Supplementary information

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

Authors: Georgina L. Brennan[1*], Caitlin Potter[2], Natasha de Vere[2,3], Gareth W. Griffith[2], Carsten A. Skjøth[4], Nicholas J. Osborne[5,6], Benedict W. Wheeler[5], Rachel N. McInnes[7], Yolanda Clewlow[7], Adam Barber[7], Helen M. Hanlon[7], Matthew Hegarty[2], Laura Jones[1,3], Alexander Kurganskiy[4], Francis M. Rowney[5], Charlotte Armitage[8], Beverley Adams-Groom[4], Col R. Ford[3], Geoff M. Petch[4], The PollerGEN Consortium, and Simon Creer[1*].

Consortium Angela Elliot[9], Carl A. Frisk[4], Roy Neilson[10], Stephen Potter[11], Abdullah M. Rafiq[1], David, B. Roy[12], Katherine Selby[13], Natascha Steinberg[9]

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 species[2] 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 pollen[5].

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 collected[6]. 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 plants[8], whereas *rbcL* primers are highly universal but the marker shows lower resolution between closely related plants[9]. 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.

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

Positive and negative controls

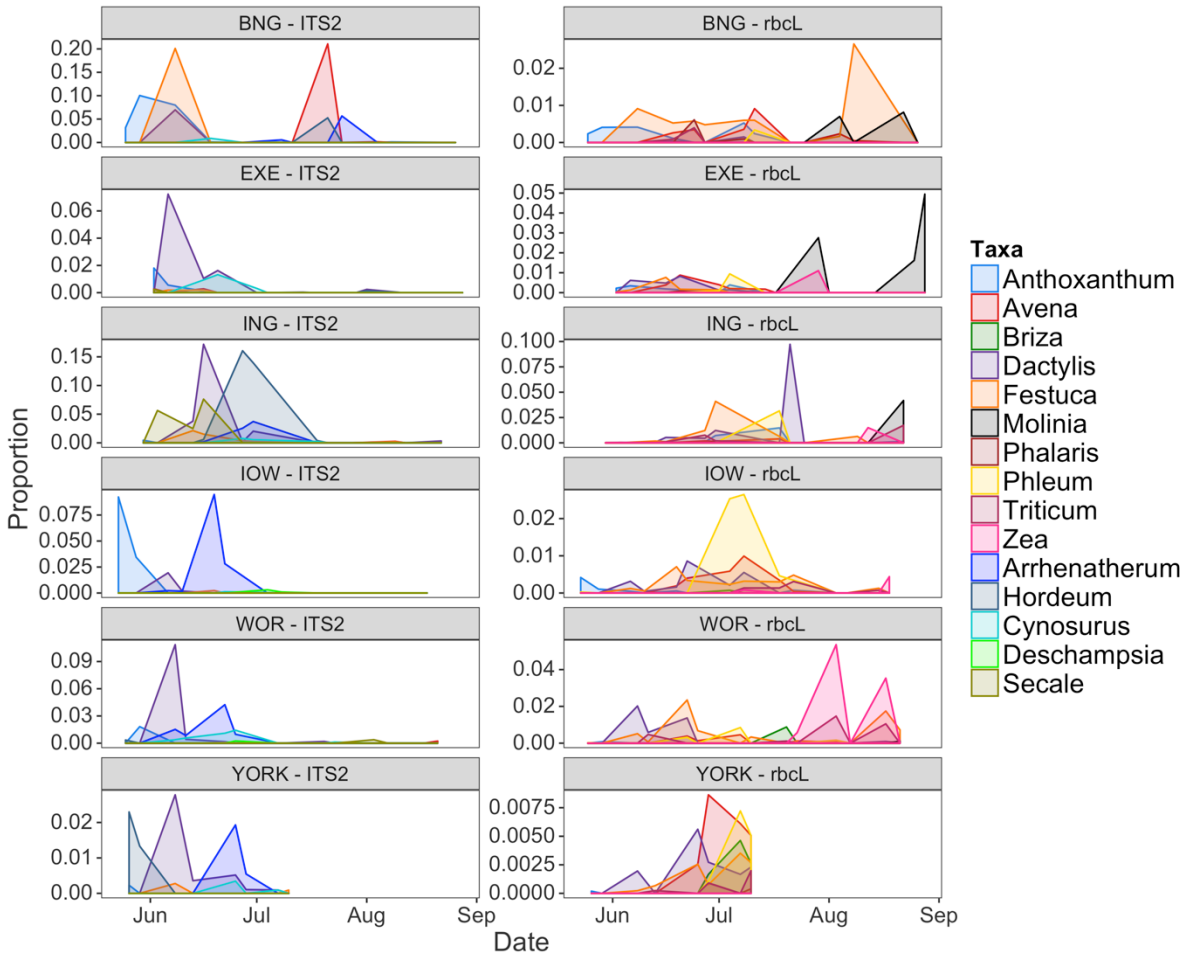
Negative controls, with all reagents and no DNA were used to identify any cross-contamination. Of the six negative controls, four contained no reads following quality control filtering, one contained a single read and one contained nine reads in the *rbcL* database. None of the negative controls in the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

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

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

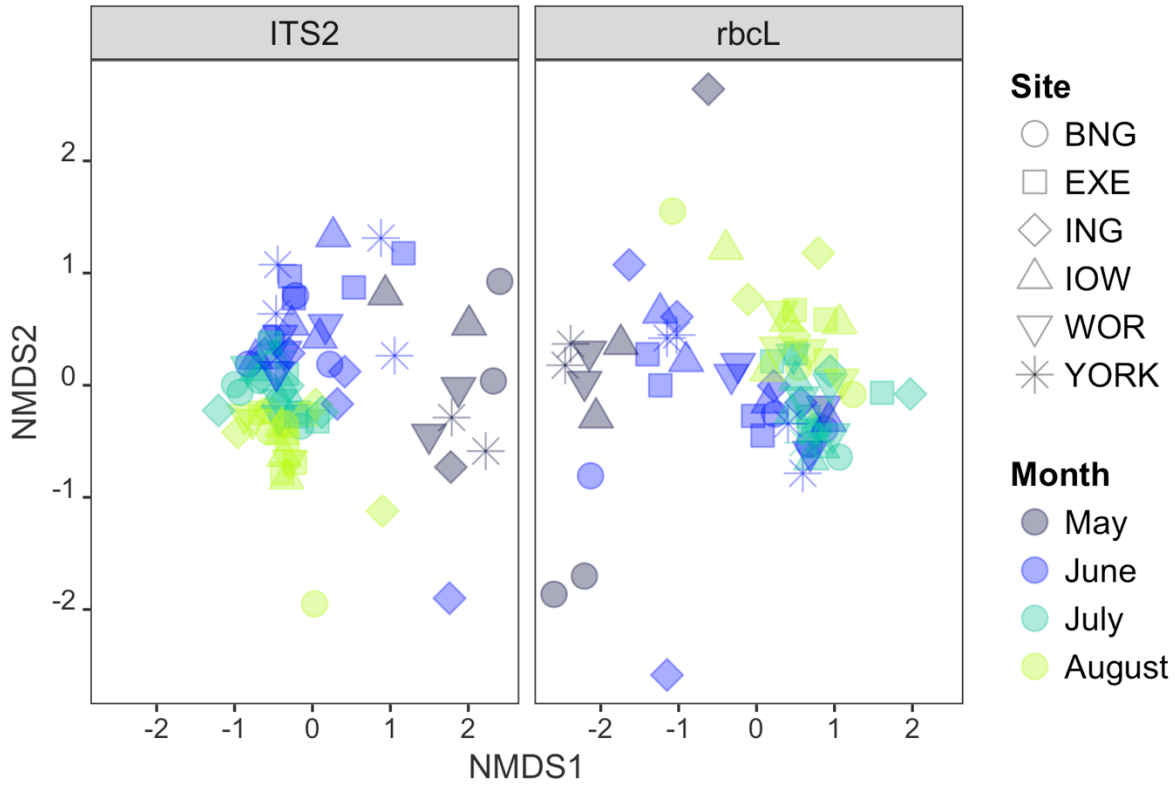
and *rbcL* markers respectively). However, three of the genera not detected in the positive control samples, despite being included, were detected in airborne samples (*Agrostis*, *Anthoxanthum*, *Alopecurus*), likely reflecting higher local abundances of airborne pollen (Supplementary Table 2). The 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.

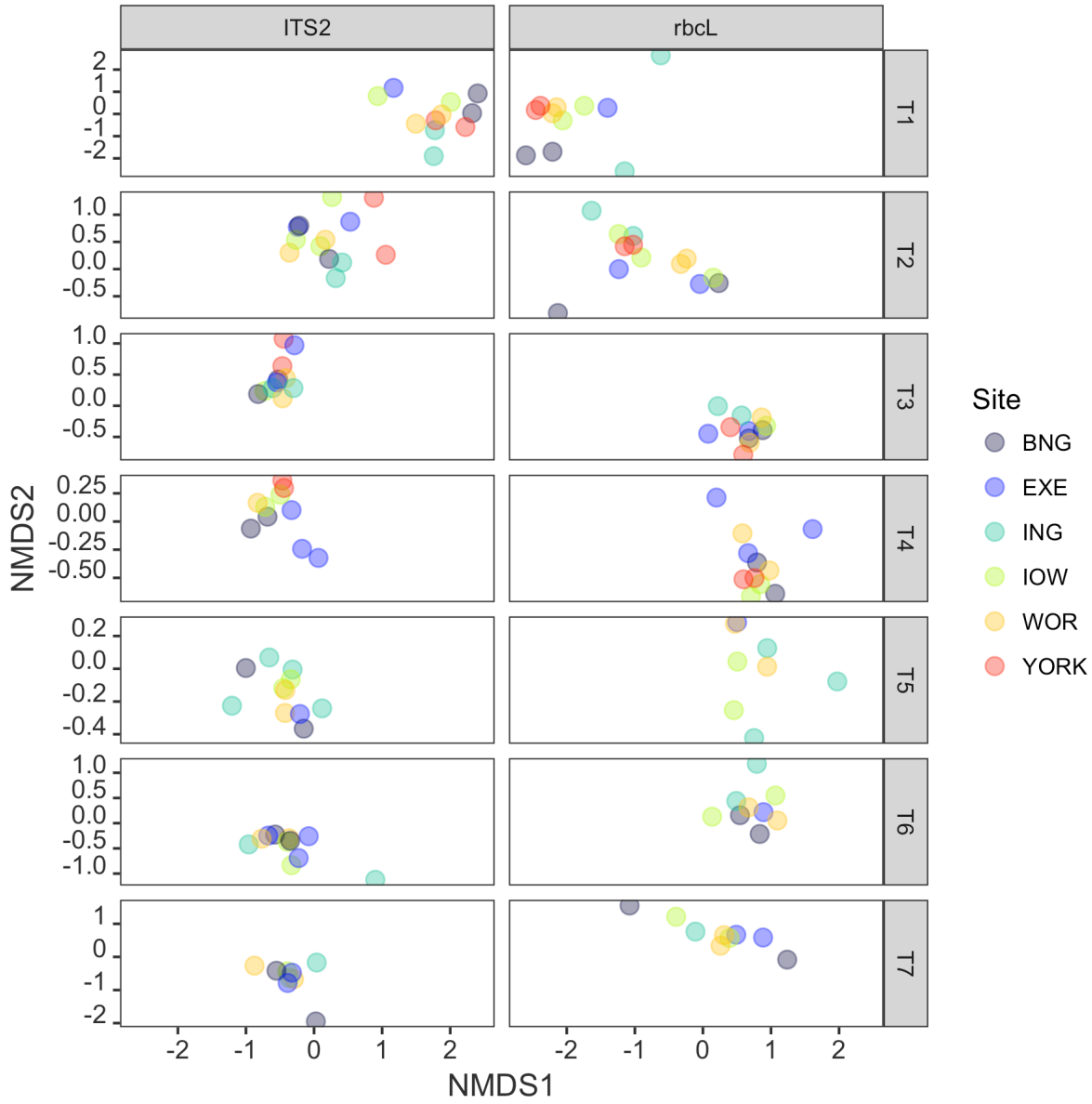


Supplementary Figure 1 Abundance of airborne grass pollen taxa throughout the grass pollen season.

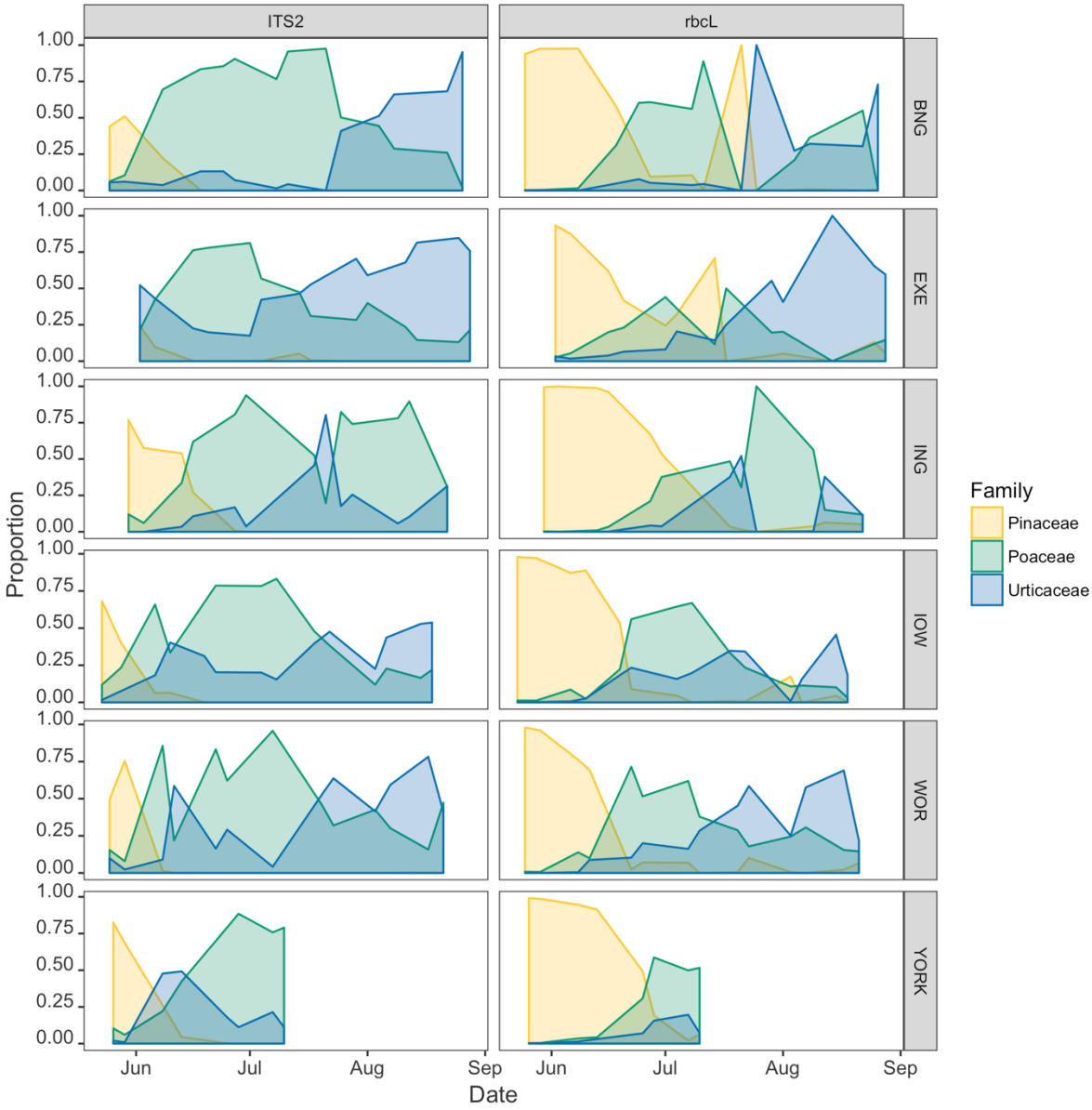
Abundance of rare grasses (expressed as proportion of total reads). Sampling sites are indicated in the top panel, followed by the marker used to identify grass pollen. Due to errors in sampling equipment, only 4 alternate weeks (out of a possible 7 alternate weeks) of samples were collected at the York 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.



Supplementary Figure 2 Grass community composition is structured by time across the UK. Non-metric multidimensional scaling (NMDS) ordination of grass community identified using ITS2 and *rbcL* barcode makers. Different shapes indicate the sampling location and colour of symbols indicate the month that pollen was collected. Refer to figure 2 for site name abbreviations.



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.



143

144 **Supplementary Figure 4 Abundance of the most common families of airborne pollen.** The three most

145 abundant families of airborne pollen, Pinaceae, Poaceae and Urticaceae, expressed as proportion of

146 total reads, collected between May and September. Markers used to identify pollen are stated in the

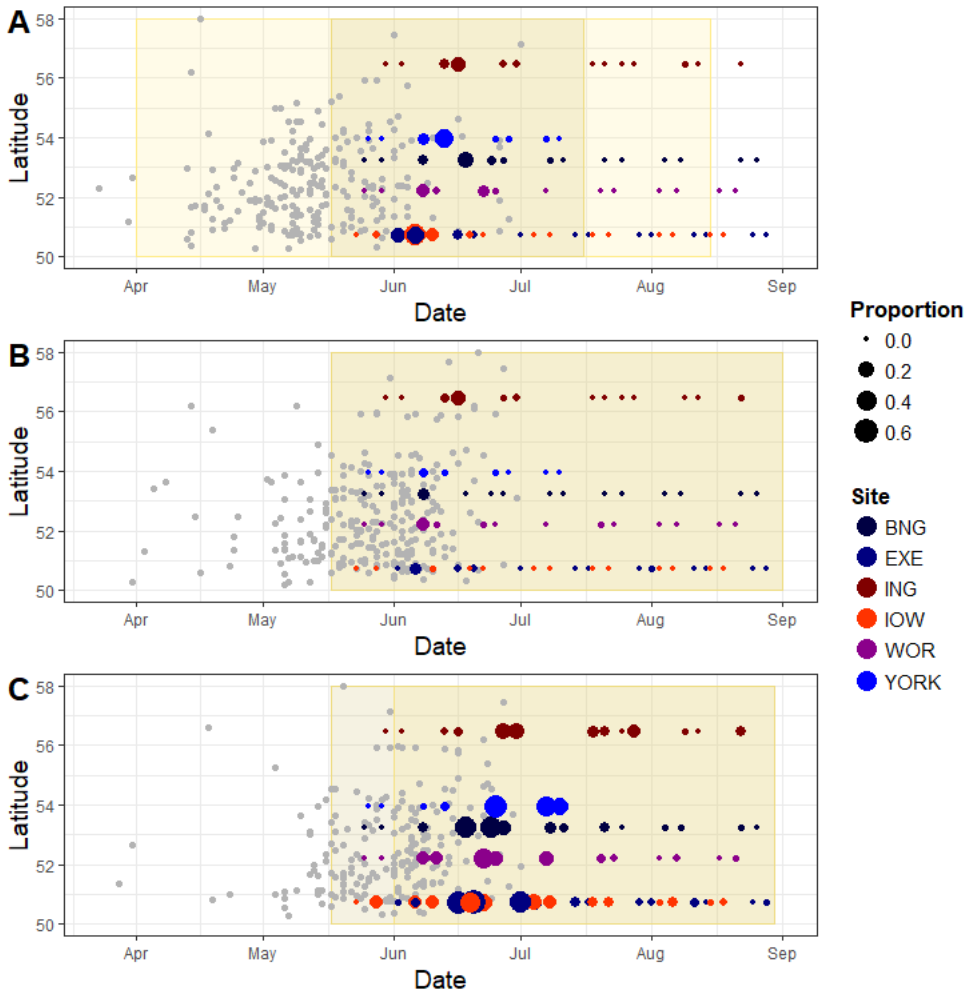
147 top panel label. Due to errors in sampling equipment, only 4 alternate weeks (out of a possible 7

148 alternate weeks) of samples were collected at the York sampling site. Sampling sites are indicated in

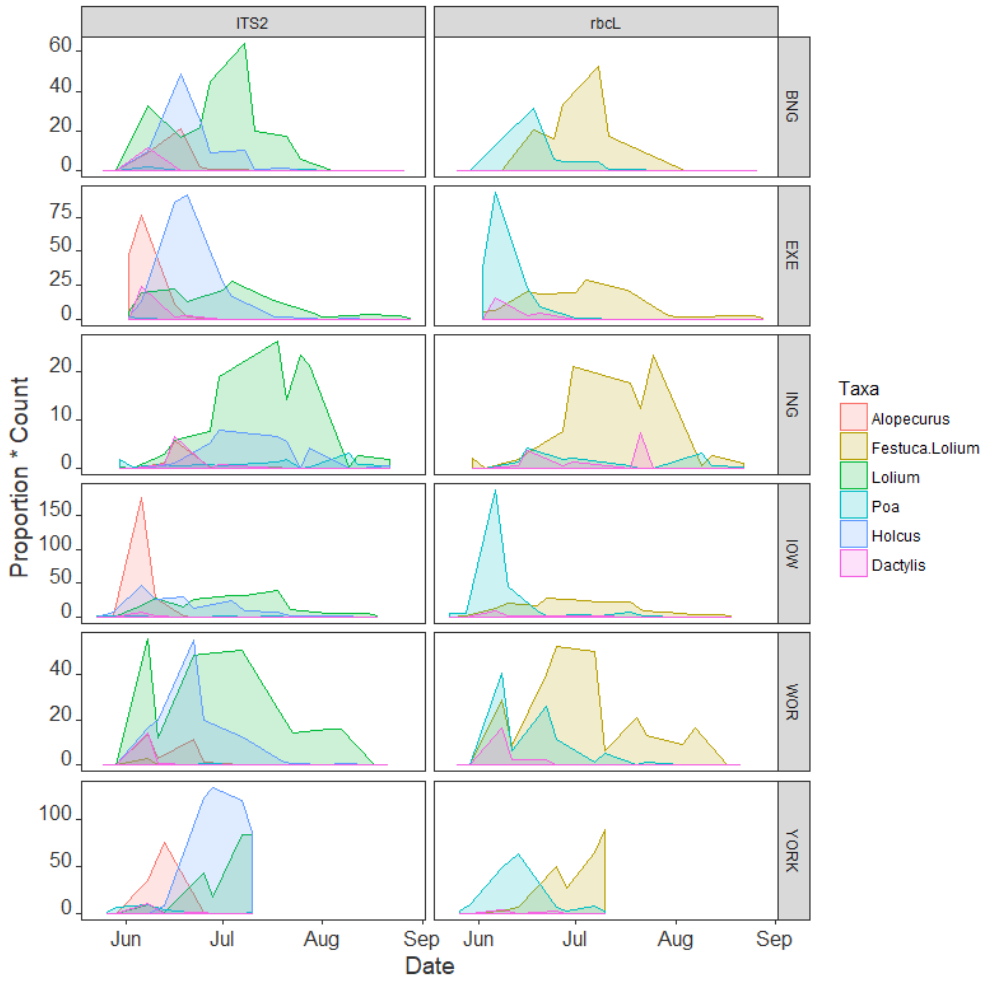
149 the right panel label abbreviated as follows: BNG = Bangor; EXE = Exeter; ING = Invergowrie; IOW =

150 Isle of Wight; WOR = Worcester; YORK = York. A map of sampling locations can be found in Figure 1.

151

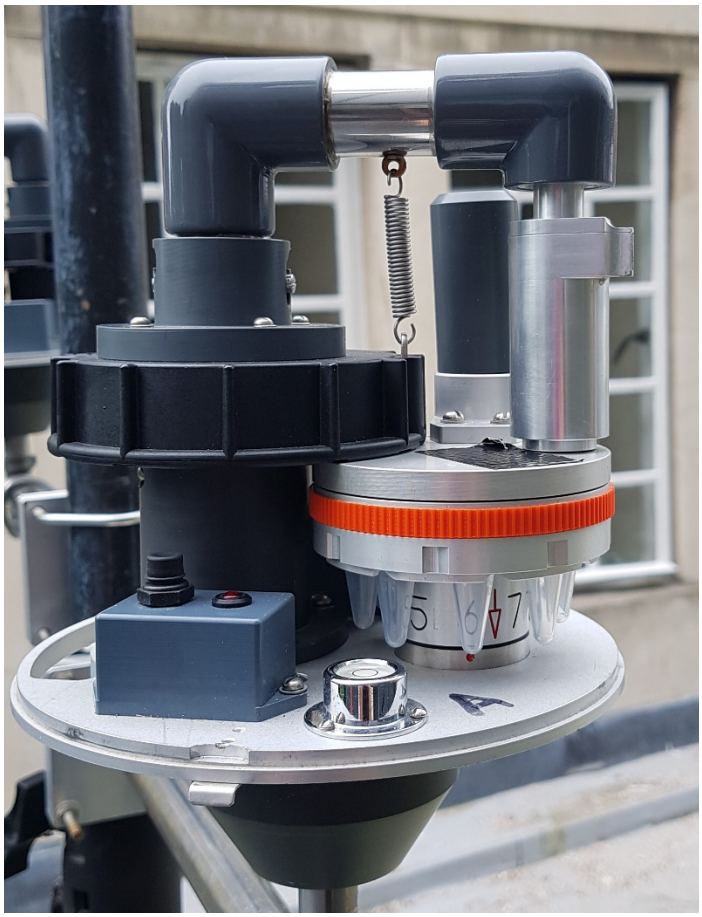


Supplementary Figure 5 Airborne grass pollen observed 3-4 weeks after first flowering dates. Comparison of genus incidence in metabarcoding data with records of first flowering dates in 2016 from the citizen science project Nature's Calendar (www.naturescalendar.org.uk) for (A) *Alopecurus pratensis*, (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.



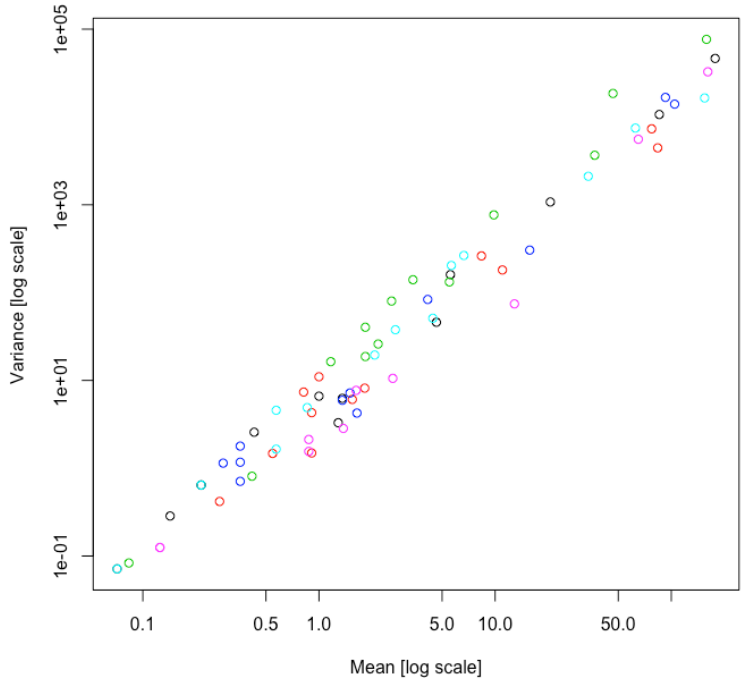
163

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

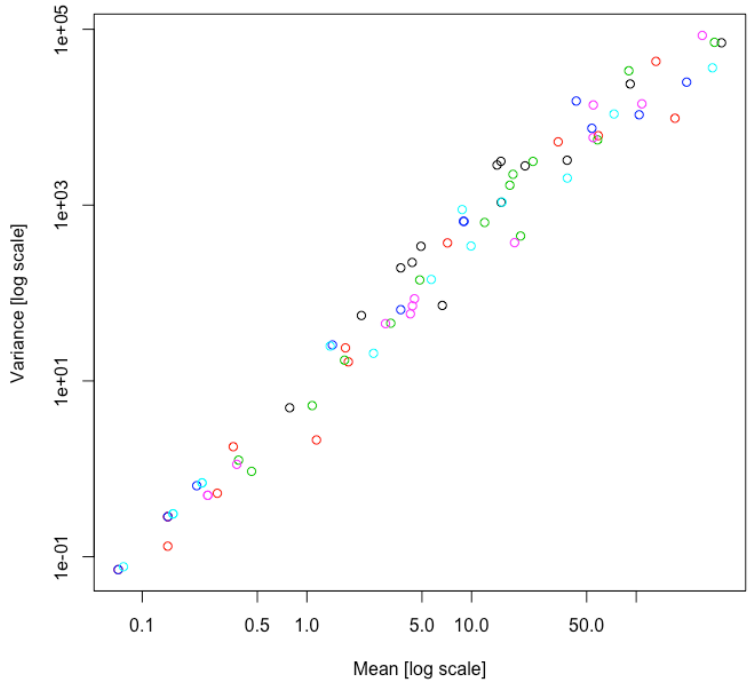


Supplementary Figure 7 Photograph of 1.5 ml microcentrifuge tubes mounted onto carousel on Burkard Automatic Multi-Vial Cyclone Sampler. Author provided.

A)

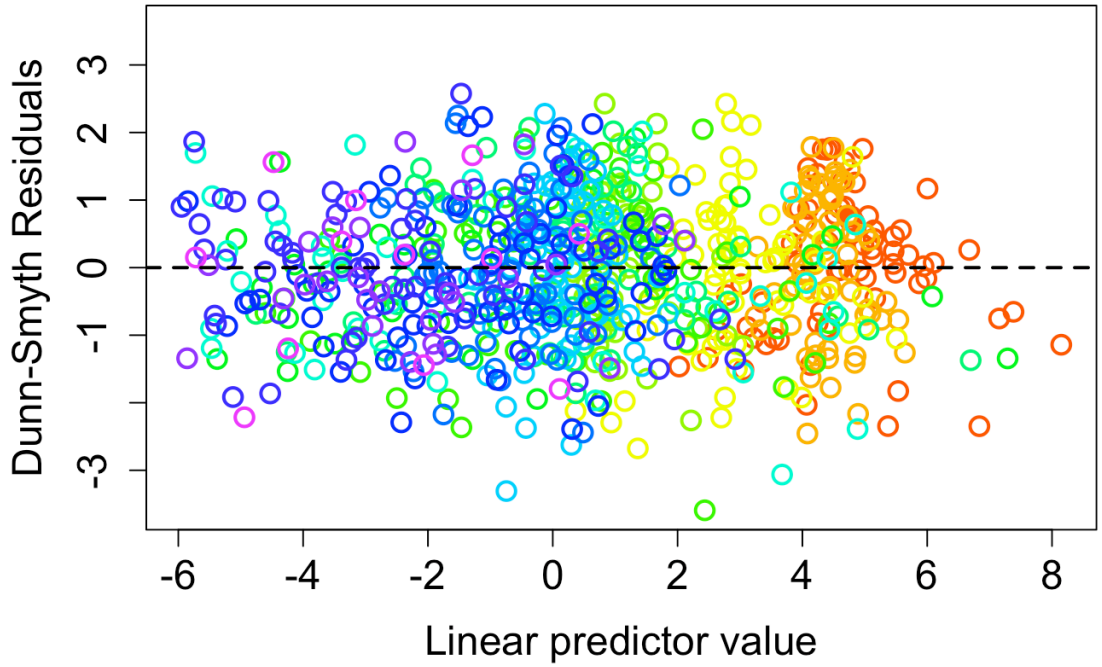


B)

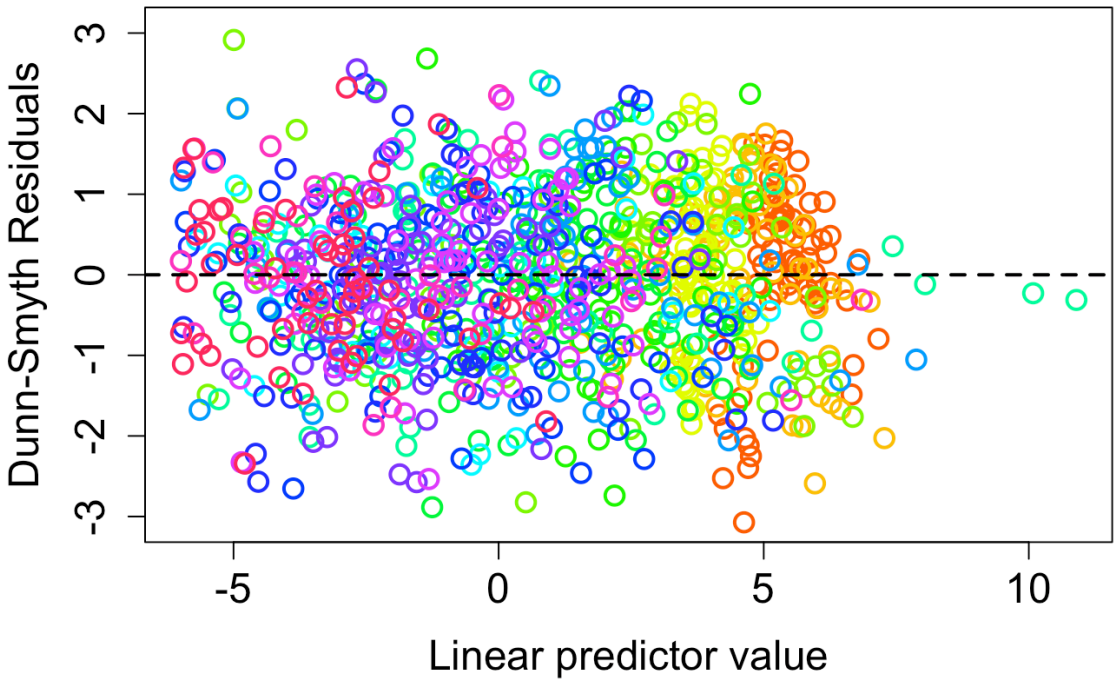


Supplementary Figure 8 There is a strong relationship between the mean proportion of sequences and the variance of the proportion of sequences from each sampling site using both A) *rbcL* and B) ITS2 markers. Coloured circles denote sampling site. The plots were produced using the meanvar.plot function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

A)

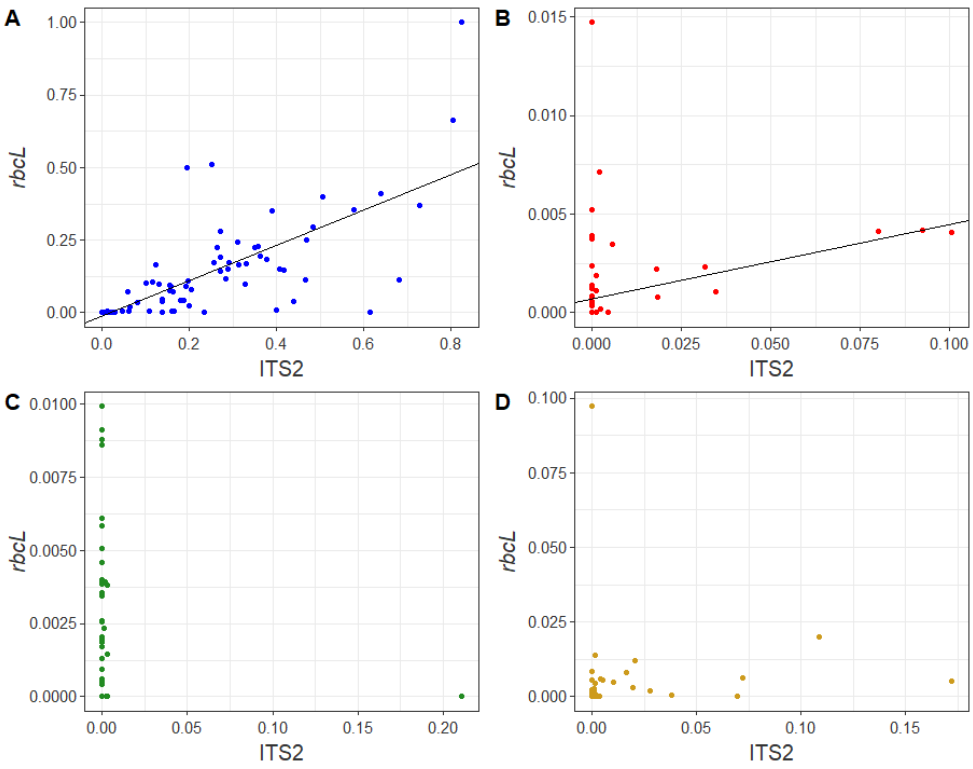


B)



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.

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

Aglaia sp.
Antidesma sp.
Baccaurea stipulata
Cynometra sp.
Dalbergia sp.
Dehaasia sp.
Dillenia excelsa
Diospryos sp.
Kleinhovia hospita
Lagerstroemia sp.
Lophopyxis sp.
Madhuca dubardii
Mallotus muticus
Microcos crassifolia
Pternandra sp.
Pterospermum macrocarpum
Syzygium sp.
Uncaria sp.
Urophyllum sp.
Vatica sp.
Xylosma sp.

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

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

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

226

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

228 Table 5.

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

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

233

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	-	-	-	-	-

241

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 ***rbcL*aF** [ACACTCTTTCCCTACACGACGCTCTTCCGATCT]-[NNNNNN]-[ATGTCACCACAAACAGAGACTAAAGC]
Reverse Universal Tail - Template Specific Primer ***rbcL*r506** [GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT]-[AGGGGACGACCATACTTGTTCA]
Forward Universal Tail - NNNNNN - Template Specific Primer **ITS2F** [ACACTCTTTCCCTACACGACGCTCTTCCGATCT]-[NNNNNN]-[ATGCGATACTTGGTGTGAAT]
Reverse Universal Tail - Template Specific Primer **ITS3R** [GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT]- [GACGCTTCTCCAGACTACAAT]
Round 2 PCR
P5 Illumina adapter - **i5 index -** Forward Universal Tail [AATGATACGGCGACCACCGAGATCTACAC]-**[i5 index]-**[ACACTCTTTCCCTACACGACGCTC]
P7 Illumina adapter - **i7 index -** Reverse Universal Tail [CAAGCAGAAGACGGCATACGAGAT]-**[i7 index]-**[GTGACTGGAGTTCAGACGTGTGCTC]

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

rbcL	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

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).