarcoding to reveal and quantify plant–pollinator interactions. *Scientific Reports*, *6*, 27282. https://doi.org/10.1038/srep27282

Prokopowich, C. D., Gregory, T. R., & Crease, T. J. (2003). The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, *46*(1), 48–50. https://doi.org/10.1139/g02-103

Prosser, S. W., & Hebert, P. D. (2017). Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding. *Food Chemistry*, *214*, 183–191. https://doi.org/10.1016/j.foodchem.2016.07.077

Richardson, R. T., Bengtsson-Palme, J., & Johnson, R. M. (2017). Evaluating and optimizing the performance of software commonly used for the taxonomic classification of DNA metabarcoding sequence data. *Molecular Ecology Resources*, *17*(4), 760–769. https://doi.org/10.1111/1755-0998.12628

Richardson, R. T., Lin, C.-H., 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.

Richardson, R. T., Lin, C.-H., Sponsler, D. B., Quijia, J. O., Goodell, K., & Johnson, R. M. (2015). Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. *Applications in Plant Sciences*, *3*, 1400066.

Richardson, B. A., Page, J. T., Bajgain, P., Sanderson, S. C., & Udall, J. A. (2012). Deep sequencing of amplicons reveals widespread intraspecific hybridization and multiple origins of polyploidy in big sagebrush (*Artemisia tridentata*; Asteraceae). *American Journal of Botany*, *99*(12), 1962–1975.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. https://doi.org/10.7717/peerj.2584

Rolph, G., Stein, A., & Stunder, B. (2017). Real-time environmental applications and display system: Ready. *Environmental Modelling & Software*, *95*, 210–228. https://doi.org/10.1016/j.envsoft.2017.06.025

Sanderson, M. J., & Wojciechowski, M. F. (1996). Diversification rates in a temperate legume clade: Are there "so many species" of *Astragalus* (Fabaceae)? *American Journal of Botany*, *83*, 1488–1502.

Shields, O. (1974). Toward a theory of butterfly migration. *Journal of Research on the Lepidoptera*, *13*(4), 217–238.

Shields, O. (1992). World distribution of the *Vanessa cardui* group (Nymphalidae). *Journal of the Lepidopterists' Society*, *46*, 235–238.

Sickel, W., Ankenbrand, M. J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., & Keller, A. (2015). Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology*, *15*(1), 20. https://doi.org/10.1186/s12898-015-0051-y

Simel, E. J., Saidak, L. R., & Tuskan, G. A. (1997). Method of extracting genomic DNA from non-germinated gymnosperm and angiosperm pollen. *Bio Techniques*, *22*,390–392, 394.

Sipos, R., Székely, A. J., Palatinszky, M., Révész, S., Márialigeti, K., & Nikolausz, M. (2007). Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial

community analysis. *FEMS Microbiology Ecology*, *60*(2), 341–350. https://doi.org/10.1111/j.1574-6941.2007.00283.x

Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual Review of Plant Biology*, *60*, 561–588. https://doi.org/10.1146/annurev.arplant.043008.092039

Stefanescu, C., Alarcón, M., & Àvila, A. (2007). Migration of the painted lady butterfly, *Vanessa cardui*, to north-eastern Spain is aided by African wind currents. *Journal of Animal Ecology*, *76*(5), 888–898. https://doi.org/10.1111/j.1365-2656.2007.01262.x

Stefanescu, C., Alarcón, M., Izquierdo, R., Páramo, F., & Ávila, A. (2011). Moroccan source areas of the painted lady butterfly *Vanessa cardui* (Nymphalidae: Nymphalinae) migrating into Europe in spring. *Journal of the Lepidopterists' Society*, *65*, 15–26.

Stefanescu, C., Páramo, F., Åkesson, S., Alarcón, M., Ávila, A., Brereton, T., … Hill, J. K. (2013). Multi-generational long-distance migration of insects: Studying the painted lady butterfly in the Western Palaearctic. *Ecography*, *36*(4), 474–486. https://doi.org/10.1111/j.1600-0587.2012.07738.x

Stefanescu, C., Soto, D. X., Talavera, G., Vila, R., & Hobson, K. A. (2016). Long-distance autumn migration across the Sahara by painted lady butterflies: Exploiting resource pulses in the tropical savannah. *Biology Letters*, *12*(10), 20160561. https://doi.org/10.1098/rsbl.2016.0561

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J., Cohen, M. D., & Ngan, F. (2015). NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bulletin of the American Meteorological Society*, *96*(12), 2059–2077. https://doi.org/10.1175/BAMS-D-14-00110.1

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Talavera, G., Bataille, C., Benyamini, D., Gascoigne-Pees, M., & Vila, R. (2018). Round-trip across the Sahara: Afrotropical Painted Lady butterflies recolonize the Mediterranean in early spring. *Biology Letters*, *14*(6), 20180274.

Talavera, G., & Vila, R. (2016). Discovery of mass migration and breeding of the painted lady butterfly *Vanessa cardui* in the Sub-Sahara: The Europe-Africa migration revisited. *Biological Journal of the Linnean Society*, *120*(2), 274–285.

Valentini, A., Miquel, C., & Taberlet, P. (2010). DNA barcoding for honey biodiversity. *Diversity*, *2*(4), 610–617. https://doi.org/10.3390/d2040610

Westbrook, J. K., & Eyster, R. S. (2017). Doppler weather radar detects emigratory flights of noctuids during a major pest outbreak. *Remote Sensing Applications: Society and Environment*, *8*, 64–70. https://doi.org/10.1016/j.rsase.2017.07.009

Westbrook, J. K., Wolf, W. W., Lingren, P. D., Raulston, J. R., Lopez, J. D. Jr, Matis, J. H., … Schleider, P. G. (1997). Early-season migratory flights of corn earworm (Lepidoptera: Noctuidae). *Environmental Entomology*, *26*, 12–20. https://doi.org/10.1093/ee/26.1.12

White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, & T. J. White (Eds.), *PCR protocols: A guide to methods and applications* (pp. 315–322). New York, NY: Academic Press.

Wikelski, M., Moskowitz, D., Adelman, J. S., Cochran, J., Wilcove, D. S., & May, M. L. (2006). Simple rules guide dragonfly migration. *Biology Letters*, *2*(3), 325–329. https://doi.org/10.1098/rsbl.2006.0487

Wong, W. H., Tay, Y. C., Puniamoorthy, J., Balke, M., Cranston, P. S., & Meier, R. (2014). 'Direct PCR'optimization yields a rapid, cost-effective, nondestructive and efficient method for obtaining DNA barcodes without DNA extraction. *Molecular Ecology Resources*, *14*(6), 1271–1280. https://doi.org/10.1111/1755-0998.12275

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, *3*(4), 613–623. https://doi.org/10.1111/j.2041-210X.2012.00198.x

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2013). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Suchan T, Talavera G, Sáez L, Ronikier M, Vila R. Pollen metabarcoding as a tool for tracking long-distance insect migrations. *Mol Ecol Resour*. 2018;00:1–14. https://doi.org/10.1111/1755-0998.12948