

Temperate airborne grass pollen defined by spatio-temporal shifts in community composition

Georgina L. Brennan¹*, Caitlin Potter², Natasha de Vere^{2,3}, Gareth W. Griffith², Carsten A. Skjøth⁴, Nicholas J. Osborne^{5,6,7}, Benedict W. Wheeler⁵, Rachel N. McInnes⁸, Yolanda Clewlow⁸, Adam Barber⁸, Helen M. Hanlon⁸, Matthew Hegarty², Laura Jones^{1,3}, Alexander Kurganskiy⁴, Francis M. Rowney⁵, Charlotte Armitage⁹, Beverley Adams-Groom⁴, Col R. Ford³, Geoff M. Petch⁴, The PollerGEN Consortium¹⁰ and Simon Creer¹⁰*

Grass pollen is the world's most harmful outdoor aeroallergen. However, it is unknown how airborne pollen assemblages change across time and space. Human sensitivity varies between different species of grass that flower at different times, but it is not known whether temporal turnover in species composition match terrestrial flowering or whether species richness steadily accumulates over the grass pollen season. Here, using targeted, high-throughput sequencing, we demonstrate that all grass genera displayed discrete, temporally restricted peaks of incidence, which varied with latitude and longitude throughout Great Britain, revealing that the taxonomic composition of grass pollen exposure changes substantially across the grass pollen season.

Allergens carried in airborne pollen are associated with both asthma¹ and allergic rhinitis (hay fever), negatively affecting 400 million people worldwide². Pollen from the grass family (Poaceae) constitutes the most important outdoor aeroallergen^{3,4}, and more people are sensitized to grass pollen than to any other pollen type⁵. However, despite the harmful impact of grass pollen on human health, as reported in developed nations, current studies and forecasts categorize grass pollen at the family level (Poaceae)^{6,7} owing to the difficulty of differentiating species based on morphology⁸. Furthermore, although different species of temperate grass flower at different time points^{9,10}, it is unknown whether the disparate phenology of local grass taxa at the ground level is useful for making predictions on the seasonal variation in airborne pollen. Airborne pollen is highly mobile^{11,12} and pollen concentrations often do not directly correlate with local flowering times¹¹. Persistence and mobility of grass pollen could result in a steadily increasing species richness of airborne pollen over the grass pollen season. Conversely, if grass pollen does not persist for an extended time in the air, pollen assemblages should reflect temporal turnover in the composition of species over the summer months. Understanding the taxon-specific phenology of airborne pollen would fill an important gap in our knowledge of allergen triggers and would have associated benefits for healthcare providers, pharmaceutical industries and the public.

Many species within the subfamilies Pooideae, Chloridoideae and Panicoideae release allergenic pollen into the atmosphere⁵,

including *Phleum* spp. (for example, timothy grasses), *Dactylis* spp. (cocksfoot or orchard grasses), *Lolium* spp. (ryegrasses), *Festuca* spp. (fescues), *Poa* spp. (meadow grasses and bluegrasses) and *Anthoxanthum* spp. (vernal grasses). Furthermore, some grass taxa, such as diverse cultivars and hybrids of *Lolium* spp., are widely sown in agricultural grasslands and are likely to contribute disproportionately to airborne pollen. However, it is unknown whether particular grass species, or varieties or cultivars within species, contribute more to the prevalence of allergic symptoms and related diseases than others¹³. Although some grass species have been identified as more allergenic than others in vitro (triggering higher levels of immunoglobulin E (IgE) antibody production), there is a high degree of cross-reactivity between grass species^{14–16}. In addition, the allergen profiles (the characterization of the different allergens that are common to different grass) and the degree of sensitization differ between grass species^{14,17} and the overall allergenicity of grass pollen in the air varies across seasons¹⁸. Therefore, family-level estimates of grass pollen concentrations cannot be considered to be a reliable proxy for either the concentration of pollen-derived aeroallergens or pollen-induced public health outcomes.

The identification of biodiversity through high-throughput analysis of taxonomic marker genes (also termed metabarcoding) provides an emerging solution to semi-quantitatively identify complex mixtures of airborne pollen grains^{19–22}. Furthermore, recent global DNA barcoding initiatives and coordinated regional efforts have now resulted in near-complete genetic databases of national native plants, including grass species. In Great Britain, the vast majority of angiosperms are included in mature DNA barcoding databases for multiple markers²³, meaning that we are now in a position to investigate the aerial composition of pollen during the grass pollen season at a national scale.

Here, using two complementary DNA barcode marker genes (*rbcL* and *ITS2*), we characterize the spatial and temporal distribution of airborne grass pollen throughout the temperate summer grass-pollen season (May–August) across the latitudinal and longitudinal range of Great Britain (Fig. 1). We hypothesize that the composition of airborne grass pollen, from different grass taxa will be broadly homogenous across the grass pollen season, regardless of terrestrial Poaceae phenology and that they are homogenous across

¹Molecular Ecology and Fisheries Genetics Laboratory, School of Natural Sciences, Bangor University, Bangor, UK. ²IBERS, Aberystwyth University, Aberystwyth, UK. ³National Botanic Garden of Wales, Llanarthne, UK. ⁴University of Worcester, Worcester, UK. ⁵University of Exeter, Truro, UK. ⁶University of New South Wales, Sydney, New South Wales, Australia. ⁷School of Public Health, Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia. ⁸Met Office, Exeter, UK. ⁹The Woodland Trust, Kempton Way, Grantham, UK. ¹⁰A full list of authors and their affiliations appears at the end of the paper. *e-mail: g.l.b.doonan@gmail.com; s.creer@bangor.ac.uk

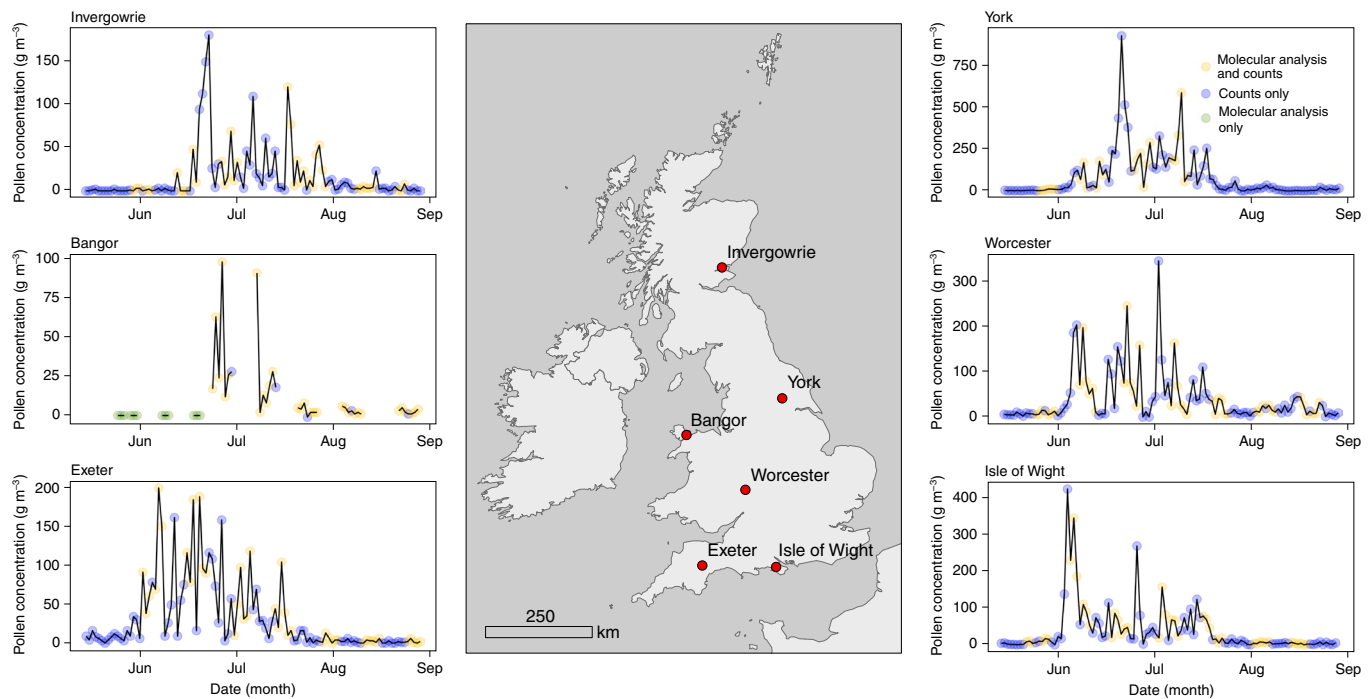


Fig. 1 | Locations where pollen was collected and the temporal concentrations and composition of Poaceae. The map shows the location of the six sampling sites and daily Poaceae pollen concentrations (number of grains per m^3) throughout the grass pollen season (May–August 2016). Yellow filled circles indicate dates when pollen was collected for both observational concentrations and molecular analysis; blue circles indicate days when pollen was collected for observational concentrations only; green circles indicate when pollen was collected for molecular analysis only at the Bangor site. Note that Bangor is not part of the UK pollen-monitoring network, and observational concentrations were only performed alongside pollen collections for molecular analysis between 24 June and 28 August 2016. Contains OS data Crown copyright and database right (2018). Image Crown Copyright, The Met Office (2018).

Great Britain owing to the potential for long-distance transport of windborne pollen grains.

Airborne grass pollen from each genus occupied distinct temporal windows across the grass pollen season in 2016 (May to August), thus we reject the first hypothesis that the composition of airborne grass pollen from different grass taxa is broadly homogenous across the grass pollen season (Fig. 2 and Supplementary Fig. 1). Time, measured as the number of days after the first sample was collected, is a good predictor of the composition of airborne grass pollen taxa using both markers (Fig. 2 and Supplementary Fig. 1; ITS2, likelihood ratio (LR)_{1,74} = 128.8, P = 0.001; *rbcL*, LR _{1,71} = 46.71, P = 0.001). Community-level ordination reveals that the airborne grass pollen community as a whole changed across the grass pollen season (Supplementary Figs. 2, 3); similar overarching trends were observed for the most-abundant airborne pollen families, including Poaceae, Pinaceae and Urticaceae (Supplementary Fig. 4). In addition, observations of the dates of first flowering from a citizen science project (www.naturescalendar.org.uk) and metabarcoding data show similar sequences of seasonal progression with a lag time similar to that found in observational studies¹¹ (see Supplementary Text and Supplementary Fig. 5), suggesting that there is a link between local phenology of Poaceae and the composition of airborne grass pollen.

Focusing on the more taxonomically specific ITS2 marker dataset, *Alopecurus* and *Holcus* typically dominated the early grass pollen season (Fig. 2), which coincides with typical peaks in allergic rhinitis²⁴; however, further research will be required to confirm this association. *Lolium* featured prominently for the majority of the later grass season. The popularity of *Lolium* species as a forage crop means that it is widely sown in agricultural grasslands²⁵, although most agricultural grasslands are managed by grazing, silage cutting or mowing, which prevents the growth of flowering heads²⁵.

The length of time over which *Lolium* pollen dominated may be because many varieties have been bred with the potential to mature at different times throughout the year²⁶, although it should be noted that *Lolium* species frequently hybridize with each other and that it is therefore difficult to distinguish these genera using genetic material alone. Additionally, although there is some evidence that some species of grass appear to be more allergenic than others¹⁸, it is unknown how much allergenicity differs within a species (that is, at the cultivar or hybrid level)¹⁶. Although *Lolium* was the dominant species in airborne grass pollen from July to the end of the sampling period, the total grass pollen concentration declined in August, indicating that the absolute number of *Lolium* pollen grains at this time is low (Fig. 1 and Supplementary Fig. 6).

The top five genera that contributed to airborne pollen, indicated by the relative abundance of taxonomic marker genes, were *Alopecurus*, *Festuca*, *Holcus*, *Lolium* and *Poa* (Fig. 2 and Supplementary Fig. 6). Each of these genera is widespread in the United Kingdom, although long-distance pollen transport means that they may also originate further afield²⁷. These dominant genera have all been shown to provoke IgE-mediated responses in grass-sensitized patients¹⁴, providing candidate species for links with hay fever and asthma exacerbation. Conversely, less prevalent species in the dataset could contribute disproportionately to the allergenic load. Species such as *Phleum pratense* have been identified to be a major allergen^{5,28}. However, we found that *Phleum* made up a very small proportion of metabarcoding reads (Supplementary Fig. 1), corresponding with the results of an earlier phenological study⁹. Most genera, such as *Phleum*, *Anthoxanthum* and *Dactylis*, show distinct and narrow temporal incidences (Supplementary Fig. 1) and could enable researchers to identify the grass species that are associated with allergenic windows with greater accuracy.

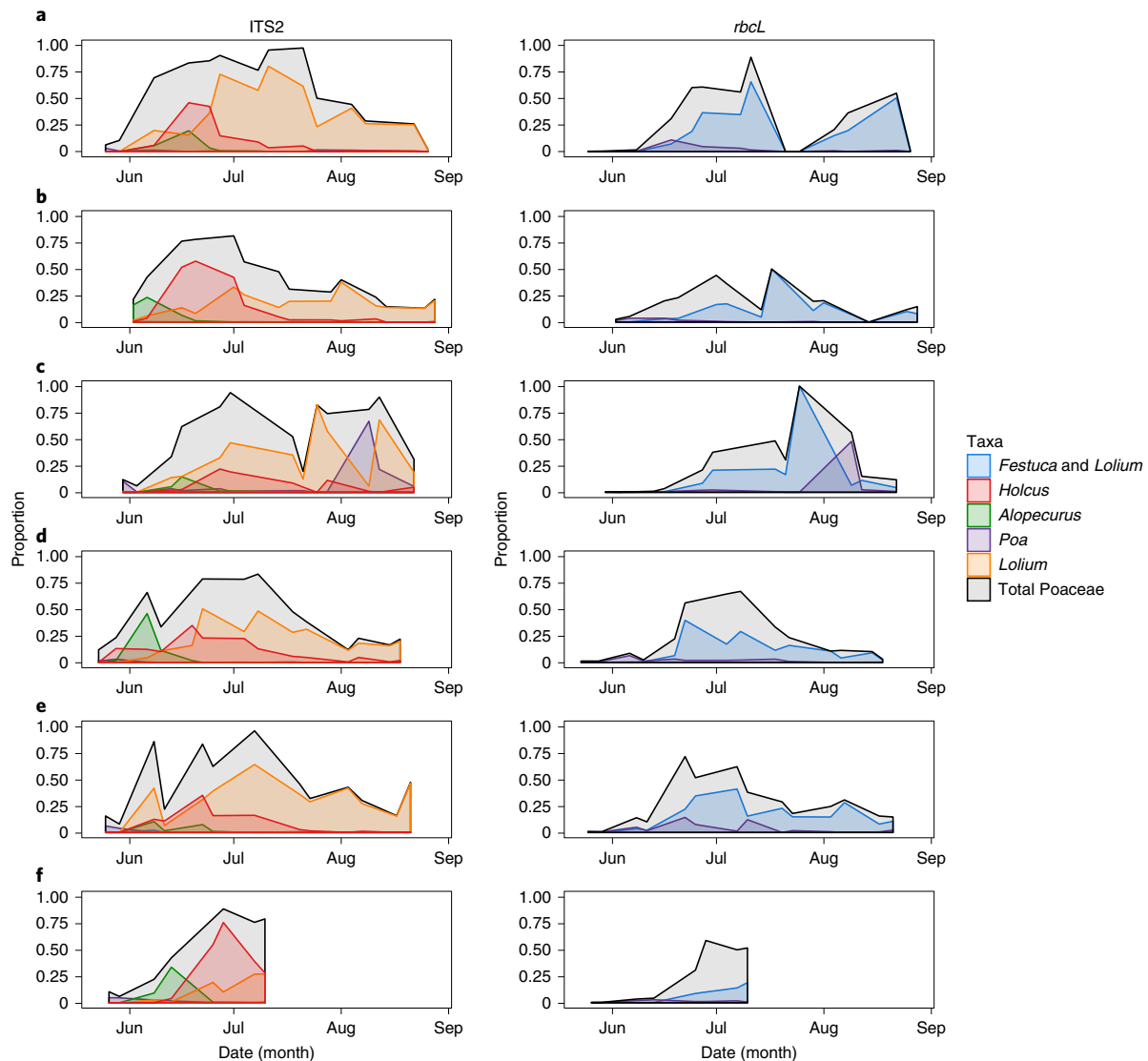


Fig. 2 | Abundance of the most common taxa of airborne grass pollen throughout the grass pollen season. a–f, The five most abundant grass taxa (expressed as proportion of total reads) are depicted alongside the total proportion of reads assigned to family Poaceae using the markers ITS2 (left) and *rbcL* (right). **a,** Bangor. **b,** Exeter. **c,** Invergowrie. **d,** Isle of Wight. **e,** Worcester. **f,** York. Owing to errors in sampling equipment, only four alternate weeks (out of a possible seven alternate weeks) of samples were collected at the York sampling site. A map of sampling locations and daily Poaceae pollen concentrations can be found in Fig. 1.

Changes in species composition over time were localized. We found that peaks in abundance of airborne pollen occurred at different times at each location during the summer (Fig. 2 and Supplementary Fig. 1). For example, the relative abundance of airborne grass pollen from the genus *Poa* peaked in mid-June in Worcester and Bangor, but 6–8 weeks later in Invergowrie (Fig. 2), probably owing to latitudinal effects on flowering time^{27,27}. This is supported by a significant interaction between latitude and time of year for both markers (Fig. 2 and Supplementary Fig. 1; ITS2, $LR_{68,1} = 34.2$, $P = 0.002$; *rbcL*, $LR_{68,1} = 47.36$, $P = 0.001$). Differences in species composition of airborne grass pollen between the six sampling sites is supported by a significant effect of latitude (Fig. 2 and Supplementary Fig. 1; ITS2, $LR_{1,73} = 73.2$, $P = 0.001$; *rbcL*, $LR_{1,70} = 26.4$, $P = 0.025$) and longitude (Fig. 2 and Supplementary Fig. 1; ITS2, $LR_{1,69} = 33$, $P = 0.003$; *rbcL*, $LR_{1,69} = 27.10$, $P = 0.010$), which are proxies for a broad range of environmental variables. These results do not support our second hypothesis that the composition of airborne grass pollen will be homogenous across the United Kingdom, and instead suggest that taxon-specific effects of regional geography, climate and environmental conditions

underpin distributions that have been demonstrated for Poaceae pollen as a whole⁷. Further investigations into the mechanisms of pollen production and transport, and interactions with a range of climatic, seasonal and meteorological effects, will therefore provide valuable future research foci that will help to expand our mechanistic knowledge of the deposition of grass pollen across time and space.

Enabled by contemporary molecular biodiversity assessments and mature, curated DNA barcoding databases, here we provide a taxonomic overview of the distribution of airborne grass pollen, throughout an entire grass pollen season and across geographical scales. The grass pollen season is defined by discrete temporal windows of different grass species and some species showed geographical variation. Temporal pollen distributions in metabarcoding data follow observed flowering times. The data provide an important step towards developing genera-, and in certain cases, species-level grass pollen forecasting. Additionally, the research presented here enables future studies to elucidate the relationships between grass pollen and disease, the understanding of which will have considerable global public health relevance and socioeconomic importance.

Methods

Sampling and experimental design. We collected aerial samples from six sites across Great Britain (Fig. 1 and Supplementary Table 3) using Burkard Automatic Multi-Vial Cyclone Samplers (V2; Burkard Manufacturing) designed to simplify collection of pollen and spores by sampling directly into a microcentrifuge tube²⁹. The volumetric aerial sampler uses a turbine to draw in air (16.5 l min⁻¹) and aerial particles, using mini-cyclone technology. The aerial particles are collected into 1.5-ml sterile microcentrifuge tubes located on a carousel; the carousel is programmed to sample into a new tube every 24 h, thus providing daily samples of airborne pollen (Supplementary Fig. 7). Sample tubes were sent to Bangor and stored at -20 °C before processing. Each sampling unit was mounted alongside a seven-day volumetric trap of the Hirst design³⁰ (Burkard Manufacturing) belonging to the Met Office UK Pollen Monitoring Network, which provided daily pollen concentrations (Fig. 1; map produced using ArcGIS). In the seven-day volumetric trap, a turbine draws air in (10 l min⁻¹) and particles are impacted upon an adhesive-coated tape carried on a clockwork-driven drum. The tape is cut into 24-h sections and mounted on glass slides using a gelatine-glycerol mountant containing basic fuchsin to stain the pollen grains. Pollen were identified and counted under a microscope and converted to volumetric concentrations⁷. Although the high cost of the pollen samplers precluded routine replicate sampling, our methodologies mirror methodologies that have been used for several decades in the UK network^{31,32} and are in agreement with recommended terminology described previously³³. All pollen samplers were situated in elevated positions on flat-roofed buildings between 4 to 6 floors in height to sample from a mixed air flow. Fins on the samplers (both Burkard Multi-Vial Cyclone and Hirst-type seven-day volumetric samplers) direct the cyclone inlet port into the wind. Bangor was the only sampling site that was not part of the pollen monitoring network, but the same methodology for observational concentrations and molecular analysis of pollen was also used at the Bangor site (which began on 24 June 2016, Fig. 1).

Sampling began in late May 2016 and during alternate weeks, aerial samples were collected for seven days for a total of seven weeks between 25 May and 28 August. Exact sampling dates varied slightly between sites and a total of 279 aerial samples were collected (Supplementary Table 4).

DNA extraction, PCR and sequencing. From the 279 daily aerial samples, 231 were selected for downstream molecular analysis, as described below. Within each sampling week, two series of three consecutive days were pooled. Pooled samples were selected based on grass pollen concentrations on the basis of microscopy analyses. The final, unselected day was not used in downstream molecular analysis. In total, 77 pools of DNA were created. In one instance, three consecutive days of pollen samples were unavailable (Invergowie, week 2, pool 2) owing to trap errors. For this sample, the next sampling day was selected for pooling (Supplementary Table 4). DNA was extracted from daily samples using a DNeasy Plant Mini kit (Qiagen), with some modifications to the standard protocol that has been previously described³⁴. DNA from daily samples was pooled and eluted into 60 µl of elution buffer at the binding stage of the DNeasy Plant Mini kit.

Illumina MiSeq paired-end indexed amplicon libraries were prepared following a two-step protocol. Two marker genes were amplified with universal primer pairs *RBCLaf* and *RBCLr506*^{23,35}, and *ITS2* and *ITS3*¹⁸ (Supplementary Table 5). A 5' universal tail was added to the forward and reverse primers and a 6N sequence was added between the forward universal tail and the template-specific primer; this is known to improve clustering and cluster detection on MiSeq sequencing platforms³⁶ (Integrated DNA Technologies). The first round of PCR was carried out in a final volume of 25 µl, including forward and reverse primers (0.2 µM), 1× Q5 HS High-Fidelity Master Mix (New England Biolabs) and 1 µl of template DNA. Thermal cycling conditions were an initial denaturation step at 98 °C for 30 s; 35 cycles of 98 °C for 10 s, 50 °C for 30 s, 72 °C for 30 s; and a final annealing step of 72 °C for 5 min. Products from the first PCR were purified using Agencourt AMPure XP beads (Beckman Coulter) with a 1:0.6 ratio of product:AMPure XP beads.

The second round of PCR added the unique identical i5 and i7 indexes and the P5 and P7 Illumina adaptors, along with universal tails complementary to the universal tails used in the first round of PCR (Supplementary Tables 4, 5) (Ultramer, Integrated DNA Technologies). The second round of PCR was carried out in a final volume of 25 µl, including forward and reverse index primers (0.2 µM), 1× Q5 HS High-Fidelity Master Mix (New England Biolabs) and 5 µl of purified PCR product. Thermal cycling conditions were: 98 °C for 3 min; 98 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s (10 cycles); 72 °C for 5 min, 4 °C for 10 min. Both PCRs were run in triplicate. The same set of unique indices were added to the triplicates, which were then pooled following visual inspection on an agarose gel (1.5%) to ensure that indices were added successfully. Pooled metabarcoding libraries were cleaned a second time using Agencourt AMPure magnetic bead purification, run on an agarose gel (1.5%) and quantified using the Qubit high-sensitivity kit (Thermo Fisher Scientific). Positive and negative controls were amplified in triplicate with both primer pairs and sequenced alongside airborne plant community DNA samples using the MiSeq. Sequence data, including metadata, are available at the Sequence Read Archive (SRA) using the project accession number SUB4136142.

Bioinformatics analysis. Initial sequence processing was carried out following a modified version of the workflow described previously³⁷. In brief, raw sequences were trimmed using Trimmomatic v.0.33³⁸ to remove short reads (<200 bp), adaptors and low-quality regions. Reads were merged using FLASH v.1.2.11^{37,39} and merged reads shorter than 450 bp were excluded. Identical reads were merged using fastx-toolkit (v.0.0.14) and reads were split into *ITS2* and *rbcl* on the basis of the primer sequences.

To prevent spurious BLAST hits, custom reference databases that contained *rbcl* and *ITS2* sequences from UK plant species were generated. Although all native species of the UK have been DNA-barcoded²³, a list of all species found in the United Kingdom was generated to gain coverage of non-native species. A list of UK plant species was generated by combining lists of native and alien species⁴⁰ with a list of cultivated plants obtained from Botanic Gardens Conservation International, which represented horticultural species. All available *rbcl* and *ITS2* records were downloaded from NCBI GenBank, and sequences belonging to UK species were extracted using the script 'creatingselctedfastadatabase.py', archived on GitHub.

Metabarcoding data was searched against the relevant sequence database using blastn⁴¹, using the script 'blast_with_ncbi.py'. The top 20 blast hits (identified using the highest bit-score) were tabulated (using 'blast_summary.py'), then manually filtered to limit results to species currently present in Great Britain. Reads occurring fewer than four times were excluded from further analysis. All scripts used are archived on GitHub: <https://doi.org/10.5281/zenodo.1305767>.

Statistical analysis. To understand how the composition of grass pollen changed across space and time, the effect of time (measured as the number of days after the first sampling date), and latitude and longitude of sampling location were included in a two-tailed generalized linear model using the 'manyglm' function in the package 'mvabund'⁴². The proportion of sequences was set as the response variable; proportion data was used as this has been shown to be an effective way of controlling for differences in read numbers⁴³. The effect of time, latitude, longitude and the interaction between time and latitude were included as explanatory variables in the models to test our hypotheses. The effect of longitude is also consistent when York, the most easterly sampling site, with missing data from mid-July until the end of the sampling period, is removed from the analysis (Supplementary Table 6).

The data best fit a negative binomial distribution, most likely owing to the large number of zeros (zeros indicate that a grass genus is absent from a sample), resulting in a strong mean-variance relationship in the data (Supplementary Fig. 8). The proportion of sequences was scaled by 1,000 and values were converted to integers so that a generalized linear model with a negative binomial distribution could be used. Overfitting of the models was tested using 'dropterm' in R and based on the lowest Akaike information criterion score, no terms were removed from the models. In addition, the appropriateness of the models was checked by visual inspection of the residuals against predicted values from the models (Supplementary Fig. 9).

In order to compare the metabarcoding data with flowering time data, we used phenological records of first flowering collected in 2016 by citizen scientists from the UK's Nature's Calendar (www.naturescalendar.org.uk). First flowering time was compared to genus-level *ITS2* metabarcoding data for three species: *Alopecurus pratensis*, *Dactylis glomerata* and *Holcus lanatus*. Because grass pollen could only be reliably identified to genus level in the metabarcoding data, the taxa compared may not have been exactly equivalent since both *Alopecurus* and *Holcus* contain other widespread species within the United Kingdom. However, *A. pratensis* and *H. lanatus* are the most abundant species within their respective genera. The comparison was only carried out for *ITS2* data, because two of the three genera were not identified by the *rbcl* marker.

Non-metric multidimensional scaling (NMDS) ordination was carried out using package 'vegan' in R⁴⁴, based on the proportion of total high-quality reads contributed by each grass genus, using Bray-Curtis dissimilarity (Supplementary Figs. 2, 3). Ordination is used to reduce multivariate datasets (for example, abundances of many species) into fewer variables that reflect overall similarities between samples. A linear model was carried out using the 'lm' function within the 'stats' package in R to investigate the relationship between the number of reads obtained for each genus using the *rbcl* and *ITS2* marker.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All sequence data are available at the Sequence Read Archive (SRA) using the project accession number SUB4136142. Archived sequence data were used to generate Fig. 2 and Supplementary Figs. 1–6, 8–10). Data on first flowering dates used in Supplementary Fig. 5 were obtained from Nature's Calendar, Woodland Trust and are available upon request. The sequence analysis pipeline is available at <https://github.com/colford/nbgw-plant-illumina-pipeline>.

Received: 5 October 2018; Accepted: 18 February 2019;

Published online: 08 April 2019

References

- Blackley, C. H. *Experimental Researches on the Causes and Nature of Catarrhus Aëstivus (hay-fever Or Hay-asthma)* (Bailliere, 1873).
- Marks, G., Pearce, N., Strachan, D. & Asher, I. *The Global Asthma Report 2014* 16–21 (Global Asthma Network, 2014).
- Bauchau, V. *Eur. Respir. J.* **24**, 758–764 (2004).
- Bousquet, P.-J. et al. *Allergy* **62**, 301–309 (2007).
- García-Mozo, H. *Allergy* **72**, 1849–1858 (2017).
- Emberlin, J. et al. *Aerobiologia* **16**, 373–379 (2000).
- Emberlin, J. et al. *Grana* **33**, 94–99 (1994).
- Mander, L., Li, M., Mio, W., Fowlkes, C. C. & Punyasena, S. W. *Proc. R. Soc. Lond. B* **280**, 20131905 (2013).
- Kmenta, M. et al. *World Allergy Organ. J.* **10**, 31 (2017).
- Cope, C. & Gray, A. *Grasses of the British Isles* (Botanical Society of the British Isles, 2009).
- Estrella, N., Menzel, A., Krämer, U. & Behrendt, H. *Int. J. Biometeorol.* **51**, 49–59 (2006).
- Skjøth, C. A., Sommer, J., Stach, A., Smith, M. & Brandt, J. *Clin. Exp. Allergy* **37**, 1204–1212 (2008).
- McInnes, R. N. et al. *Sci. Total Environ.* **599–600**, 483–499 (2017).
- van Ree, R., van Leeuwen, W. A. & Aalberse, R. C. *J. Allergy Clin. Immunol.* **102**, 184–190 (1998).
- Petersen, A. et al. *J. Allergy Clin. Immunol.* **107**, 856–862 (2001).
- Jung, S. et al. *PLoS ONE* **13**, 1–12 (2018).
- Moingeon, P., Peltre, G. & Bergmann, K.-C. *Clin. Exp. Allergy Rev.* **8**, 12–14 (2008).
- de Weger, L. A. et al. *Clin. Transl. Allergy* **1**, 18 (2011).
- Kraaijeveld, K. et al. *Mol. Ecol. Resour.* **15**, 8–16 (2015).
- Korpelainen, H. & Pietiläinen, M. *Nord. J. Bot.* **35**, 602–608 (2017).
- Deiner, K. et al. *Mol. Ecol.* **26**, 5872–5895 (2017).
- Creer, S. et al. *Methods Ecol. Evol.* **7**, 1008–1018 (2016).
- de Vere, N. et al. *PLoS ONE* **7**, e37945 (2012).
- England, P. H. *GP In-Hours Consultations Bulletin: 25 August 2016 Week 33* (PHE Real-time Syndromic Surveillance Team, 2016).
- Frame, J. & Laidlaw, A. S. *Improved Grassland Management* (The Crowood Press, 2011).
- RGCL. *Recommended Grass and Clover Lists* (British Grassland Society, 2017).
- Rousseau, D.-D. et al. *Geophys. Res. Lett.* **30**, 1765 (2003).
- D'Amato, G. et al. *Allergy* **62**, 976–990 (2007).
- West, J. S. & Kimber, R. B. E. *Ann. Appl. Biol.* **166**, 4–17 (2015).
- Hirst, J. M. *Ann. Appl. Biol.* **39**, 257–265 (1952).
- Adams-Groom, B., Emberlin, J., Corden, J., Millington, W. & Mullins, J. *Aerobiologia* **18**, 117–123 (2002).
- Skjøth, C. A., Baker, P., Sadyś, M. & Adams-Groom, B. *Urban Climate* **14**, 414–428 (2015).
- Galán, C. et al. *Aerobiologia* **33**, 293–295 (2017).
- Hawkins, J. et al. *PLoS ONE* **10**, e0134735 (2015).
- Kress, W. J. & Erickson, D. L. *PLoS ONE* **2**, e508 (2007).
- Miya, M. et al. *R. Soc. Open Sci.* **2**, 150088 (2015).
- de Vere, N. et al. *Sci. Rep.* **7**, 42838 (2017).
- Bolger, A. M., Lohse, M. & Usadel, B. *Bioinformatics* **30**, 2114–2120 (2014).
- Magoč, T. & Salzberg, S. L. *Bioinformatics* **27**, 2957–2963 (2011).
- Stace, C. *New Flora of the British Isles* (Cambridge Univ. Press, 2010).
- Camacho, C. et al. *BMC Bioinformatics* **10**, 421 (2008).
- Wang, Y., Naumann, U., Wright, S. T. & Warton, D. I. *Methods Ecol. Evol.* **3**, 471–474 (2012).
- McMurdie, P. J. & Holmes, S. *PLoS Comput. Biol.* **10**, e1003531 (2014).
- Dixon, P. J. *J. Veg. Sci.* **14**, 927–930 (2003).

Acknowledgements

We thank J. Kenny, P. Koldkjær, R. Gregory and A. Lucaci for sequencing support; J. Winn for ArcGIS assistance with Fig. 1; W. Grail and the technical support staff at Bangor University; the Botanic Gardens Conservation International (BGCI) for access to the list of plant collections in the National Gardens in the United Kingdom and Ireland; the Met Office network for providing additional observational grass pollen count data; the Woodland Trust and the Centre for Ecology & Hydrology for supplying the UK Phenology Network data and the citizen scientists who have contributed to the latter scheme. We acknowledge the computational services and support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) via Welsh Government. This work was supported by the Natural Environment Research Council (<https://nerc.ukri.org/>), awarded to S.C. (NE/N003756/1), C.A.S. (NE/N002431/1), N.J.O. (NE/N002105/1), and G.W.G., N.d.V. and M.H. (NE/N001710/1). IBERS Aberystwyth receives strategic funding from the BBSRC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

S.C., N.d.V., G.W.G., R.N.M., N.J.O., C.A.S., Y.C., B.W.W. and G.L.B. conceived and designed the study; B.A.-G., G.L.B., G.M.P., A.E., R.N., S.P., K.S. and N.S. collected samples and counted pollen; G.L.B. performed laboratory work, supported by S.C.; N.d.V., C.P., C.R.F., L.J., G.L.B. and S.C. contributed methods; C.A. and D.B.R. contributed materials; C.P. and G.L.B. analysed the data; and G.L.B., C.P. and S.C. produced the first draft of the manuscript. All authors contributed substantially to the final submitted manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41559-019-0849-7>.

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

Correspondence and requests for materials should be addressed to G.L.B. or S.C.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

The PollerGEN Consortium

Angela Elliot¹¹, Carl A. Frisk⁴, Roy Neilson¹², Stephen Potter¹³, Abdullah M. Rafiq¹, David B. Roy¹⁴, Katherine Selby¹⁵ and Natascha Steinberg¹¹

¹¹College of Life and Environmental Science, University of Exeter, Exeter, UK. ¹²Ecological Sciences, The James Hutton Institute, Dundee, UK. ¹³The David Hide Asthma & Allergy Research Centre, St Mary's Hospital, Newport, UK. ¹⁴Centre for Ecology & Hydrology, Crowmarsh Gifford, Wallingford, UK.

¹⁵Environment Department, University of York, York, UK.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☒ ☐ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

The MiSeq platform was used to generate DNA sequences from amplicon libraries. This was carried out at Liverpool Genomics, UK (<https://www.liverpool.ac.uk/genomic-research/technologies/next-generation-sequencing/>). The raw, unfiltered sequences were used in downstream analysis.

Data analysis

Quality filtering of sequence data and merging of pair-end reads was carried out in the Linux environment using a pipeline which is accessible on GitHub. All available rbcl and ITS2 records were downloaded from NCBI Genbank, and sequences belonging to UK species were extracted using the script 'creatingselectedfastadatabase.py', archived on GitHub (<https://doi.org/10.5281/zenodo.1305767>). Sequence data, including metadata, are available at the Sequence Read Archive (SRA), using the project accession number SUB4136142. This information can also be found in the supplementary information. All statistical analysis of high quality reads were carried out in the R environment and details of functions and packages used are found in the methods section.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All sequence data are available at the Sequence Read Archive (SRA) using the project accession number SUB4136142. Archived sequence data was used to generate Figure 2 Figures S1-S6 and S8-S10). First flowering data used in Figure S5 was obtained from Nature's Calendar, Woodland Trust and is available upon request. The sequence analysis pipeline is available at <https://github.com/colford/nbgw-plant-illumina-pipeline>.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

This study tracked the seasonal progression of airborne grass pollen, in time and space. To do this we collected aerial samples from six sites across Britain. During alternate weeks, aerial samples were collected continuously every 24h, for seven days for a total of seven weeks. A total of 279 daily aerial samples were collected. Within each sampling week, two series of three consecutive daily samples were pooled. In total, seventy-seven pools of DNA were created and used in downstream analysis. Statistical models were used to test the effect of sampling location (latitude and longitude), sampling timepoint (number of days after the first sampling date and month) and their interactions. Statistical analysis are described in further detail in the methods section.

Research sample

The research samples in this study were aerial particles, including pollen, collected using Burkard Automatic Multi-Vial Cyclone Samplers (Burkard Manufacturing Co. Ltd. Rickmansworth, UK). The volumetric aerial sampler uses a turbine to draw in air (16.5 litres/min) and aerial particles are collected, using mini-cyclone technology into 1.5 ml sterile Eppendorf vials located on a carousel. Only aerial particles originating from plants were targeted for downstream analysis using two complementary plant DNA barcode marker genes (rbcl and ITS2).

Sampling strategy

Samples were collected for 24 h, following the same strategy employed by the UK Met Office Pollen Monitoring Network. Using this method, we obtained a representative sample of the daily airborne pollen composition. For this study, 24 h samples were collected during alternate weeks for seven days, for a total of seven weeks. From the 279 daily aerial samples, 231 were selected for pooling and downstream molecular analysis. Within each sampling week, two series of three consecutive days were pooled. Pooled samples were selected based on grass pollen counts obtained by microscopy. Pooling was performed in this way to ensure that enough material was available for DNA extractions. The final, unselected, day was not used in downstream molecular analysis. In total, seventy-seven pools of DNA were created.

Data collection

Samples were collected using Burkard Automatic Multi-Vial Cyclone Samplers (Burkard Manufacturing Co. Ltd. Rickmansworth, UK). Pollen collected for daily pollen counts were collected using a seven-day volumetric trap (Burkard Manufacturing Co. Ltd. Rickmansworth, UK). First flowering data was obtained from Nature's calendar (www.naturescalendar.org.uk). Co-authors which contributed to sample collections in the field are detailed in the author contribution section of the manuscript as follows: B.A-G., G.L.B., G.P., A.E., R.N., S.P., K.S., and N.S.

Timing and spatial scale

Samples were collected at six sites across the Great Britain. The sites were selected to cover broad geographic areas (including heterogeneous soil types and land use types across the Great Britain) (Fig. 1). Locations were selected based on the Met Office UK Pollen Monitoring Network. Each sampling unit was mounted alongside a seven-day volumetric trap (Burkard Manufacturing Co. Ltd. Rickmansworth, UK) belonging to the Met Office UK Pollen Monitoring Network, which provided daily pollen count data. Bangor (Fig. 1) was the only sampling site which was not part of the pollen monitoring network, but we deployed the same pollen count methodology at the Bangor site. Samples were collected during the grass pollen season, informed by the pollen counts provided by Met Office UK Pollen Monitoring Network. Sampling began in late May 2016 (Table S4) and during alternate weeks, aerial samples were collected for seven days for a total of seven weeks between 25th May and 28th August. Exact sampling dates varied slightly between sites (Table S4).

Data exclusions

Following DNA sequencing, raw sequences were trimmed to remove short reads (<200bp), adaptors and low-quality regions to avoid misidentification during taxonomic assignment. Paired end reads were merged and merged reads shorter than 450bp were excluded. This is common practice as the presence of high length variation could influence the clustering process (identical reads are merged) when creating an operational taxonomic unit (OTU) table. Identical reads were merged and reads were split into ITS2 and rbcl based on primer sequences. The top twenty blast hits were tabulated, then manually filtered to limit results to species currently present in Great Britain. Reads occurring fewer than four times were excluded from further analysis to avoid incorporation of background sequencing errors. Since grasses are the focus of the study, following taxonomic assignment, only data obtained for Poaceae were

Reproducibility	This study involved sampling airborne pollen in order to establish patterns in species composition in time and space. To achieve this, we collected daily pollen samples for 7 weeks, across six locations. For each daily sample, the air was sampled continuously for 24 hours into a single microcentrifuge tube and we analyzed the entire contents of the sample. Consequently, replication of the biological sample is not applicable to our study design. However, the reliability of our findings was evaluated based on the positive and negative controls.
Randomization	Aerial samples shipped to Bangor were randomized for DNA extractions. DNA extracts were randomised on a single 96-well plate in order to prepare amplicon libraries for ITS2 and rbcL barcode markers. Finally, all amplicon libraries were pooled at equal concentration and sequenced on a single MiSeq sequencing run.
Blinding	This study focuses on measuring how pollen composition changes in space and time in the natural environment and as such there are no control or treatment groups. The typical use of “blinding” observers to the identity of treatment and control groups was not applicable in our study.
Did the study involve field work?	☒ Yes ☐ No

Field work, collection and transport

Field conditions	Samples were collected during the grass pollen season between 25th May and 28th August 2016. See Table S4 for exact dates. The sampling sites are located on exposed roof tops between 4 to 6 floors in height.
Location	Latitude and longitude of the sampling locations are given in Supplementary Table S3. Sampling location were selected based on the UK Met Office Pollen Monitoring Network requirements. The sampling sites are located on exposed roof tops, between 4 to 6 floors high, in order to sample from a mixed air flow. The site needs to be well exposed and not sheltered for example, by parapets or adjacent taller buildings. Sampling equipment also needs to be away from ventilation outlets/ inlets or anything else which could influence the catch.
Access and import/export	The sampling sites are located on roof tops and permission to access roof tops was granted by the building owner/manager. Aerial pollen samples were shipped securely to Bangor University for downstream analysis. No permit was required.
Disturbance	No disturbance was caused during sampling. Samples were collected from the air using samplers powered by batteries and solar panels or plugged directly into the mains.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involvement in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involvement in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Nature Research, brought to you courtesy of Springer Nature Limited (“Nature Research”)

Terms and Conditions

Nature Research supports a reasonable amount of sharing of content by authors, subscribers and authorised or authenticated users (“Users”), for small-scale personal, non-commercial use provided that you respect and maintain all copyright, trade and service marks and other proprietary notices. By accessing, viewing or using the nature content you agree to these terms of use (“Terms”). For these purposes, Nature Research considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). By sharing, or receiving the content from a shared source, Users agree to be bound by these Terms.

We collect and use personal data to provide access to the nature content. ResearchGate may also use these personal data internally within ResearchGate and share it with Nature Research, in an anonymised way, for purposes of tracking, analysis and reporting. Nature Research will not otherwise disclose your personal data unless we have your permission as detailed in the Privacy Policy.

Users and the recipients of the nature content may not:

1. use the nature content for the purpose of providing other users with access to content on a regular or large scale basis or as a means to circumvent access control;
2. use the nature content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by either Nature Research or ResearchGate in writing;
4. use bots or other automated methods to access the nature content or redirect messages; or
5. override any security feature or exclusionary protocol.

These terms of use are reviewed regularly and may be amended at any time. We are not obligated to publish any information or content and may remove it or features or functionality at our sole discretion, at any time with or without notice. We may revoke this licence to you at any time and remove access to any copies of the shared content which have been saved.

Sharing of the nature content may not be done in order to create substitute for our own products or services or a systematic database of our content. Furthermore, we do not allow the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Nature content cannot be used for inter-library loans and librarians may not upload nature content on a large scale into their, or any other, institutional repository.

To the fullest extent permitted by law Nature Research makes no warranties, representations or guarantees to Users, either express or implied with respect to the nature content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Nature Research that we license from third parties.

If you intend to distribute our content to a wider audience on a regular basis or in any other manner not expressly permitted by these Terms please contact us at

onlineservice@springernature.com

The Nature trademark is a registered trademark of Springer Nature Limited.

In the format provided by the authors and unedited.

Temperate airborne grass pollen defined by spatio-temporal shifts in community composition

Georgina L. Brennan¹^{*}, Caitlin Potter², Natasha de Vere^{2,3}, Gareth W. Griffith²^{*}, Carsten A. Skjøth⁴, Nicholas J. Osborne^{5,6,7}, Benedict W. Wheeler⁵, Rachel N. McInnes⁸, Yolanda Clewlow⁸, Adam Barber⁸, Helen M. Hanlon⁸, Matthew Hegarty², Laura Jones^{1,3}, Alexander Kurganskiy⁴, Francis M. Rowney⁵, Charlotte Armitage⁹, Beverley Adams-Groom⁴, Col R. Ford³, Geoff M. Petch⁴, The PollerGEN Consortium¹⁰ and Simon Creer¹^{*}

¹Molecular Ecology and Fisheries Genetics Laboratory, School of Natural Sciences, Bangor University, Bangor, UK. ²IBERS, Aberystwyth University, Aberystwyth, UK. ³National Botanic Garden of Wales, Llanarthne, UK. ⁴University of Worcester, Worcester, UK. ⁵University of Exeter, Truro, UK. ⁶University of New South Wales, Sydney, New South Wales, Australia. ⁷School of Public Health, Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia. ⁸Met Office, Exeter, UK. ⁹The Woodland Trust, Kempton Way, Grantham, UK. ¹⁰A full list of authors and their affiliations appears at the end of the paper. *e-mail: g.l.b.doonan@gmail.com; s.creer@bangor.ac.uk

Supplementary information

Title: Temperate grass flowering season defined by spatio-temporal shifts in airborne pollen communities

Authors: Georgina L. Brennan^{1*}, Caitlin Potter², Natasha de Vere^{2,3}, Gareth W. Griffith², Carsten A. Skjøth⁴, Nicholas J. Osborne^{5,6}, Benedict W. Wheeler⁵, Rachel N. McInnes⁷, Yolanda Clewlow⁷, Adam Barber⁷, Helen M. Hanlon⁷, Matthew Hegarty², Laura Jones^{1,3}, Alexander Kurganskiy⁴, Francis M. Rowney⁵, Charlotte Armitage⁸, Beverley Adams-Groom⁴, Col R. Ford³, Geoff M. Petch⁴, The PollerGEN Consortium, and Simon Creer^{1*}.

Consortium Angela Elliot⁹, Carl A. Frisk⁴, Roy Neilson¹⁰, Stephen Potter¹¹, Abdullah M. Rafiq¹, David, B. Roy¹², Katherine Selby¹³, Natascha Steinberg⁹

Corresponding authors: Simon Creer (s.creer@bangor.ac.uk) and Georgina L. Brennan (g.l.b.doonan@gmail.com)

Supplementary text

Observations of first flowering dates from a citizen science project (UKPN; www.naturescalendar.org.uk) and metabarcoding data show similar sequences of seasonal progression (Supplementary Figure 5). First flowering dates of each genus started almost 3-4 weeks prior to the observation of peaks of grass pollen in the metabarcoding data (Supplementary Figure 5). Pollen release (anthesis) occurs approximately 2-3 weeks after the production of flowering heads (heading)¹, and this is reflected in the metabarcoding data suggesting that local flowering data are informative for predicting the composition of airborne pollen. At the family level, the transition of the dominant airborne pollen from Pinaceae, to Poaceae, to Urticaceae reflects the flowering and pollinating season of their contributing flora, providing further confidence that the airborne community is represented by the phenology of local flora at the ground level (Supplementary Figure 4). This study substantially extends research into the relationship between airborne pollen and terrestrial phenology. Previous studies have suggested that patterns in airborne pollen concentrations are related to phenological behaviour of dominant species² and others have demonstrated that strong correlations exist between local phenology and airborne pollen concentrations (e.g.^{3,4}). However, for these previous studies, it was not possible to quantify the contribution to the daily grass pollen concentrations from individual genera as optical recognition of grass pollen using microscopy only allow for a detection at the family level. Continuing this study over multiple years would allow us to track long-term, phenological changes in airborne pollen communities and improve our ability to forecast the seasonal progression of airborne pollen⁵.

Sequencing statistics

In total, 460,981 high-quality sequences were assigned to ITS2 and 1,054,333 to *rbcl*. There was considerable variation in the number of reads per sample, with the total number of high quality *rbcl* sequences varying from 3 to 56,981 and the number of ITS2 sequences ranging

from 260 to 10,180 (excluding a single sample that contained zero sequences following quality control).

Excluding control samples, ITS2 reads included 179,450 grass sequences (Poaceae), assigned to 13 genera. Whilst not the focus of our study, the remaining 162,540 sequences consisted of 33 families and 31 genera of terrestrial plants. Of these families, 17 contained only a single genus and four contained reads which could not be identified confidently to genus level. Within the *rbcL* marker, 179,330 grass reads were assigned to 13 grass genera and the remaining 867,823 reads belonged to 68 families and 84 genera of terrestrial plants. Of these families, 33 contained only a single genus and 13 contained reads which could not be identified confidently to genus level. The top three most abundant pollen belonged to the families Poaceae, Pinaceae and Urticaceae, accounting for 82% and 90% of total reads, for *rbcL* and ITS2 respectively (Supplementary Figure 2). Poaceae was the dominant pollen during the central part of the grass pollen season (June and July) (Supplementary Figure 2), where the grass pollen season is identified using the 95% method. Using this method, the pollen season is defined using the period of time when 2.5% and 97.5% of the yearly total airborne pollen is collected⁶. This approach is generally used in UK (e.g.⁷) and many other European countries. Pinaceae was the dominant airborne pollen during late spring/early summer (May to June or early July at some locations), a period typically characterised by the late tree pollen season. Urticaceae was the dominant airborne pollen during the end of the summer (late July to August), indicating the onset of the weed pollen season.

ITS2 and *rbcL* detect different grass species

The contrasting characteristics of the ITS2 and *rbcL* markers makes them an ideal pairing. The ITS2 marker shows high specificity between species but cannot detect all plants⁸, whereas *rbcL* primers are highly universal but the marker shows lower resolution between closely related plants⁹. For example, we found that the Pinaceae pollen season is longer when using *rbcL* compared with ITS2 (Supplementary Figure 4), and it is likely that *rbcL* is picking up a broader range of species within each family, compared to ITS2.

72

73 Of the grass genera identified, only four were present in both ITS2 and *rbcL* datasets:
74 *Dactylis*, *Lolium/Festuca*, *Anthoxanthum*, and *Avena*. While the proportion of reads assigned
75 to *Lolium/Festuca* and *Anthoxanthum* were correlated between the two markers
76 (*Lolium/Festuca*: $t_{72}=8.6$, adjusted p-value < 0.001, $r^2 = 0.5$, Supplementary Figure 10A;
77 *Anthoxanthum*: $t_{72}=2.9$, adjusted p-value = 0.006, $r^2 = 0.09$, Supplementary Figure 10B), this
78 was not the case for *Dactylis* or *Avena* (Supplementary Figure 10C; Supplementary Figure
79 10D). However, both of the latter species were detected at relatively low levels in both
80 datasets, potentially increasing the degree of stochasticity introduced by library
81 preparation¹⁰.

82

83 **Positive and negative controls**

84 Negative controls, with all reagents and no DNA were used to identify any cross-
85 contamination. Of the six negative controls, four contained no reads following quality control
86 filtering, one contained a single read and one contained nine reads in the *rbcL* database.
87 None of the negative controls in the ITS2 dataset contained any reads.

88

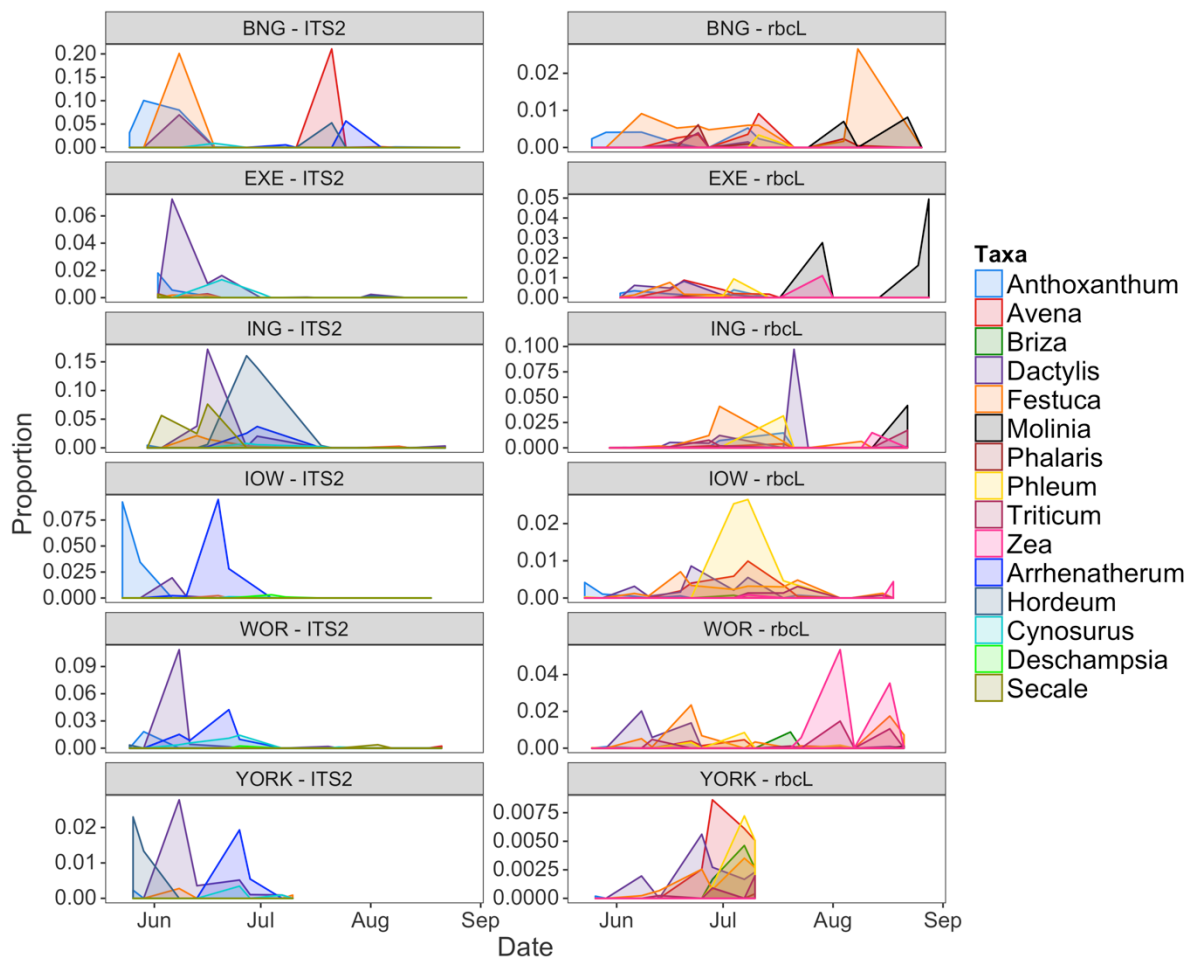
89 Two positive control samples were also included, a grass positive control (Supplementary
90 Table 1) and an exotic plant positive control (list S1). Both sets of positive controls were
91 diluted to 0.3 ng μl^{-1} , similar to that of the aerial eDNA samples.

92

93 The grass positive control contained a mixture of fifty-two species of grass from herbarium
94 collections held at the National Botanic Garden of Wales. The mixture of grasses contained
95 thirty-three genera, with twenty-four of these genera represented by a single species and the
96 remaining nine genera represented by between two and five species (S1 Table). Of the
97 thirty-three genera in the grass positive control sample, four were detected across both
98 markers, eight were detected by the *rbcL* marker and twelve by the ITS2 marker. The
99 remaining sequences were too similar to be identified to genus level (27% and 37% of reads
100 in the grass positive control samples could not be reliably assigned to genus level, using ITS2

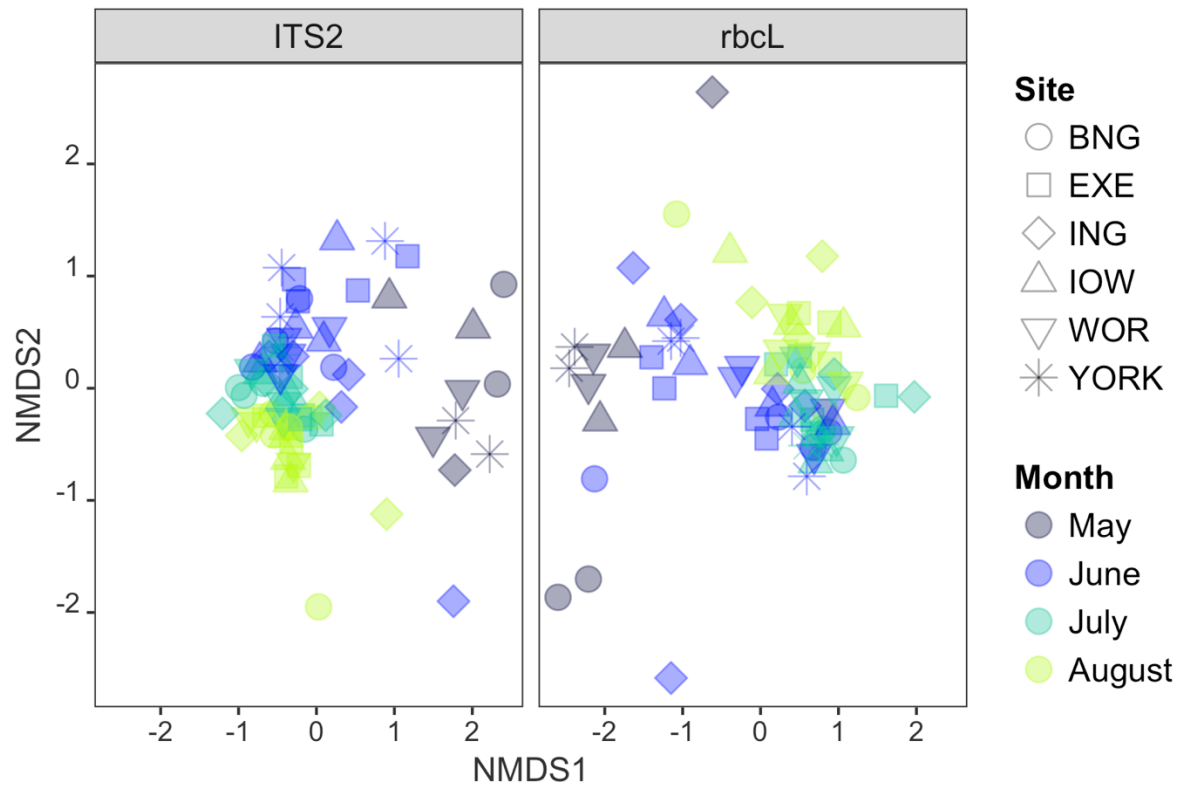
and *rbcL* markers respectively). However, three of the genera not detected in the positive control samples, despite being included, were detected in airborne samples (*Agrostis*, *Anthoxanthum*, *Alopecurus*), likely reflecting higher local abundances of airborne pollen (Supplementary Table 2). The number of species in the grass positive control is much higher than the number of species predicted to contribute to airborne pollen concentrations according to phenological studies^{11,12}. Differences in taxon diversity between the grass positive control and the airborne samples will likely lead to differences in taxonomic assignment due to taxon-specific PCR amplification biases^{13–15}. While sample coverage (i.e. number of reads) obtained for the grass positive control samples was comparable to the airborne samples, the high diversity of the positive control and variation in the number of species between genera may have led to a higher likelihood of amplification for certain genera.

In order to check for cross-contamination between samples, an exotic plant positive control sample was used containing DNA extracted from twenty-one tropical tree species samples held at the National Botanic Garden of Wales (list S1). None of the genera identified in this positive control were present in the experimental samples.



Supplementary Figure 1 Abundance of airborne grass pollen taxa throughout the grass pollen season.

Abundance of rare grasses (expressed as proportion of total reads). Sampling sites are indicated in the top panel, followed by the marker used to identify grass pollen. Due to errors in sampling equipment, only 4 alternate weeks (out of a possible 7 alternate weeks) of samples were collected at the York sampling site. Note that the y axes differ between panels. Sampling sites are indicated in the right panel label abbreviated as follows: BNG = Bangor; EXE = Exeter; ING = Invergowrie; IOW = Isle of Wight; WOR = Worcester; YORK = York. A map of sampling locations and daily Poaceae pollen concentrations can be found in Figure 1.



129

130 Supplementary Figure 2 Grass community composition is structured by time across the UK. Non-

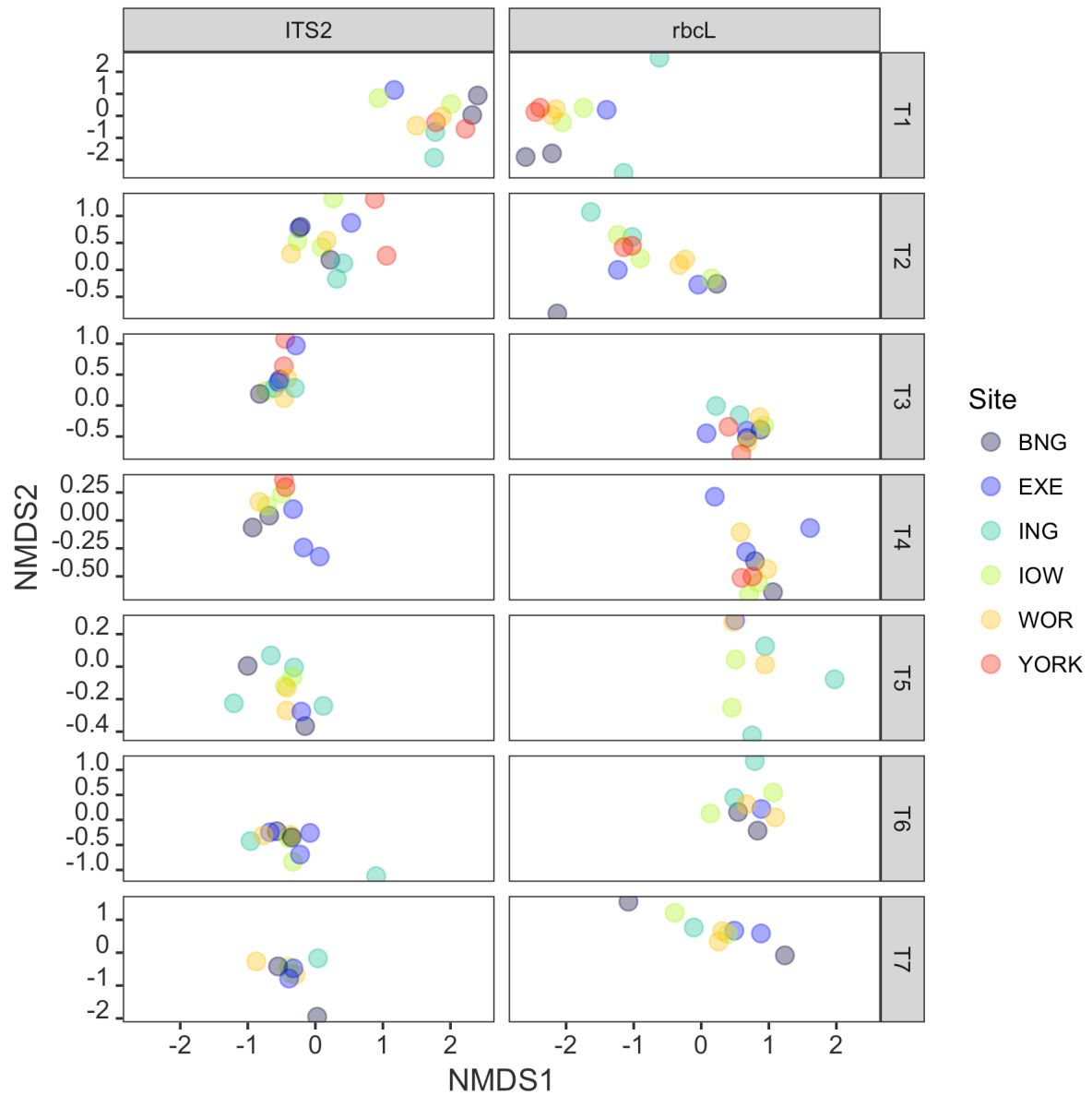
131 metric multidimensional scaling (NMDS) ordination of grass community identified using ITS2 and *rbcL*

132 barcode makers. Different shapes indicate the sampling location and colour of symbols indicate the

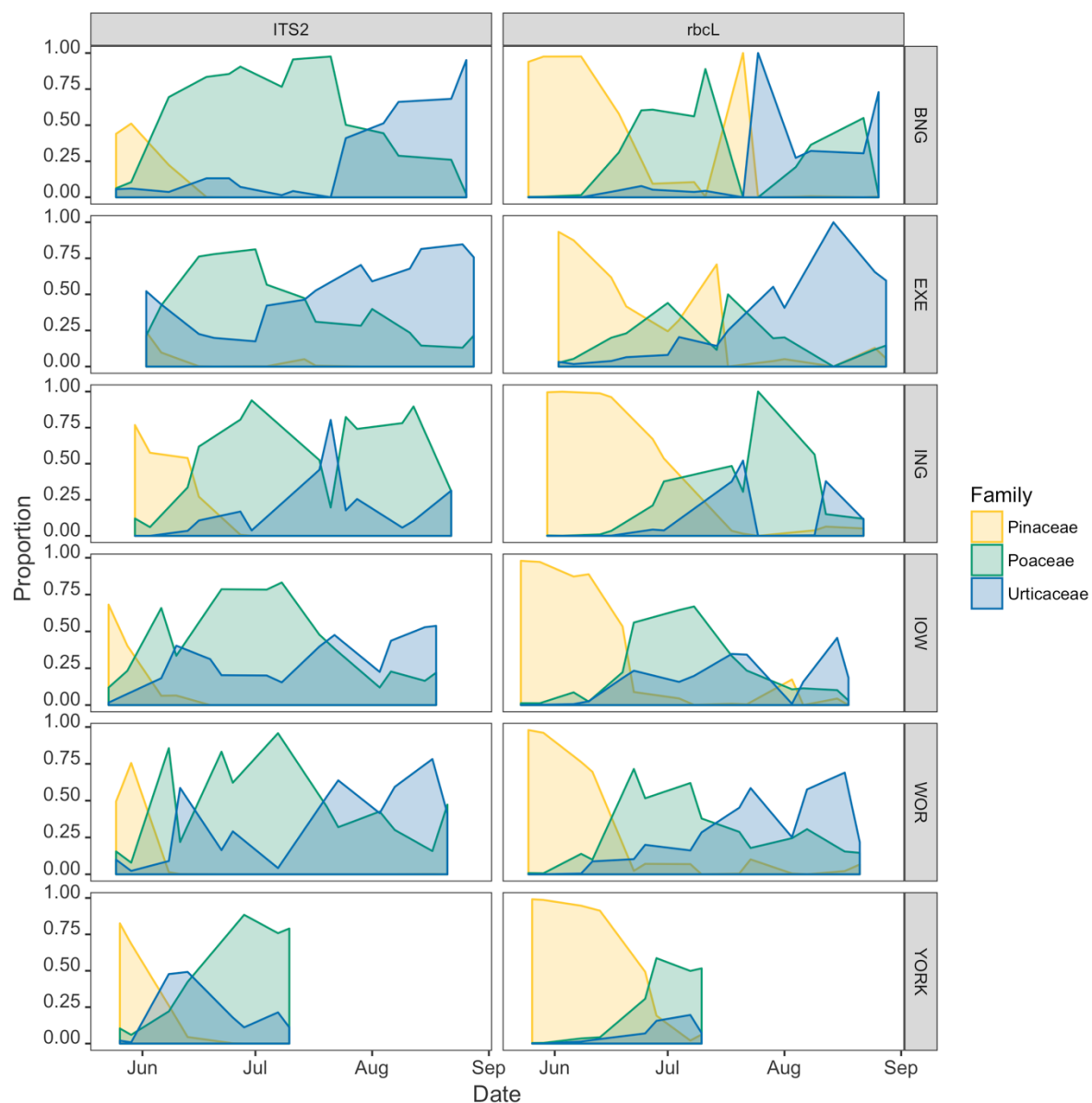
133 month that pollen was collected. Refer to figure 2 for site name abbreviations.

134

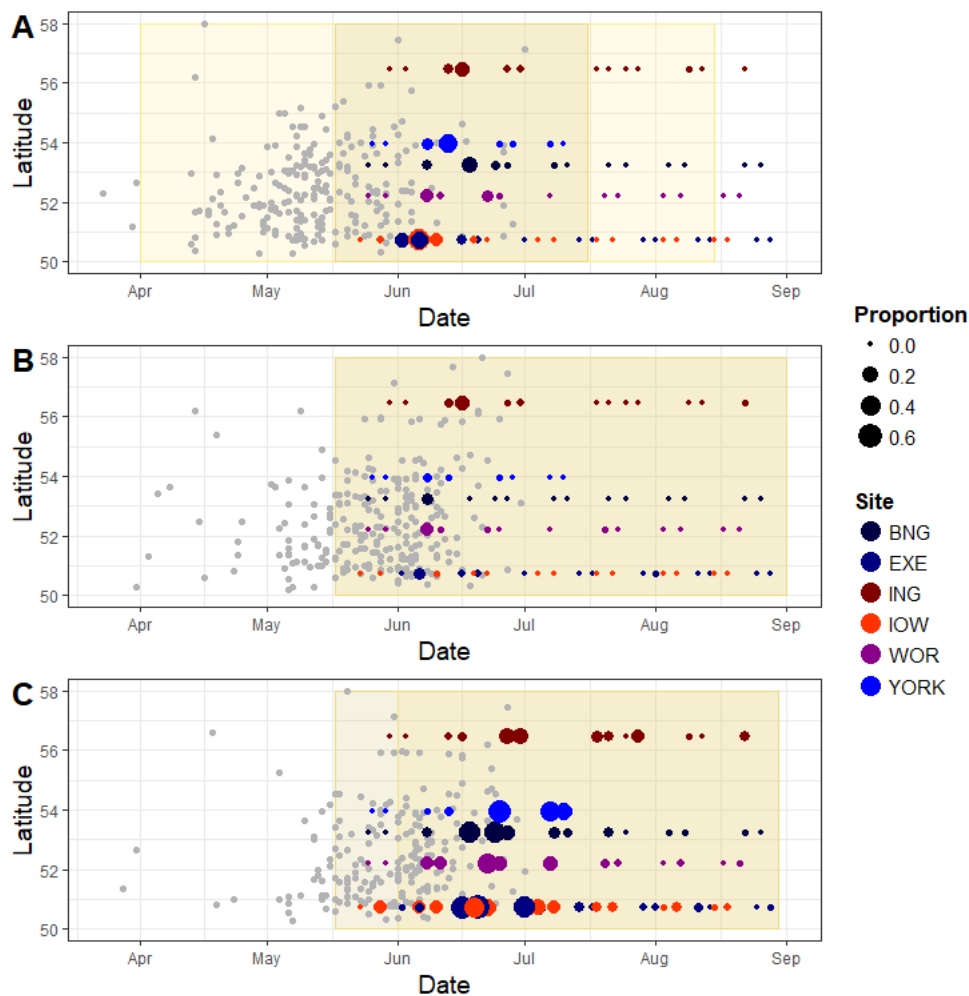
135



Supplementary Figure 3 Non-metric multidimensional scaling (NMDS) ordination of grass community identified using ITS2 and *rbcL* barcode makers. Each timepoint T1 to T7 indicates two consecutive weeks when pollen samples were collected (May to August), in chronological order. Coloured circles indicate sampling sites. Site labels are abbreviated as follows: BNG = Bangor; EXE = Exeter; ING = Invergowrie; IOW = Isle of Wight; WOR = Worcester; YORK = York.

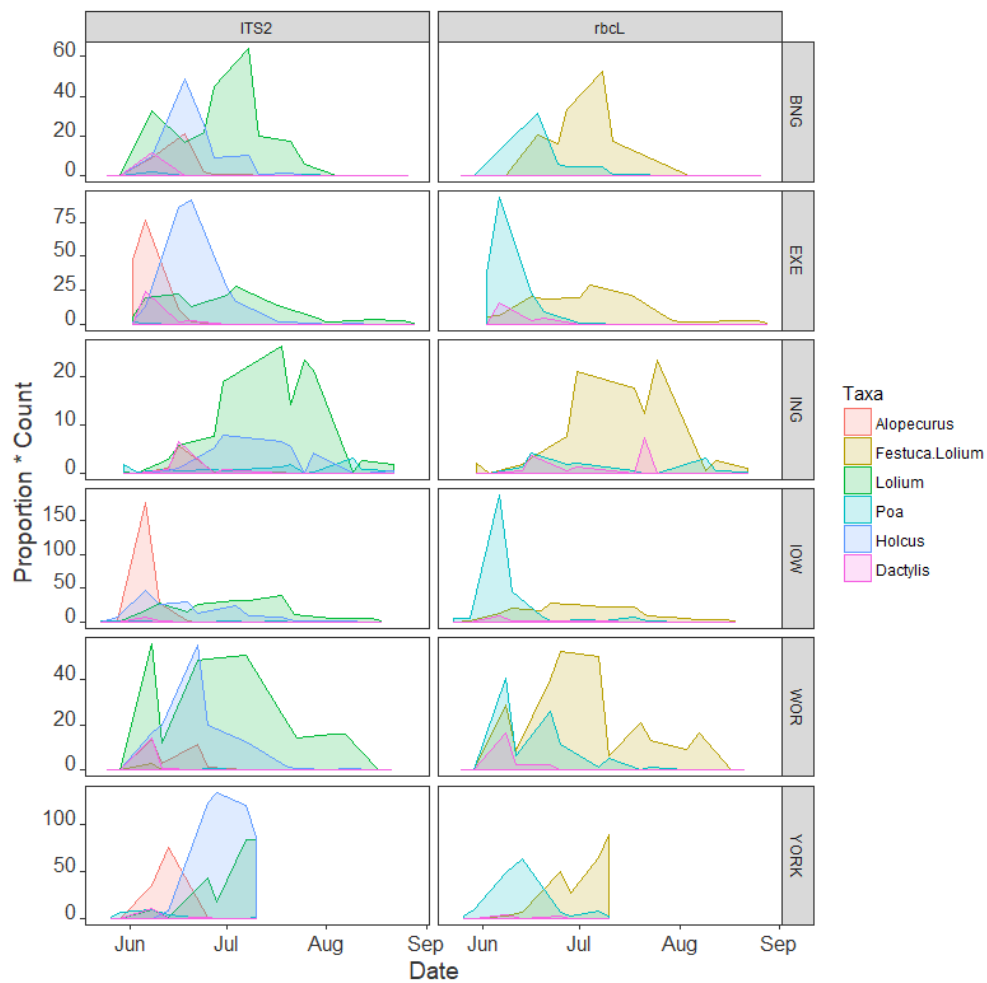


Supplementary Figure 4 Abundance of the most common families of airborne pollen. The three most abundant families of airborne pollen, Pinaceae, Poaceae and Urticaceae, expressed as proportion of total reads, collected between May and September. Markers used to identify pollen are stated in the top panel label. Due to errors in sampling equipment, only 4 alternate weeks (out of a possible 7 alternate weeks) of samples were collected at the York sampling site. Sampling sites are indicated in the right panel label abbreviated as follows: BNG = Bangor; EXE = Exeter; ING = Invergowrie; IOW = Isle of Wight; WOR = Worcester; YORK = York. A map of sampling locations can be found in Figure 1.



Supplementary Figure 5 Airborne grass pollen observed 3-4 weeks after first flowering dates.

Comparison of genus incidence in metabarcoding data with records of first flowering dates in 2016 from the citizen science project Nature's Calendar (www.naturescalendar.org.uk) for (A) *Alopecurus pratensis*, (B) *Dactylis glomerata* and (C) *Holcus lanatus*. Each grey point represents the earliest time of flower heading as observed by a participant in the project. Coloured points represent metabarcoding samples, with the size of the point representing the proportion of total reads assigned to the relevant genus. Yellow shaded areas represent the expected flowering period as described in ¹⁰, with darker shades showing the 'main' flowering period.



Supplementary Figure 6 Relative abundance of the five most abundant grasses at genus level, normalized according to airborne pollen concentration data. Relative abundances were calculated as a proportion of reads assigned to Poaceae, rather than of reads as a whole, then multiplied by mean pollen concentration across the three pooled days. Markers used to identify grass pollen are stated in the top panel label. Due to errors in sampling equipment, only 4 alternate weeks (out of a possible 7 alternate weeks) of samples were collected at the York sampling site. Sampling sites are indicated in the right panel label abbreviated as follows: BNG = Bangor; EXE = Exeter; ING = Invergowrie; IOW = Isle of Wight; WOR = Worcester; YORK = York.

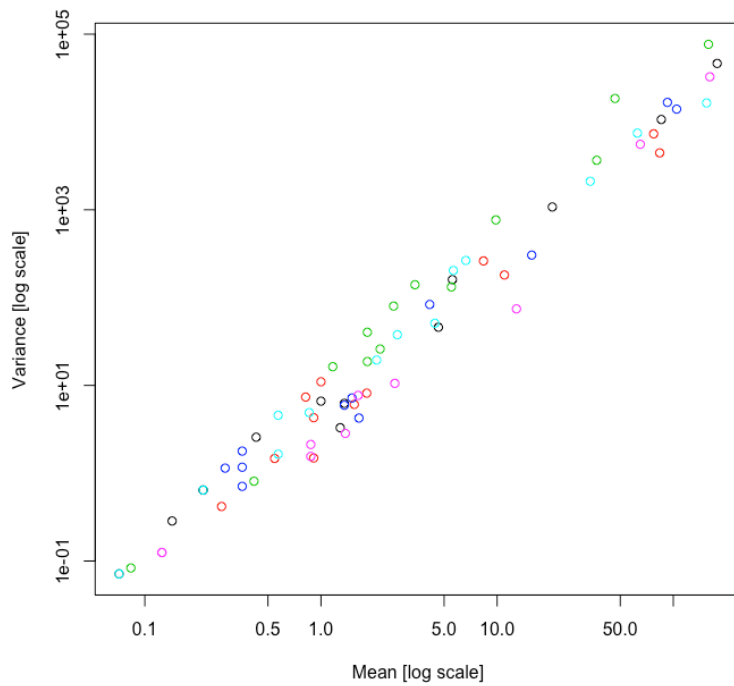


172

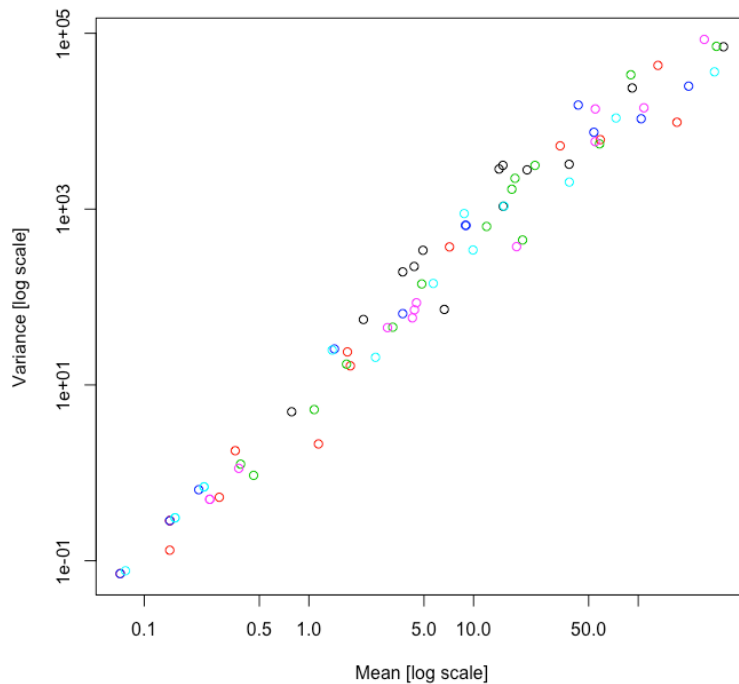
173 **Supplementary Figure 7** Photograph of 1.5 ml microcentrifuge tubes mounted onto carousel on
174 Burkard Automatic Multi-Vial Cyclone Sampler. Author provided.

175

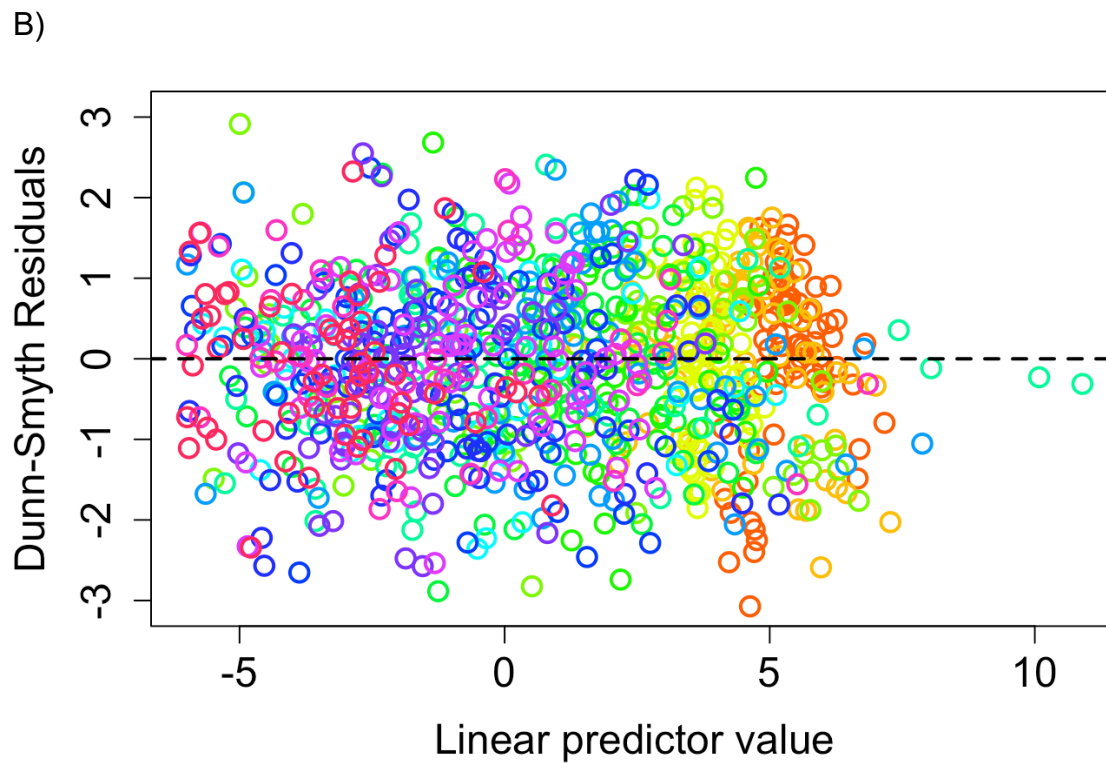
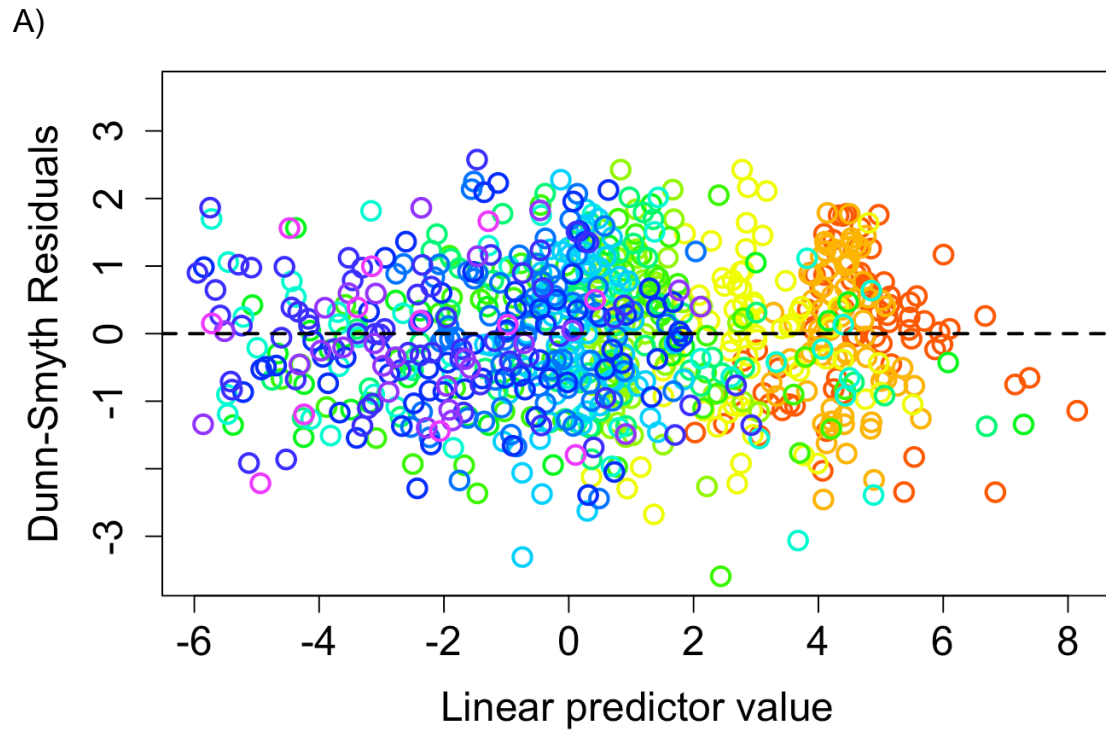
A)



B)

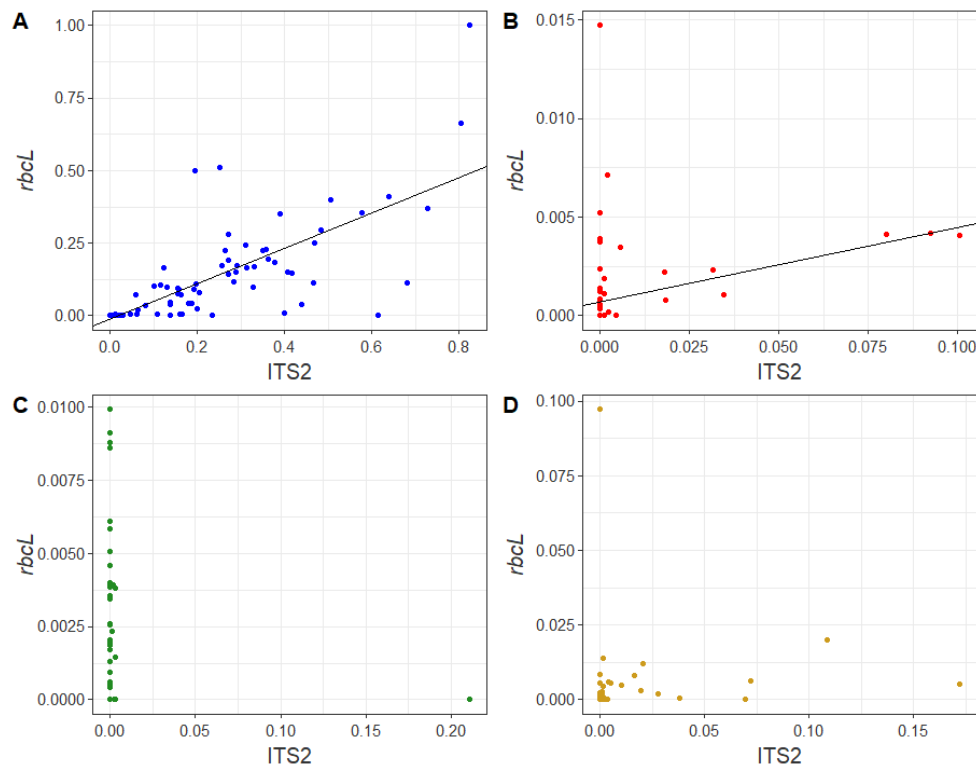


Supplementary Figure 8 There is a strong relationship between the mean proportion of sequences and the variance of the proportion of sequences from each sampling site using both A) *rbcL* and B) ITS2 markers. Coloured circles denote sampling site. The plots were produced using the `meanvar.plot` function in the `mvabund` package in R (21).



Supplementary Figure 9 Scatter plot of linear predictor values and the residuals output from the models selected to analyse the abundance data produced by the A) *rbcL* marker and B) ITS2 marker. Little association suggests that the models selected are plausible and the mean-variance assumption of the negative binomial regression is correct. Coloured circles denote different genera in the

abundance data. The plots were produced using the `plot.manyglm` function in the `mvabund` package in R (21).



Supplementary Figure 10 Correlations between proportions of reads made up by the same genus in the two marker gene datasets. All four genera present in both datasets are shown: (A) *Lolium/Festuca*, (B) *Anthoxanthum*, (C) *Avena*, and (D) *Dactylis*. For cases where there was a significant relationship between relative abundances in both datasets, black lines show the intercept and slope.

201 **List S1** Borneo plant taxa pooled for the exotic plant positive control.

202

203 *Aglaia sp.*

204 *Antidesma sp.*

205 *Baccaurea stipulata*

206 *Cynometra sp.*

207 *Dalbergia sp.*

208 *Dehaasia sp.*

209 *Dillenia excelsa*

210 *Diospyros sp.*

211 *Kleinhovia hospita*

212 *Lagerstroemia sp.*

213 *Lophopyxis sp.*

214 *Madhuca dubardii*

215 *Mallotus muticus*

216 *Microcos crassifolia*

217 *Pternandra sp.*

218 *Pterospermum macrocarpum*

219 *Syzygium sp.*

220 *Uncaria sp.*

221 *Urophyllum sp.*

222 *Vatica sp.*

223 *Xylosma sp.*

224

225 **Supplementary Table 1** Grass species pooled for the Grass Positive Control at equal volumes.

Grass positive control	Concentration of DNA (ng/μl)
<i>Agrostis canina</i>	0.121
<i>Agrostis capillaris</i>	1.29
<i>Agrostis gigantea</i>	9.48
<i>Agrostis stolonifera</i>	3.7
<i>Agrostis vinealis</i>	1.03
<i>Aira praecox</i>	0.8
<i>Alopecurus geniculatus</i>	1.18
<i>Alopecurus pratensis</i>	1.42
<i>Anisantha sterilis</i>	0.848
<i>Anthoxanthum odoratum</i>	0.804
<i>Arrhenatherum elatius</i>	1.36
<i>Brachypodium sylvaticum</i>	0.804
<i>Briza media</i>	2
<i>Bromopsis ramosa</i>	0.35
<i>Bromus hordeaceus</i>	0.098
<i>Catapodium rigidum</i>	3.96
<i>Cynosurus cristatus</i>	0.0736
<i>Dactylis glomerata</i>	13.8
<i>Danthonia decumbens</i>	1.34
<i>Deschampsia cespitosa</i>	2.52
<i>Deschampsia flexuosa</i>	0.648
<i>Elymus caninus</i>	1.62
<i>Elytrigia repens</i>	3.47
<i>Festuca arundinacea</i>	1.21
<i>Festuca gigantea</i>	1.32
<i>Festuca ovina</i>	0.592
<i>Festuca pratensis</i>	1.34
<i>Festuca rubra</i>	2.68
<i>Glyceria declinata</i>	0.226
<i>Glyceria fluitans</i>	0.892
<i>Glyceria maxima</i>	8.32
<i>Glyceria notata</i>	0.992
<i>Holcus lanatus</i>	0.42
<i>Holcus mollis</i>	below detection limit*
<i>Hordeum murinum</i>	0.476
<i>Hordeum secalinum</i>	0.416
<i>Lolium perenne</i>	0.452
<i>Milium effusum</i>	0.524
<i>Molinia caerulea</i>	2.24
<i>Nardus stricta</i>	0.246
<i>Phalaris arundinacea</i>	below detection limit*

<i>Phleum bertolonii</i>	0.444
<i>Phleum pratense</i>	18
<i>Phragmites australis</i>	13.2
<i>Poa annua</i>	0.0844
<i>Poa humilis</i>	2.37
<i>Poa pratensis</i>	1.13
<i>Poa trivialis</i>	below detection limit*
<i>Puccinellia distans</i>	11.1
<i>Trisetum flavescens</i>	0.736
<i>Triticum aestivum</i>	1.47
<i>Vulpia myuros</i>	0.199

226

227 * note these samples successfully amplified using *rbcL* and ITS2 primers shown in Supplementary

228 Table 5.

Supplementary Table 2 Genera included in the grass positive control, and genera detected using metabarcoding of both marker genes in both the positive control and in actual aerial DNA extracts. Genera with a grey background were detected by at least one marker gene; genera with a white background were not.

Expected	rbcl- Control	ITS2- Control	rbcl- Samples	ITS2- Samples
<i>Agrostis</i>				<i>Agrostis</i>
<i>Aira</i>				
<i>Alopecurus</i>				<i>Alopecurus</i>
<i>Anisantha</i>				
<i>Anthoxanthum</i>			<i>Anthoxanthum</i>	<i>Anthoxanthum</i>
<i>Arrhenatherum</i>		<i>Arrhenatherum</i>		<i>Arrhenatherum</i>
<i>Avena</i>	<i>Avena</i>		<i>Avena</i>	<i>Avena</i>
<i>Brachypodium</i>				
<i>Briza</i>	<i>Briza</i>	<i>Briza/Bromus</i>	<i>Briza</i>	
<i>Bromopsis</i>				
<i>Bromus</i>		<i>Briza/Bromus</i>		
<i>Catapodium</i>				
<i>Cynosurus</i>		<i>Cynosurus</i>		<i>Cynosurus</i>
<i>Dactylis</i>	<i>Dactylis</i>	<i>Dactylis</i>	<i>Dactylis</i>	
<i>Danthonia</i>				
<i>Deschampsia</i>		<i>Deschampsia</i>		<i>Deschampsia</i>
<i>Elymus</i>				
<i>Elytrigia</i>				
<i>Festuca</i>	<i>Festuca/Lolium</i>	<i>Festuca/Lolium</i>	<i>Festuca/Lolium</i>	<i>Festuca/Lolium</i>
<i>Glyceria</i>		<i>Glyceria</i>		
<i>Holcus</i>				<i>Holcus</i>
<i>Hordeum</i>		<i>Hordeum</i>		<i>Hordeum</i>
<i>Lolium</i>		<i>Lolium</i>		<i>Lolium</i>
<i>Milium</i>				
<i>Molinia</i>	<i>Molinia</i>		<i>Molinia</i>	
<i>Nardus</i>				
<i>Phalaris</i>				
<i>Phleum</i>	<i>Phleum</i>		<i>Phleum</i>	
<i>Phragmites</i>				
<i>Poa</i>	<i>Poa</i>	<i>Poa</i>	<i>Poa</i>	<i>Poa</i>
<i>Puccinellia</i>				
<i>Trisetum</i>				
<i>Triticum</i>	<i>Triticum</i>			
<i>Vulpia</i>				

234 **Supplementary Table 3** Latitude and longitude of each pollen sampling site.

Site Name	Abbreviation	Latitude	Longitude
Bangor	BNG	53.2300	-4.1300
Exeter	EXE	50.7365	-3.5322
Invergowrie	ING	56.4576	-3.0687
Isle of Wight	IOW	50.7111	-1.3009
Worcestershire	WORK	52.1976	-2.2430
York	YORK	53.9484	-1.0535

235

Supplementary Table 4 Sample collection dates of each sequenced air sample. Three consecutive days of air samples were pooled during DNA extraction (note that sample ING_w2_p2, three consecutive samples were unavailable due to sampling error and the next sampling day was selected for pooling). The mean pollen concentration for the three pooled days and the index i5 and i7 sequence for demultiplexing is shown here.

Sample	Index i5 and i7 Sequence	Week	Pool	Site	Collection date (2016)	Mean pollen conc. (grains m ⁻³)
BNG_w1_p1	CAAGTCGT	1	1	BNG	25 May - 28 May	61.7
BNG_w1_p2	TAACGTCG	1	2	BNG	29 May - 01 Jun	27
BNG_w2_p1	CTGTATGC	2	1	BNG	08 Jun - 11 Jun	NA
BNG_w2_p2	TGCTTGCT	2	2	BNG	18 Jun - 21 Jun	NA
BNG_w3_p1	GTAGTACC	3	1	BNG	24 Jun - 27 Jun	NA
BNG_w3_p2	AAGTCCTC	3	2	BNG	27 Jun - 30 Jun	NA
BNG_w4_p1	GCATAACG	4	1	BNG	08 Jul - 11 Jul	35.3
BNG_w4_p2	ATAGTCGG	4	2	BNG	11 Jul - 14 Jul	18.3
BNG_w5_p1	TAGGAGCT	5	1	BNG	21 Jul - 24 Jul	5.7
BNG_w5_p2	AGGTGTTG	5	2	BNG	25 Jul - 28 Jul	2
BNG_w6_p1	CATTGACG	6	1	BNG	04 Aug - 07 Aug	4.3
BNG_w6_p2	CCACAACA	6	2	BNG	08 Aug - 11 Aug	1.3
BNG_w7_p1	TCTAGGAG	7	1	BNG	22 Aug - 25 Aug	3.3
BNG_w7_p2	TTGCTTGG	7	2	BNG	26 Aug - 29 Aug	2.3
EXE_w1_p1	TGATCACG	1	1	EXE	02 Jun - 05 Jun	63
EXE_w1_p2	TCTGGACA	1	2	EXE	06 Jun - 09 Jun	139.3
EXE_w2_p1	CAGTGCTT	2	1	EXE	16 Jun - 19 Jun	126
EXE_w2_p2	ATAGGTCC	2	2	EXE	20 Jun - 23 Jun	124.7
EXE_w3_p1	CTGTACCA	3	1	EXE	01 Jul - 04 Jul	52.3
EXE_w3_p2	AAGCATCG	3	2	EXE	04 Jul - 07 Jul	61.3
EXE_w4_p1	CCTGTCAA	4	1	EXE	14 Jul - 17 Jul	56
EXE_w4_p2	AATGGTCG	4	2	EXE	17 Jul - 20 Jul	21.7
EXE_w5_p1	CTCCTGAA	5	1	EXE	29 Jul - 01 Aug	7
EXE_w5_p2	GACGAACT	5	2	EXE	01 Aug - 04 Aug	2.7
EXE_w6_p1	GGTCGTAT	6	1	EXE	11 Aug - 14 Aug	2.3
EXE_w6_p2	AAGTGCAG	6	2	EXE	14 Aug - 17 Aug	3.3
EXE_w7_p1	CCATGAAC	7	1	EXE	25 Aug - 28 Aug	3
EXE_w7_p2	TACTAGCG	7	2	EXE	28 Aug - 31 Aug	0.7
ING_w1_p1	GTGATCCA	1	1	ING	30 May - 02 Jun	2
ING_w1_p2	ATAACGCC	1	2	ING	03 Jun - 06 Jun	1
ING_w2_p1	ACCATAGG	2	1	ING	13 Jun - 16 Jun	7

ING_w2_p2	AGTTCGCA	2	2	ING	16 Jun, 19 Jun, 20 Jun	19.3
ING_w3_p1	CAACTTGG	3	1	ING	27 Jun - 30 Jun	19
ING_w3_p2	CGCAATGT	3	2	ING	30 Jun - 03 Jul	38
ING_w4_p1	GGCTCAAT	4	1	ING	18 Jul - 21 Jul	67.7
ING_w4_p2	GACTTGTG	4	2	ING	21 Jul - 24 Jul	22.7
ING_w5_p1	GCTACAAC	5	1	ING	25 Jul - 28 Jul	19.7
ING_w5_p2	GGTACGAA	5	2	ING	28 Jul - 31 Jul	27.3
ING_w6_p1	ACGAACGA	6	1	ING	09 Aug - 12 Aug	3.3
ING_w6_p2	AACACTGG	6	2	ING	12 Aug - 15 Aug	3.3
ING_w7_p1	TGGATGGT	7	1	ING	22 Aug - 25 Aug	3
IOW_w1_p1	TACTGCTC	1	1	IOW	23 May - 26 May	7.3
IOW_w1_p2	CTTCGCAA	1	2	IOW	28 May - 31 May	13.7
IOW_w2_p1	GATCAAGG	2	1	IOW	06 Jun - 09 Jun	253
IOW_w2_p2	GGCGAATA	2	2	IOW	10 Jun - 13 Jun	84
IOW_w3_p1	CAACGAGT	3	1	IOW	19 Jun - 22 Jun	57.7
IOW_w3_p2	ATCGGAGA	3	2	IOW	22 Jun - 25 Jun	39.7
IOW_w4_p1	TGTTCCGT	4	1	IOW	04 Jul - 07 Jul	86
IOW_w4_p2	ATCCACGA	4	2	IOW	08 Jul - 11 Jul	52.3
IOW_w5_p1	TCACCTAG	5	1	IOW	18 Jul - 21 Jul	64.3
IOW_w5_p2	AGGATAGC	5	2	IOW	22 Jul - 25 Jul	13
IOW_w6_p1	ATGACAGG	6	1	IOW	03 Aug - 06 Aug	5
IOW_w6_p2	CCGTTATG	6	2	IOW	06 Aug - 09 Aug	6.7
IOW_w7_p1	ACCTCTTC	7	1	IOW	15 Aug - 18 Aug	4.3
IOW_w7_p2	ACAGAGGT	7	2	IOW	18 Aug - 21 Aug	2
WOR_w1_p1	CGCTACAT	1	1	WOR	25 May - 28 May	0
WOR_w1_p2	AACCAGAG	1	2	WOR	29 May - 01 Jun	0
WOR_w2_p1	GCAATTCC	2	1	WOR	08 Jun - 11 Jun	114.7
WOR_w2_p2	AGCCGTAA	2	2	WOR	11 Jun - 14 Jun	40.7
WOR_w3_p1	AACAAGGC	3	1	WOR	22 Jun - 25 Jun	131
WOR_w3_p2	GAGCAATC	3	2	WOR	25 Jun - 28 Jun	78.7
WOR_w4_p1	AGTATGCC	4	1	WOR	07 Jul - 10 Jul	76
WOR_w4_p2	TCGATGAC	4	2	WOR	10 Jul - 13 Jul	16
WOR_w5_p1	GATACCTG	5	1	WOR	20 Jul - 23 Jul	26.3
WOR_w5_p2	ACCGACAA	5	2	WOR	23 Jul - 26 Jul	16
WOR_w6_p1	ACGAATCC	6	1	WOR	03 Aug - 06 Aug	0
WOR_w6_p2	TCGAGAGT	6	2	WOR	07 Aug - 10 Aug	0
WOR_w7_p1	GTTCTTCG	7	1	WOR	17 Aug - 20 Aug	0
WOR_w7_p2	CCTTCCAT	7	2	WOR	21 Aug - 24 Aug	0
YORK_w1_p1	TCCACGTT	1	1	YORK	26 May - 29 May	3
YORK_w1_p2	TTACCGAC	1	2	YORK	29 May - 01 Jun	9.7
YORK_w2_p1	TTCGCCAT	2	1	YORK	08 Jun - 11 Jun	84.7

YORK_w2_p2	TATGGCAC	2	2	YORK	13 Jun - 16 Jun	96.7
YORK_w3_p1	CGCGTATT	3	1	YORK	25 Jun - 28 Jun	178
YORK_w3_p2	AGCCTATC	3	2	YORK	28 Jun - 01 Jul	157
YORK_w4_p1	GACACAGT	4	1	YORK	07 Jul - 10 Jul	234.3
YORK_w4_p2	GAGAGTAC	4	2	YORK	10 Jul - 13 Jul	245.3
Negative control 1	CCACTAAG	-	-	-	-	-
Negative control 2	CCACATTG	-	-	-	-	-
Negative control 3	CCGATGTA	-	-	-	-	-
Negative control 4	CTCGGTAA	-	-	-	-	-
Negative control 5	AACCGTGT	-	-	-	-	-
Negative control 6	CGGTTGTT	-	-	-	-	-
Negative control 7	CTAGCAGT	-	-	-	-	-
Negative control 8	ACAACAGC	-	-	-	-	-
Negative control 9	GATTGTCC	-	-	-	-	-
Exotic positive control	ACAGGCAT	-	-	-	-	-
Grass positive control	TTCGTACG	-	-	-	-	-

Supplementary Table 5 Primer sequences used in library preparation. Round 1 PCR primer sequences contain forward or reverse template primer, the forward primer sequence contains a series of N's in order to improve clustering and cluster detection on MiSeq sequencing. Round 1 and round 2 sequences contain complementary universal tails. Round 2 PRC primers sequences also contain the P5 or P7 Illumina adaptors and an 8 bp unique index on both the forward and reverse primers used for demultiplexing samples (see Supplementary Table 4 for index i5 and i7 sequence).

Round 1 PCR	
Forward Universal Tail - NNNNNN - Template Specific Primer <i>rbclLaF</i>	
[ACACTCTTCCCTACACGACGCTCTCCGATCT]-[NNNNNN]-[ATGTCACCACAAACAGAGACTAAAGC]	
Reverse Universal Tail - Template Specific Primer <i>rbclr506</i>	
[GTGACTGGAGTTCAGACGTGTGCTCTCCGATCT]-[AGGGGACGACCATACTTGTTCA]	
Forward Universal Tail - NNNNNN - Template Specific Primer <i>ITS2F</i>	
[ACACTCTTCCCTACACGACGCTCTCCGATCT]-[NNNNNN]-[ATGCGATACTTGGTGTGAAT]	
Reverse Universal Tail - Template Specific Primer <i>ITS3R</i>	
[GTGACTGGAGTTCAGACGTGTGCTCTCCGATCT]- [GACGCTTCTCCAGACTACAAT]	
Round 2 PCR	
P5 Illumina adapter - i5 index - Forward Universal Tail	
[AATGATACGGCGACCACCGAGATCTACAC]-[i5 index]-[ACACTCTTCCCTACACGACGCTC]	
P7 Illumina adapter - i7 index - Reverse Universal Tail	
[CAAGCAGAAGACGGCATACGAGAT]-[i7 index]-[GTGACTGGAGTTCAGACGTGTGCTC]	

249 **Supplementary Table 6** Output from the models used to analyse the *rbcL* and ITS2 data, without the
 250 York sampling site.

<i>rbcL</i>	Res.Df	Df.diff	LR	Pr(>Dev)
Time	71	1	46.71	0.001
Latitude	70	1	26.4	0.024
Longitude	69	1	27.1	0.014
Time:Latitude	68	1	47.36	0.001
ITS2				
Time	74	1	128.8	0.001
Latitude	73	1	73.2	0.001
Longitude	72	1	33.0	0.011
Time:Latitude	71	1	34.2	0.04

251

252

References

1. D'Amato, G. *et al.* in *Allergy and Allergen Immunotherapy* **62**, 261–279 (2017).
2. Peel, R. G. *et al.* Seasonal variation in diurnal atmospheric grass pollen concentration profiles. *Biogeosciences* **11**, 821–832 (2014).
3. Cebrino, J., Galán, C. & Domínguez-Vilches, E. Aerobiological and phenological study of the main Poaceae species in Córdoba City (Spain) and the surrounding hills. *Aerobiologia* **32**, 595–606 (2016).
4. León-Ruiz, E., Alcázar, P., Domínguez-Vilches, E. & Galán, C. Study of Poaceae phenology in a Mediterranean climate. Which species contribute most to airborne pollen counts?
doi:10.1007/s10453-010-9174-y
5. Chapman, D. S. Greater phenological sensitivity to temperature on higher Scottish mountains: new insights from remote sensing. *Global Change Biology* **19**, 3463–3471 (2013).
6. Goldberg, C., Buch, H., Moseholm, L. & Weeke, E. R. Airborne Pollen Records in Denmark, 1977–1986. *Grana* **27**, 209–217 (1988).
7. Skjøth, C. A., Baker, P., Sadyś, M. & Adams-Groom, B. Pollen from alder (*Alnus* sp.), birch (*Betula* sp.) and oak (*Quercus* sp.) in the UK originate from small woodlands. *Urban Climate* **14**, 414–428 (2015).
8. Cheng, T. *et al.* Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. *Molecular Ecology Resources* **16**, 138–149 (2015).
9. Hollingsworth, M. L. *et al.* Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol. Ecol. Resour.* **9**, 439–457 (2009).
10. Alberdi, A., Aizpurua, O., Gilbert, M. T. P. & Bohmann, K. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution* **9**, 134–147 (2018).
11. Ghitarrini, S., Galán, C., Frenguelli, G. & Tedeschini, E. Phenological analysis of grasses (Poaceae) as a support for the dissection of their pollen season in Perugia (Central Italy). *Aerobiologia* **33**, 339–349 (2017).
12. Romero-Morte, J., Rojo, J., Rivero, R., Fernández-González, F. & Pérez-Badia, R. Standardised index for measuring atmospheric grass-pollen emission. *Science of the Total Environment* **612**, 180–191 (2018).
13. Bista, I. *et al.* Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Molecular Ecology Resources* 1–15 (2018).
14. Pawluczyk, M. *et al.* Quantitative evaluation of bias in PCR amplification and next-generation sequencing derived from metabarcoding samples. *Analytical and Bioanalytical Chemistry* **407**, 1841–1848 (2015).
15. Suzuki, M. T. & Giovannoni, S. J. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology* **62**, 625–630 (1996).