

## Article (refereed) - postprint

---

Brennan, Georgina L.; Potter, Caitlin; de Vere, Natasha; Griffith, Gareth W.; Skjøth, Carsten A.; Osborne, Nicholas J.; Wheeler, Benedict W.; McInnes, Rachel N.; Clewlow, Yolanda; Barber, Adam; Hanlon, Helen M.; Hegarty, Matthew; Jones, Laura; Kurganskiy, Alexander; Rowney, Francis M.; Armitage, Charlotte; Adams-Groom, Beverley; Ford, Col R.; Petch, Geoff M.; Elliot, Angela; Frisk, Carl A.; Neilson, Roy; Potter, Stephen; Rafiq, Abdullah M.; Roy, David B.; Selby, Katherine; Steinberg, Natascha; Creer, Simon. 2019. **Temperate airborne grass pollen defined by spatio-temporal shifts in community composition.** *Nature Ecology & Evolution*, 3 (5). 750-754.  
<https://doi.org/10.1038/s41559-019-0849-7>

© 2019 Springer Nature Publishing AG

This version available <http://nora.nerc.ac.uk/id/eprint/524587/>

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at <http://nora.nerc.ac.uk/policies.html#access>

**This document is the author's final manuscript version of the journal article following the peer review process. Some differences between this and the publisher's version may remain. You are advised to consult the publisher's version if you wish to cite from this article.**

<https://link.springer.com/>

Contact CEH NORA team at  
[noraceh@ceh.ac.uk](mailto:noraceh@ceh.ac.uk)

1 **Title: Temperate airborne grass pollen defined by spatio-temporal shifts in community**  
2 **composition**

3 **Authors:** Georgina L. Brennan<sup>1\*</sup>, Caitlin Potter<sup>2</sup>, Natasha de Vere<sup>2,3</sup>, Gareth W. Griffith<sup>2</sup>,  
4 Carsten A. Skjøth<sup>4</sup>, Nicholas J. Osborne<sup>5,6,7</sup>, Benedict W. Wheeler<sup>5</sup>, Rachel N. McInnes<sup>8</sup>,  
5 Yolanda Clewlow<sup>8</sup>, Adam Barber<sup>8</sup>, Helen M. Hanlon<sup>8</sup>, Matthew Hegarty<sup>2</sup>, Laura  
6 Jones<sup>1,3</sup>, Alexander Kurganskiy<sup>4</sup>, Francis M. Rowney<sup>5</sup>, Charlotte Armitage<sup>9</sup>, Beverley Adams-  
7 Groom<sup>4</sup>, Col R. Ford<sup>3</sup>, Geoff M. Petch<sup>4</sup>, The PollerGEN Consortium, and Simon Creer<sup>1\*</sup>.

8

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

11

12 **Affiliations:**

13 <sup>1</sup>Molecular Ecology and Fisheries Genetics Laboratory, School of Natural Sciences, Bangor  
14 University, Bangor, Gwynedd, Wales, UK

15 <sup>2</sup>IBERS, Aberystwyth University, Aberystwyth, Ceredigion, Wales, UK

16 <sup>3</sup>National Botanic Garden of Wales, Llanarthne, Carmarthenshire, Wales, UK

17 <sup>4</sup>University of Worcester, Worcester, Worcestershire, UK

18 <sup>5</sup>University of Exeter, Truro, Cornwall, UK

19 <sup>6</sup>University of New South Wales, Sydney, New South Wales, Australia

20 <sup>7</sup>School of Public Health, Faculty of Medicine, The University of Queensland, Herston,

21 Brisbane, Australia

22 <sup>8</sup>Met Office, Exeter, Devon, UK

23 <sup>9</sup>The Woodland Trust, Kempton Way, Grantham, Lincolnshire, UK.

24 <sup>10</sup>College of Life and Environmental Science, University of Exeter, Exeter, UK

25 <sup>11</sup>Ecological Sciences, The James Hutton Institute, Dundee, Scotland, UK

26 <sup>12</sup>The David Hide Asthma & Allergy Research Centre, St Mary's Hospital, Newport, Isle of  
27 Wight, UK

28 <sup>13</sup>Centre for Ecology & Hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire, UK.

29 <sup>14</sup>Environment Department, University of York, Heslington, York, UK

30

31 **Corresponding authors:** Simon Creer ([s.creer@bangor.ac.uk](mailto:s.creer@bangor.ac.uk)) and Georgina L. Brennan

32 ([g.l.b.doonan@gmail.com](mailto:g.l.b.doonan@gmail.com))

33

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

43

44 Allergens carried in airborne pollen are associated with both asthma<sup>1</sup> and allergic rhinitis (hay  
45 fever), negatively affecting 400 million people worldwide<sup>2</sup>. Pollen from the grass family  
46 (Poaceae) constitutes the most significant outdoor aeroallergen<sup>3,4</sup>, and more people are  
47 sensitised to grass pollen than to any other pollen type<sup>5</sup>. However, despite the harmful  
48 impact of grass pollen on human health, especially as reported in developed nations, current  
49 studies and forecasts categorize grass pollen at the family level (Poaceae)<sup>6,7</sup> due to difficulties  
50 in differentiating species based on morphology<sup>8</sup>. Furthermore, while different species of  
51 temperate grass flower at different timepoints<sup>9,10</sup>, it is unknown if the disparate phenology of  
52 local grass taxa at ground level are useful for making predictions on the seasonal variation in  
53 airborne pollen. Airborne pollen is highly mobile<sup>11,12</sup> and pollen concentrations often do not  
54 directly correlate to local flowering times<sup>11</sup>. Persistence and mobility of grass pollen could  
55 result in steadily increasing species richness of airborne pollen over the grass pollen season.  
56 Conversely, if grass pollen does not persist for an extended time in the air, pollen  
57 assemblages should reflect temporal turnover in species composition over the summer

58 months. Understanding the taxon-specific phenology of airborne pollen would fill a  
59 significant knowledge gap in our understanding of allergen triggers, with associated benefits  
60 to healthcare providers, pharmaceutical industries and the public.

61

62 Many species within the subfamilies Pooideae, Chloridoideae, and Panicoideae release  
63 allergenic pollen into the atmosphere<sup>5</sup>, including *Phleum* spp. (e.g. Timothy grasses), *Dactylis*  
64 spp. (Cocksfoot/Orchard grasses), *Lolium* spp. (Ryegrasses), *Festuca* spp. (Fescues), *Poa* spp.  
65 (Meadow-grasses and Bluegrasses), and *Anthoxanthum* spp. (Vernal grasses). Furthermore,  
66 some grass taxa, notably diverse cultivars and hybrids of *Lolium* spp., are widely sown in  
67 agricultural grasslands and are likely to contribute disproportionately to airborne pollen.

68 However, it is unknown whether particular grass species, or varieties/cultivars within species,  
69 contribute more to the prevalence of allergic symptoms and related diseases than others<sup>13</sup>.

70 Whilst some grass species have been identified as more allergenic than others *in vitro*  
71 (triggering higher levels of Immunoglobulin E (IgE) antibody production), there is a high  
72 degree of cross-reactivity between grass species<sup>14-16</sup>. In addition, the allergen profiles (the  
73 characterisation of the different allergens common to different grass), the degree of  
74 sensitisation differ between grass species<sup>14,17</sup>, and the overall allergenicity of grass pollen in  
75 the air varies across seasons<sup>18</sup>. Family-level estimates of grass pollen concentrations cannot  
76 therefore be considered a reliable proxy for either the concentration of pollen-derived  
77 aeroallergens or pollen-induced public health outcomes.

78

79 The identification of biodiversity via the high-throughput analysis of taxonomic marker genes  
80 (popularly termed metabarcoding) provides an emerging solution to semi-quantitatively  
81 identify complex mixtures of airborne pollen grains<sup>19-22</sup>. Further, recent global DNA

82 barcoding initiatives and co-ordinated regional efforts have now resulted in near complete  
83 genetic databases of national native plants, including grass species. In Great Britain, the vast  
84 majority of angiosperms are included in mature DNA barcoding databases for multiple  
85 markers<sup>23</sup>, meaning that we are now in a position to investigate the aerial composition of  
86 pollen over the grass pollen season, at a national scale.

87

88 Here, using two complementary DNA barcode marker genes (*rbcL* and ITS2), we characterise  
89 the spatial and temporal distribution of airborne grass pollen throughout the temperate  
90 summer grass pollen season (May-August) across the latitudinal and longitudinal range of  
91 Great Britain (Fig. 1). We hypothesise that the composition of airborne grass pollen, from  
92 different grass taxa will be (i) broadly homogenous across the grass pollen season, regardless  
93 of terrestrial Poaceae phenology and (ii) homogenous across Great Britain due to the  
94 potential for long distance transport of windborne pollen grains.

95

96 Airborne grass pollen from each genus occupied distinct temporal windows across the grass  
97 pollen season in 2016 (May to August), thereby rejecting our hypothesis (i) (Fig. 2,  
98 Supplementary Figure 1). Time, measured as number of days after the first sample was  
99 collected, is a good predictor of airborne grass pollen taxon composition using both markers  
100 (Fig. 2, Supplementary Figure 1; *ITS2*,  $LR_{1,74} = 128.8$ ,  $P = 0.001$ ; *rbcL*,  $LR_{1,71} = 46.71$ ,  $P =$   
101  $0.001$ ). Community-level ordination reveals that the airborne grass pollen community as a  
102 whole changed across the grass pollen season (Supplementary Figure 2, Supplementary  
103 Figure 3), with similar overarching trends observed for the most abundant airborne pollen  
104 families including, Poaceae, Pinaceae and Urticaceae (Supplementary Figure 4). In addition,  
105 observations of first flowering dates from a citizen science project (UKPN;

106 [www.naturescalendar.org.uk](http://www.naturescalendar.org.uk)) and metabarcoding data show similar sequences of seasonal  
107 progression with a lag time similar to that found in observational studies<sup>11</sup> (See  
108 supplementary text; Supplementary Figure 5), suggesting that there is a link between local  
109 phenology of Poaceae and composition of airborne grass pollen.

110

111 Focusing on the more taxonomically specific ITS2 marker dataset, *Alopecurus* and *Holcus*  
112 typically dominated the early grass pollen season (Fig. 2), which coincides with typical peaks  
113 in allergic rhinitis<sup>24</sup>, but further research will be required to confirm this association. *Lolium*  
114 featured prominently for the majority of the later grass season. The popularity of *Lolium*  
115 species as a forage crop means that it is widely sown in agricultural grasslands<sup>25</sup>, although the  
116 majority of agricultural grasslands are managed by grazing silage-cutting or mowing which  
117 prevents the growth of flowering heads<sup>25</sup>. The length of time over which *Lolium* pollen  
118 dominated may be because many varieties have been bred with the potential to mature at  
119 different times throughout the year<sup>26</sup>, although it should be noted that *Lolium* species  
120 frequently hybridise with each other and therefore it is difficult to distinguish these genera  
121 using genetic material alone. Additionally, while there is some evidence that some species of  
122 grass appear to be more allergenic than others<sup>18</sup>, it is unknown how much they may differ  
123 within a species (i.e. at the cultivar/hybrid level)<sup>16</sup>. Although *Lolium* was the dominant species  
124 in airborne grass pollen from July to the end of the sampling period, the total grass pollen  
125 concentration declined in August, indicating that the absolute number of *Lolium* pollen grains  
126 at this time is low (Fig. 1, Supplementary Figure 6).

127

128 The top five genera contributing to airborne pollen, indicated by the relative abundance of  
129 taxonomy marker genes, were *Alopecurus*, *Festuca*, *Holcus*, *Lolium* and *Poa* (Fig. 2;

130 Supplementary Figure 6). Each of these genera is widespread in the UK, although long-  
131 distance pollen transport means they may also originate further afield<sup>27</sup>. These dominant  
132 genera have all been shown to provoke IgE-mediated responses in grass-sensitised patients<sup>14</sup>,  
133 providing candidate species for links with hay fever and asthma exacerbation. Conversely,  
134 less prevalent species in the dataset could contribute disproportionately to the allergenic  
135 load. Species such as *Phleum pratense* have been identified to be a major allergen<sup>5,28</sup>.  
136 However, we found that *Phleum* made up a very small proportion of metabarcoding reads  
137 (Supplementary Figure 1), corresponding with the results of an earlier phenological study<sup>9</sup>.  
138 Most genera, such as *Phleum*, *Anthoxanthum* and *Dactylis*, show distinct and narrow  
139 temporal incidence (Supplementary Figure 1), and could allow researchers to identify grass  
140 species associated with allergenic windows with greater accuracy.

141

142 Changes in species composition over time were localised. We found that peaks in abundance  
143 of airborne pollen occurred at different times at each location during the summer (Fig. 2,  
144 Supplementary Figure 1). For example, the relative abundance of airborne grass pollen from  
145 the genus *Poa* peaked in mid-June in Worcester and Bangor but 6-8 weeks later in  
146 Invergowrie (Fig. 2), probably due to latitudinal effects on flowering time<sup>7,27</sup>. This is  
147 supported by a significant interaction between latitude and time of year for both markers  
148 (Fig. 2, Supplementary Figure 1; *ITS2*,  $LR_{68,1} = 34.2$ ,  $P = 0.002$ ; *rbcl*,  $LR_{68,1} = 47.36$ ,  $P = 0.001$ ).

149 Differences in species composition of airborne grass pollen between the six sampling sites is  
150 supported by a significant effect of latitude (Fig. 2, Supplementary Figure 1; *ITS2*,  $LR_{1,73} =$   
151  $73.2$ ,  $P = 0.001$ ; *rbcl*,  $LR_{1,70} = 26.4$ ,  $P = 0.025$ ) and longitude (Fig. 2, Supplementary Figure 1;  
152 *ITS2*,  $LR_{1,69} = 33$ ,  $P = 0.003$ ; *rbcl*,  $LR_{1,69} = 27.10$ ,  $P = 0.010$ ), that are proxies for a broad range  
153 of environmental variables. These results do not support our hypothesis (ii) that the



154 composition of airborne grass pollen will be homogenous across the UK, and instead suggest  
155 that taxon-specific effects of regional geography, climate and environmental conditions  
156 underpin distributions which have been demonstrated for Poaceae pollen as a whole<sup>7</sup>.  
157 Further investigations into the mechanisms of pollen production and transport, interacting  
158 with a range of climatic, seasonal and meteorological effects will therefore provide valuable  
159 future research foci to elucidate our mechanistic knowledge of the deposition of grass pollen  
160 in time and space.

161

162 Enabled by contemporary molecular biodiversity assessment and mature, curated DNA  
163 barcoding databases, here we provide a comprehensive taxonomic overview of airborne  
164 grass pollen distribution, throughout an entire grass pollen season and across large  
165 geographic scales. The grass pollen season is defined by discrete temporal windows of  
166 different grass species, with some species displaying geographical variation. Temporal pollen  
167 distributions in metabarcoding data follow observed flowering times. The data provide an  
168 important step towards developing genera-, and in certain cases, species-level grass pollen  
169 forecasting. Additionally, the research presented here leads the way for future studies  
170 facilitating understanding of the relationships between grass pollen and disease, which have  
171 significant global public health relevance and socioeconomic importance.

172

173 **Figure 1 Location of pollen collection and temporal Poaceae concentrations and composition.**

174 Map showing location of the six sampling sites and daily Poaceae pollen concentrations  
175 (grains/m<sup>3</sup>) throughout the grass pollen season (May to August, 2016). Yellow filled circles  
176 indicate dates when pollen was collected for both observational concentrations and  
177 molecular analysis, blue circles indicate days when pollen was collected for observational

178 concentrations only and green circles indicate when pollen was collected for molecular  
179 analysis only at the Bangor site. Note that Bangor is not part of the UK pollen monitoring  
180 network and observational concentrations were only performed alongside pollen collections  
181 for molecular analysis between 24<sup>th</sup> June to 28<sup>th</sup> August 2016. Contains OS data Crown  
182 copyright and database right (2018). Image Crown Copyright, 2018, The Met Office.

183

184 **Figure 2 Abundance of the most common airborne grass pollen taxa throughout the grass**  
185 **pollen season.** The five most abundant grass taxa (expressed as proportion of total reads),  
186 depicted alongside the total proportion of reads assigned to family Poaceae. Due to errors in  
187 sampling equipment, only 4 alternate weeks (out of a possible 7 alternate weeks) of samples  
188 were collected at the York sampling site. Markers used to identify grass pollen are stated in  
189 the top panel label. Sampling sites are indicated in the right panel label abbreviated as  
190 follows: BNG = Bangor; EXE = Exeter; ING = Invergowrie; IOW = Isle of Wight; WOR =  
191 Worcester; YORK = York. A map of sampling locations and daily Poaceae pollen  
192 concentrations can be found in Figure 1.

193

## 194 **Methods**

195

### 196 **Sampling and Experimental Design**

197 We collected aerial samples from six sites across Great Britain (Supplementary Table 3; Fig. 1)  
198 using Burkard Automatic Multi-Vial Cyclone Samplers (V2; Burkard Manufacturing Co. Ltd.  
199 Rickmansworth, UK) designed to simplify collection of pollen and spores by sampling directly  
200 into a microcentrifuge tube (e.g.<sup>29</sup>). The volumetric aerial sampler uses a turbine to draw in  
201 air (16.5 litres/min) and aerial particles, using mini-cyclone technology. The aerial particles

202 are collected into 1.5 ml sterile microcentrifuge tubes located on a carousel, where the  
203 carousel is programmed to sample into a new tube every 24 h, thereby providing daily  
204 samples of airborne pollen (Supplementary Figure 7). Sample tubes were sent to Bangor, and  
205 stored at -20°C before processing. Each sampling unit was mounted alongside a seven-day  
206 volumetric trap of the Hirst design (1952) belonging to the Met Office UK Pollen Monitoring  
207 Network, which provided daily pollen concentrations (Fig. 1; map produced using ArcGIS). In  
208 the seven-day volumetric trap, a turbine draws air in (10 litres/min) and particles are  
209 impacted upon an adhesive coated tape carried on a clockwork-driven drum. The tape is cut  
210 into 24 h sections, and mounted on glass slides using a gelatine/glycerol mountant containing  
211 basic fuchsin to stain the pollen grains. Pollen are identified and counted under a microscope  
212 and converted to volumetric concentrations<sup>7</sup>. Although the high cost of the pollen samplers  
213 preclude routine replicate sampling, our methodologies mirror methodologies that have  
214 been used for several decades in the UK network<sup>30,31</sup> and are in agreement with  
215 recommended terminology described by Galan et al (2017)<sup>32</sup>. All pollen samplers were sited  
216 in elevated positions on flat-roofed buildings between 4 to 6 floors in height in order to  
217 sample from a mixed air flow. Fins on the samplers (both Burkard Multi-Vial Cyclone and  
218 Hirst type seven-day volumetric samplers) direct the cyclone inlet port into the wind. Bangor  
219 was the only sampling site which was not part of the pollen monitoring network, but we  
220 deployed the same methodology at the Bangor site (which began on 24<sup>th</sup> June 2016 Fig. 1).  
221  
222 Sampling began in late May 2016 and during alternate weeks, aerial samples were collected  
223 for seven days for a total of seven weeks between 25<sup>th</sup> May and 28<sup>th</sup> August. Exact sampling  
224 dates varied slightly between sites and a total of 279 aerial samples were collected  
225 (Supplementary Table 4).

226

## 227 **DNA Extraction, PCR and Sequencing**

228 From the 279 daily aerial samples, 231 were selected for downstream molecular analysis, as  
229 described below. Within each sampling week, two series of three consecutive days were  
230 pooled. Pooled samples were selected based on grass pollen concentrations based on  
231 microscopy. The final, unselected day was not used in downstream molecular analysis. In  
232 total, seventy-seven pools of DNA were created. In one instance, three consecutive days of  
233 pollen samples were unavailable (Invergowrie, week 2, pool 2) due to trap errors. For this  
234 sample, the next sampling day was selected for pooling (Supplementary Table 4). DNA was  
235 extracted from daily samples using a DNeasy Plant Mini kits (Qiagen, Valencia, CA, USA), with  
236 some modifications to the standard protocol as described by Hawkins *et al.*<sup>33</sup>. DNA from daily  
237 samples was pooled and eluted into 60 µl of elution buffer at the binding stage of the DNeasy  
238 Plant Mini kit.

239

240 Illumina MiSeq paired end indexed amplicon libraries were prepared following a two-step  
241 protocol. Two marker genes were amplified with universal primer pairs *rbcLaf* and  
242 *rbcLr506*<sup>23,34</sup>, and ITS2 and ITS3<sup>18</sup> (Supplementary Table 5). A 5' universal tail was added to  
243 the forward and reverse primers and a 6N sequence was added between the forward  
244 universal tail and the template-specific primer, which is known to improve clustering and  
245 cluster detection on MiSeq sequencing platforms<sup>35</sup> (Integrated DNA Technologies, Coralville,  
246 USA). Round 1 PCR was carried out in a final volume of 25 µL, including forward and reverse  
247 primers (0.2 µM), 1X Q5 HS High-Fidelity Master Mix (New England Biolabs) and 1 µL of  
248 template DNA. Thermal cycling conditions were an initial denaturation step at 98 °C for 30s;  
249 35 cycles of 98 °C for 10s, 50 °C for 30s, 72 °C for 30s; and a final annealing step of 72 °C for 5

250 minutes. Products from the first PCR were purified using Agencourt AMPure XP beads  
251 (Beckman Coulter) with a 1:0.6 ratio of product to AMPure XP beads.  
252  
253 The second round PCR added the unique identical i5 and i7 indexes and the P5 and P7  
254 Illumina adaptors, along with universal tails complementary to the universal tails used in  
255 round 1 PCR (Supplementary Table 4, Supplementary Table 5) (Ultramer, by IDT, Integrated  
256 DNA Technologies). Round 2 PCR was carried out in a final volume of 25  $\mu$ L, including forward  
257 and reverse index primers (0.2  $\mu$ M), 1X Q5 HS High-Fidelity Master Mix (New England  
258 Biolabs) and 5  $\mu$ L of purified PCR product. Thermal cycling conditions were: 98  $^{\circ}$ C for 3 min;  
259 98  $^{\circ}$ C for 30 s, 55  $^{\circ}$ C for 30 s, 72  $^{\circ}$ C for 30 s (10 cycles); 72  $^{\circ}$ C for 5 min, 4  $^{\circ}$ C for 10 min. Both  
260 PCRs were run in triplicate. The same set of unique indices were added to the triplicates  
261 which were then pooled following visual inspection on an agarose gel (1.5%) to ensure that  
262 indices were added successfully. Pooled metabarcoding libraries were cleaned a second time  
263 using Agencourt AMPure magnetic bead purification, run on an agarose gel (1.5%) and  
264 quantified using the Qubit high sensitivity kit (Thermo Fisher Scientific, Massachusetts, USA).  
265 Positive and negative controls were amplified in triplicate with both primer pairs and  
266 sequenced alongside airborne plant community DNA samples using the MiSeq. Sequence  
267 data, including metadata, are available at the Sequence Read Archive (SRA) using the project  
268 accession number SUB4136142.

269

## 270 **Bioinformatic Analysis**

271 Initial sequence processing was carried out following a modified version of the workflow  
272 described by de Vere *et al.*<sup>36</sup>. Briefly, raw sequences were trimmed using Trimmomatic  
273 v0.33<sup>37</sup> to remove short reads (<200bp), adaptors and low quality regions. Reads were

274 merged using FLASH v 1.2.11<sup>36,38</sup>, and merged reads shorter than 450bp were excluded.  
275 Identical reads were merged using fastx-toolkit (v0.0.14), and reads were split into ITS2 and  
276 *rbcL* based on primer sequences.  
277  
278 To prevent spurious BLAST hits, custom reference databases containing *rbcL* and ITS2  
279 sequences from UK plant species were generated. While all native species of the UK have  
280 been DNA barcoded<sup>23</sup>, a list of all species found in the UK was generated in order to gain  
281 coverage of non-native species. A list of UK plant species was generated by combining lists of  
282 native and alien species<sup>39</sup> with a list of cultivated plants obtained from Botanic Gardens  
283 Conservation International (BGCI) which represented horticultural species. All available *rbcL*  
284 and ITS2 records were downloaded from NCBI GenBank, and sequences belonging to UK  
285 species were extracted using the script '[creatingselectedfastadatabase.py](#)', archived on  
286 GitHub.  
287  
288 Metabarcoding data was searched against the relevant sequence database using blastn<sup>40</sup>, via  
289 the script 'blast\_with\_ncbi.py'. The top twenty blast hits (identified using the highest bit-  
290 score) were tabulated ('blast\_summary.py'), then manually filtered to limit results to species  
291 currently present in Great Britain. Reads occurring fewer than four times were excluded from  
292 further analysis. All scripts used are archived on GitHub:  
293 <https://doi.org/10.5281/zenodo.1305767>.

294

## 295 **Statistical Analysis**

296 To understand how the grass pollen composition changed with space and time, the effect of  
297 time (measured as the number of days after the first sampling date), latitude and longitude

298 of sampling location were included in a two-tailed generalized linear model using the  
299 'manyglm' function in the package 'mvabund'<sup>41</sup>. The proportion of sequences was set as the  
300 response variable; proportion data was used as this has been shown to be an effective way of  
301 controlling for differences in read numbers<sup>42</sup>. The effect of time, latitude, longitude and the  
302 interaction between time and latitude were included as explanatory variables in the models  
303 to test hypotheses (i and ii). The effect of longitude is also consistent when York, the most  
304 easterly sampling site, with missing data from mid-July until the end of the sampling period,  
305 is removed from the analysis (Supplementary Table 6).

306

307 The data best fit a negative binomial distribution, most likely due to the large number of  
308 zeros (zeros indicate that a grass genus is absent from a sample), resulting in a strong mean-  
309 variance relationship in the data (Supplementary Figure 8). The proportion of sequences was  
310 scaled by 1000 and values were converted to integers so that a generalized linear model with  
311 a negative binomial distribution could be used. Overfitting of the models was tested using  
312 'dropterm' in R, and based on the lowest Akaike Information Criterion (AIC) score, no terms  
313 were removed from the models. In addition, the appropriateness of the models was checked  
314 by visual inspection of the residuals against predicted values from the models  
315 (Supplementary Figure 9).

316

317 In order to compare the metabarcoding data with flowering time data, we used phenological  
318 records of first flowering collected in 2016 by citizen scientists from the UK's Nature's  
319 Calendar ([www.naturescalendar.org.uk](http://www.naturescalendar.org.uk)). First flowering time was compared to genus-level  
320 ITS2 metabarcoding data for three species: *Alopecurus pratensis*, *Dactylis glomerata* and  
321 *Holcus lanatus*. As grass pollen could only be reliably identified to genus level in the

322 metabarcoding data, the taxa compared may not have been exactly equivalent since both  
323 *Alopecurus* and *Holcus* contain other widespread species within the UK. However, *Alopecurus*  
324 *pratensis* and *Holcus lanatus* are the most abundant species within their respective genera.  
325 The comparison was only carried out for ITS2 data because two of the three genera were not  
326 identified by the *rbcl* marker.

327

328 NMDS ordination was carried out using package 'VEGAN' in R<sup>43</sup>, based on the proportion of  
329 total high-quality reads contributed by each grass genus, using Bray-Curtis dissimilarity  
330 (Supplementary Figure 2 and Supplementary Figure 3). Ordination is used to reduce  
331 multivariate datasets (e.g. abundances of many species) into fewer variables that reflect  
332 overall similarities between samples. A linear model was carried out using the 'lm' function  
333 within the 'stats' package in R, in order to investigate the relationship between the number  
334 of reads obtained for each genus using the *rbcl* and ITS2 marker.

335

336 **Data and materials availability:** All sequence data are available at the Sequence Read Archive  
337 (SRA) using the project accession number SUB4136142. Archived sequence data was used to  
338 generate Figure 2 Supplementary Figures 1-S6 and 8-S10). First flowering data used in  
339 Supplementary Figure 5 was obtained from Nature's Calendar, Woodland Trust and is  
340 available upon request. The sequence analysis pipeline is available at  
341 <https://github.com/colford/nbgw-plant-illumina-pipeline>.

342

### 343 **References**

- 344 1. Blackley, C.H. (London: Bailliere: London, 1873).
- 345 2. Marks, G., Pearce, N., Strachan, D. & Asher, I. *The Global Asthma Report 2014* 16–21



- 346 (2014).
- 347 3. Bauchau, V. *The European Respiratory Journal* **24**, 758–764 (2004).
- 348 4. Bousquet, P.-J. et al. *Allergy* **62**, 301–309 (2007).
- 349 5. García-Mozo, H. *Allergy* **72**, 1849–1858 (2017).
- 350 6. Emberlin, J. et al. *Aerobiologia* **16**, 373–379 (2000).
- 351 7. Emberlin, J. et al. *Grana* **33**, 94–99 (1994).
- 352 8. Mander, L., Li, M., Mio, W., Fowlkes, C.C. & Punyasena, S.W. *Proceedings of The Royal*  
353 *Society B* **280**, 20131905 (2013).
- 354 9. Kmenta, M. et al. *World Allergy Organ. J.* **10**, 31 (2017).
- 355 10. Cope, C. & Gray, A. (Botanical Society of the British Isles: London, UK, 2009).
- 356 11. Estrella, N., Menzel, A., Krämer, U. & Behrendt, H. *International Journal of*  
357 *Biometeorology* **51**, 49–59 (2006).
- 358 12. Skjøth, C.A., Sommer, J., Stach, A., Smith, M. & Brandt, J. *Clinical & Experimental*  
359 *Allergy* **37**, 1204–1212 (2008).
- 360 13. McInnes, R.N. et al. *Science of the Total Environment* **599–600**, 483–499 (2017).
- 361 14. van Ree, R., van Leeuwen, W.A. & Aalberse, R.C. *Journal of Allergy and Clinical*  
362 *Immunology* **102**, 184–190 (1998).
- 363 15. Petersen, A. et al. *Journal of Allergy and Clinical Immunology* **107**, 856–862 (2001).
- 364 16. Jung, S. et al. *PLoS ONE* **13**, 1–12 (2018).
- 365 17. Moingeon, P., Peltre, G. & Bergmann, K.-C. *Clinical and Experimental Allergy Reviews*  
366 **8**, 12–14 (2008).
- 367 18. de Weger, L.A. et al. *Clinical and Translational Allergy* **1**, 18 (2011).
- 368 19. Kraaijeveld, K. et al. *Molecular Ecology Resources* **15**, 8–16 (2015).
- 369 20. Korpelainen, H. & Pietiläinen, M. *Nordic Journal of Botany* **35**, 602–608 (2017).
- 370 21. Deiner, K. et al. *Molecular Ecology* **26**, 5872–5895 (2017).
- 371 22. Creer, S. et al. *Methods in Ecology and Evolution* **7**, 1008–1018 (2016).
- 372 23. de Vere, N. et al. *PLoS ONE* **7**, 1–12 (2012).
- 373 24. England, P.H. (PHE Real-time Syndromic Surveillance Team: Birmingham, UK, 2016).

- 374 25. Frame, J. & Laidlaw, A.S. (The Crowood Press: Marlborough, Wiltshire, UK, 2011).
- 375 26. RGCL (British Grassland Society: Cheshire, UK, 2017).
- 376 27. Rousseau, D.-D. et al. *Geophysical Research Letters* **30**, (2003).
- 377 28. D'Amato, G. et al. *Allergy and Allergen Immunotherapy* **62**, 261–279 (2017).
- 378 29. West, J.S. & Kimber, R.B.E. *Annals of Applied Biology* **166**, 4–17 (2015).
- 379 30. Adams-Groom, B., Emberlin, J., Corden, J., Millington, W. & Mullins, J. *Aerobiologia*  
380 **18**, 117–123 (2002).
- 381 31. Skjøth, C.A., Baker, P., Sadyś, M. & Adams-Groom, B. *Urban Climate* **14**, 414–428  
382 (2015).
- 383 32. Galán, C. et al. *Aerobiologia* **33**, 293–295 (2017).
- 384 33. Hawkins, J. et al. *PLoS ONE* **10**, 1–20 (2015).
- 385 34. Kress, W.J. & Erickson, D.L. *PLoS One* **2**, e508 (2007).
- 386 35. Miya, M. et al. *Royal Society open science* **2**, 150088 (2015).
- 387 36. de Vere, N. et al. *Scientific Reports* **7**, 42838 (2017).
- 388 37. Bolger, A.M., Lohse, M. & Usadel, B. *Bioinformatics* **30**, 2114–2120 (2014).
- 389 38. Magoc, T. & Salzberg, S.L. *Bioinformatics* **27**, 2957–2963 (2011).
- 390 39. Stace, C. (Cambridge University Press: Cambridge, UK, 2010).
- 391 40. Camacho, C. et al. *BMC Bioinformatics* **10**, 421 (2008).
- 392 41. Wang, Y., Naumann, U., Wright, S.T. & Warton, D.I. *Methods in Ecology and Evolution*  
393 **3**, 471–474 (2012).
- 394 42. McMurdie, P.J. & Holmes, S. *PLoS Computational Biology* **10**, e1003531 (2014).
- 395 43. Dixon, P. *Journal of Vegetation Science* **14**, 927–930 (2003).

396

### 397 Acknowledgements

398

399 We thank John Kenny, Pia Koldkjær, Richard Gregory, and Anita Lucaci of the Liverpool  
400 Centre for Genomic Research for sequencing support. We acknowledge the computational  
401 services & support of the Supercomputing Wales project, which is part-funded by the

402 European Regional Development Fund (ERDF) via Welsh Government. We thank the Botanic  
403 Gardens Conservation International (BGCI) for access to the list of plant collections in the  
404 National Gardens in the UK and Ireland. We thank the Met Office network for providing  
405 additional observational grass pollen count data and Jonathan Winn, UK Met Office for  
406 ArcGIS assistance on Figure 1. We are grateful to the Woodland Trust and Centre for Ecology  
407 & Hydrology for supplying the UK Phenology Network data and to the citizen scientists who  
408 have contributed to the latter scheme. Final thanks to Wendy Grail and technical support  
409 staff at Bangor University.

410

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

417

418 **Funding:** This work was supported by the Natural Environment Research Council  
419 (<https://nerc.ukri.org/>), awarded to SC (NE/N003756/1), CS (NE/N002431/1), NO  
420 (NE/N002105/1) and GWG/NdV/MH (NE/N001710/1). IBERS Aberystwyth receives strategic  
421 funding from the BBSRC. The funders had no role in study design, data collection and  
422 analysis, decision to publish, or preparation of the manuscript.

423

424 **Competing interests:** The authors are not aware of any competing interests.

425





