Article:
Brennan, Georgina, Potter, Caitlin, de Vere, Natasha et al. (25 more authors) (2019) Temperate airborne grass pollen defined by spatio-temporal shifts in community composition. Nature Ecology and Evolution. ISSN 2397-334X

https://doi.org/10.1038/s41559-019-0849-7

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

Authors: Georgina L. Brennan[1*], Caitlin Potter[2], Natasha de Vere[2,3], Gareth W. Griffith[2], Carsten A. Skjøth[4], Nicholas J. Osborne[5,6,7], Benedict W. Wheeler[5], Rachel N. McInnes[8], Yolanda Clewlow[8], Adam Barber[8], Helen M. Hanlon[8], Matthew Hegarty[2], Laura Jones[1,3], Alexander Kurganskiy[4], Francis M. Rowney[5], Charlotte Armitage[9], 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]

Affiliations:

[1]Molecular Ecology and Fisheries Genetics Laboratory, School of Natural Sciences, Bangor University, Bangor, Gwynedd, Wales, UK

[2]IBERS, Aberystwyth University, Aberystwyth, Ceredigion, Wales, UK

[3]National Botanic Garden of Wales, Llanarthne, Carmarthenshire, Wales, UK

[4]University of Worcester, Worcester, Worcestershire, UK

[5]University of Exeter, Truro, Cornwall, UK

[6]University of New South Wales, Sydney, New South Wales, Australia

[7]School of Public Health, Faculty of Medicine, The University of Queensland, Herston, Brisbane, Australia

[8]Met Office, Exeter, Devon, UK

[9]The Woodland Trust, Kempton Way, Grantham, Lincolnshire, UK.

[10]College of Life and Environmental Science, University of Exeter, Exeter, UK

[11]Ecological Sciences, The James Hutton Institute, Dundee, Scotland, UK

[12]The David Hide Asthma & Allergy Research Centre, St Mary's Hospital, Newport, Isle of Wight, UK

[13]Centre for Ecology & Hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire, UK.

[14]Environment Department, University of York, Heslington, York, UK

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

Grass pollen is the world's most harmful outdoor aeroallergen, yet it is unknown how airborne pollen assemblages change in time and space. Human sensitivity varies between different species of grass that flower at different times, but it is not known if temporal turnover in species composition match terrestrial flowering or if species richness steadily accumulates over the grass pollen season. Here, using targeted, high-throughput sequencing, we demonstrate that all grass genera display 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 asthma[1] and allergic rhinitis (hay fever), negatively affecting 400 million people worldwide[2]. Pollen from the grass family (Poaceae) constitutes the most significant outdoor aeroallergen[3,4], and more people are sensitised to grass pollen than to any other pollen type[5]. However, despite the harmful impact of grass pollen on human health, especially as reported in developed nations, current studies and forecasts categorize grass pollen at the family level (Poaceae)[6,7] due to difficulties in differentiating species based on morphology[8]. Furthermore, while different species of temperate grass flower at different timepoints[9,10], it is unknown if the disparate phenology of local grass taxa at ground level are 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 to local flowering times[11]. Persistence and mobility of grass pollen could result in 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 species composition over the summer

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

Many species within the subfamilies Pooideae, Chloridoideae, and Panicoideae release allergenic pollen into the atmosphere[5], including *Phleum* spp. (e.g. Timothy grasses), *Dactylis* spp. (Cocksfoot/Orchard grasses), *Lolium* spp. (Ryegrasses), *Festuca* spp. (Fescues), *Poa* spp. (Meadow-grasses and Bluegrasses), and *Anthoxanthum* spp. (Vernal grasses). Furthermore, some grass taxa, notably 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/cultivars within species, contribute more to the prevalence of allergic symptoms and related diseases than others[13]. Whilst 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 characterisation of the different allergens common to different grass), the degree of sensitisation differ between grass species[14,17], and the overall allergenicity of grass pollen in the air varies across seasons[18]. Family-level estimates of grass pollen concentrations cannot therefore be considered a reliable proxy for either the concentration of pollen-derived aeroallergens or pollen-induced public health outcomes.

The identification of biodiversity via the high-throughput analysis of taxonomic marker genes (popularly termed metabarcoding) provides an emerging solution to semi-quantitatively identify complex mixtures of airborne pollen grains[19–22]. Further, recent global DNA

barcoding initiatives and co-ordinated 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 markers[23], meaning that we are now in a position to investigate the aerial composition of pollen over the grass pollen season, at a national scale.

Here, using two complementary DNA barcode marker genes (*rbcL* and ITS2), we characterise 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 hypothesise that the composition of airborne grass pollen, from different grass taxa will be (i) broadly homogenous across the grass pollen season, regardless of terrestrial Poaceae phenology and (ii) homogenous across Great Britain due 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), thereby rejecting our hypothesis (i) (Fig. 2, Supplementary Figure 1). Time, measured as number of days after the first sample was collected, is a good predictor of airborne grass pollen taxon composition using both markers (Fig. 2, Supplementary Figure 1; *ITS2*, $LR_{1, 74} = 128.8$, $P = 0.001$; *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 Figure 2, Supplementary Figure 3), with similar overarching trends observed for the most abundant airborne pollen families including, Poaceae, Pinaceae and Urticaceae (Supplementary Figure 4). In addition, observations of first flowering dates from a citizen science project (UKPN;

106 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[11] (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[24], 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[25], although the
116 majority of agricultural grasslands are managed by grazing silage-cutting or mowing which
117 prevents the growth of flowering heads[25]. 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[26], 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[18], it is unknown how much they may differ
123 within a species (i.e. at the cultivar/hybrid level)[16]. 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;

Supplementary Figure 6). Each of these genera is widespread in the UK, although long-distance pollen transport means they may also originate further afield[27]. These dominant genera have all been shown to provoke IgE-mediated responses in grass-sensitised patients[14], 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 Figure 1), corresponding with the results of an earlier phenological study[9]. Most genera, such as *Phleum*, *Anthoxanthum* and *Dactylis*, show distinct and narrow temporal incidence (Supplementary Figure 1), and could allow researchers to identify grass species associated with allergenic windows with greater accuracy.

Changes in species composition over time were localised. We found that peaks in abundance of airborne pollen occurred at different times at each location during the summer (Fig. 2, Supplementary Figure 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 due to latitudinal effects on flowering time[7,27]. This is supported by a significant interaction between latitude and time of year for both markers (Fig. 2, Supplementary Figure 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, Supplementary Figure 1; *ITS2*, $LR_{1, 73}$ = 73.2, P = 0.001; *rbcL*, $LR_{1, 70}$ = 26.4, P = 0.025) and longitude (Fig. 2, Supplementary Figure 1; *ITS2*, $LR_{1, 69}$ = 33, P = 0.003; *rbcL*, $LR_{1, 69}$ = 27.10 P = 0.010), that are proxies for a broad range 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[7].
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^3) 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

concentrations only and 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 24th June to 28th August 2016. Contains OS data Crown copyright and database right (2018). Image Crown Copyright, 2018, The Met Office.

Figure 2 Abundance of the most common airborne grass pollen taxa throughout the grass pollen season. The five most abundant grass taxa (expressed as proportion of total reads), depicted alongside the total proportion of reads assigned to family Poaceae. 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. Markers used to identify grass pollen are stated in the top panel label. 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.

Methods

Sampling and Experimental Design

We collected aerial samples from six sites across Great Britain (Supplementary Table 3; Fig. 1) using Burkard Automatic Multi-Vial Cyclone Samplers (V2; Burkard Manufacturing Co. Ltd. Rickmansworth, UK) designed to simplify collection of pollen and spores by sampling directly into a microcentrifuge tube (e.g.[29]) . The volumetric aerial sampler uses a turbine to draw in 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[7]. 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[30,31] and are in agreement with
215 recommended terminology described by Galan et al (2017)[32]**.** 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 24th 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 25th May and 28th August. Exact sampling
224 dates varied slightly between sites and a total of 279 aerial samples were collected
225 (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 based on microscopy. The final, unselected day was not used in downstream molecular analysis. In total, seventy-seven pools of DNA were created. In one instance, three consecutive days of pollen samples were unavailable (Invergowrie, week 2, pool 2) due 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 kits (Qiagen, Valencia, CA, USA), with some modifications to the standard protocol as described by Hawkins *et al.*[33]. 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 *rbcL*af and *rbcL*r506[23,34], 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, which is known to improve clustering and cluster detection on MiSeq sequencing platforms[35] (Integrated DNA Technologies, Coralville, USA). Round 1 PCR was carried out in a final volume of 25 µL, including forward and reverse primers (0.2 µM), 1X 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 30s; 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

minutes. Products from the first PCR were purified using Agencourt AMPure XP beads (Beckman Coulter) with a 1:0.6 ratio of product to AMPure XP beads.

The second round 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 round 1 PCR (Supplementary Table 4, Supplementary Table 5) (Ultramer, by IDT, Integrated DNA Technologies). Round 2 PCR was carried out in a final volume of 25 µL, including forward and reverse index primers (0.2 µM), 1X 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, Massachusetts, USA). 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.

Bioinformatic Analysis

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

merged using FLASH v 1.2.11[36,38], and merged reads shorter than 450bp were excluded. Identical reads were merged using fastx-toolkit (v0.0.14), and reads were split into ITS2 and *rbcL* based on primer sequences.

To prevent spurious BLAST hits, custom reference databases containing *rbcL* and ITS2 sequences from UK plant species were generated. While all native species of the UK have been DNA barcoded[23], a list of all species found in the UK was generated in order to gain coverage of non-native species. A list of UK plant species was generated by combining lists of native and alien species[39] with a list of cultivated plants obtained from Botanic Gardens Conservation International (BGCI) 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 'creatingselectedfastadatabase.py', archived on GitHub.

Metabarcoding data was searched against the relevant sequence database using blastn[40], via the script 'blast_with_ncbi.py'. The top twenty blast hits (identified using the highest bit-score) were tabulated ('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 grass pollen composition changed with space and time, the effect of 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'[41]. 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[42]. 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). 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

metabarcoding data, the taxa compared may not have been exactly equivalent since both *Alopecurus* and *Holcus* contain other widespread species within the UK. However, *Alopecurus pratensis* and *Holcus 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.

NMDS ordination was carried out using package 'VEGAN' in R[43], based on the proportion of total high-quality reads contributed by each grass genus, using Bray-Curtis dissimilarity (Supplementary Figure 2 and Supplementary Figure 3). Ordination is used to reduce multivariate datasets (e.g. 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, in order to investigate the relationship between the number of reads obtained for each genus using the rbcL and ITS2 marker.

Data and materials 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 Supplementary Figures 1-S6 and 8-S10). First flowering data used in Supplementary Figure 5 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.

References

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

(2014).

3. Bauchau, V. *The European Respiratory Journal* **24**, 758–764 (2004).

4. Bousquet, P.-J. et al. *Allergy* **62**, 301–309 (2007).

5. García-Mozo, H. *Allergy* **72**, 1849–1858 (2017).

6. Emberlin, J. et al. *Aerobiologia* **16**, 373–379 (2000).

7. Emberlin, J. et al. *Grana* **33**, 94–99 (1994).

8. Mander, L., Li, M., Mio, W., Fowlkes, C.C. & Punyasena, S.W. *Proceedings of The Royal Society B* **280**, 20131905 (2013).

9. Kmenta, M. et al. *World Allergy Organ. J.* **10**, 31 (2017).

10. Cope, C. & Gray, A. (Botanical Society of the British Isles: London, UK, 2009).

11. Estrella, N., Menzel, A., Krämer, U. & Behrendt, H. *International Journal of Biometeorology* **51**, 49–59 (2006).

12. Skjøth, C.A., Sommer, J., Stach, A., Smith, M. & Brandt, J. *Clinical & Experimental Allergy* **37**, 1204–1212 (2008).

13. McInnes, R.N. et al. *Science of the Total Environment* **599–600**, 483–499 (2017).

14. van Ree, R., van Leeuwen, W.A. & Aalberse, R.C. *Journal of Allergy and Clinical Immunology* **102**, 184–190 (1998).

15. Petersen, A. et al. *Journal of Allergy and Clinical Immunology* **107**, 856–862 (2001).

16. Jung, S. et al. *PLoS ONE* **13**, 1–12 (2018).

17. Moingeon, P., Peltre, G. & Bergmann, K.-C. *Clinical and Experimental Allergy Reviews* **8**, 12–14 (2008).

18. de Weger, L.A. et al. *Clinical and Translational Allergy* **1**, 18 (2011).

19. Kraaijeveld, K. et al. *Molecular Ecology Resources* **15**, 8–16 (2015).

20. Korpelainen, H. & Pietiläinen, M. *Nordic Journal of Botany* **35**, 602–608 (2017).

21. Deiner, K. et al. *Molecular Ecology* **26**, 5872–5895 (2017).

22. Creer, S. et al. *Methods in Ecology and Evolution* **7**, 1008–1018 (2016).

23. de Vere, N. et al. *PLoS ONE* **7**, 1–12 (2012).

24. England, P.H. (PHE Real-time Syndromic Surveillance Team: Birmingham, UK, 2016).

25. Frame, J. & Laidlaw, A.S. (The Crowood Press: Marlborough, Wiltshire, UK, 2011).

26. RGCL (British Grassland Society: Cheshire, UK, 2017).

27. Rousseau, D.-D. et al. *Geophysical Research Letters* **30**, (2003).

28. D'Amato, G. et al. *Allergy and Allergen Immunotherapy* **62**, 261–279 (2017).

29. West, J.S. & Kimber, R.B.E. *Annals of Applied Biology* **166**, 4–17 (2015).

30. Adams-Groom, B., Emberlin, J., Corden, J., Millington, W. & Mullins, J. *Aerobiologia* **18**, 117–123 (2002).

31. Skjøth, C.A., Baker, P., Sadyś, M. & Adams-Groom, B. *Urban Climate* **14**, 414–428 (2015).

32. Galán, C. et al. *Aerobiologia* **33**, 293–295 (2017).

33. Hawkins, J. et al. *PLoS ONE* **10**, 1–20 (2015).

34. Kress, W.J. & Erickson, D.L. *PLoS One* **2**, e508 (2007).

35. Miya, M. et al. *Royal Society open science* **2**, 150088 (2015).

36. de Vere, N. et al. *Scientific Reports* **7**, 42838 (2017).

37. Bolger, A.M., Lohse, M. & Usadel, B. *Bioinformatics* **30**, 2114–2120 (2014).

38. Magoc, T. & Salzberg, S.L. *Bioinformatics* **27**, 2957–2963 (2011).

39. Stace, C. (Cambridge University Press: Cambridge, UK, 2010).

40. Camacho, C. et al. *BMC Bioinformatics* **10**, 421 (2008).

41. Wang, Y., Naumann, U., Wright, S.T. & Warton, D.I. *Methods in Ecology and Evolution* **3**, 471–474 (2012).

42. McMurdie, P.J. & Holmes, S. *PLoS Computational Biology* **10**, e1003531 (2014).

43. Dixon, P. *Journal of Vegetation Science* **14**, 927–930 (2003).

Acknowledgements

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

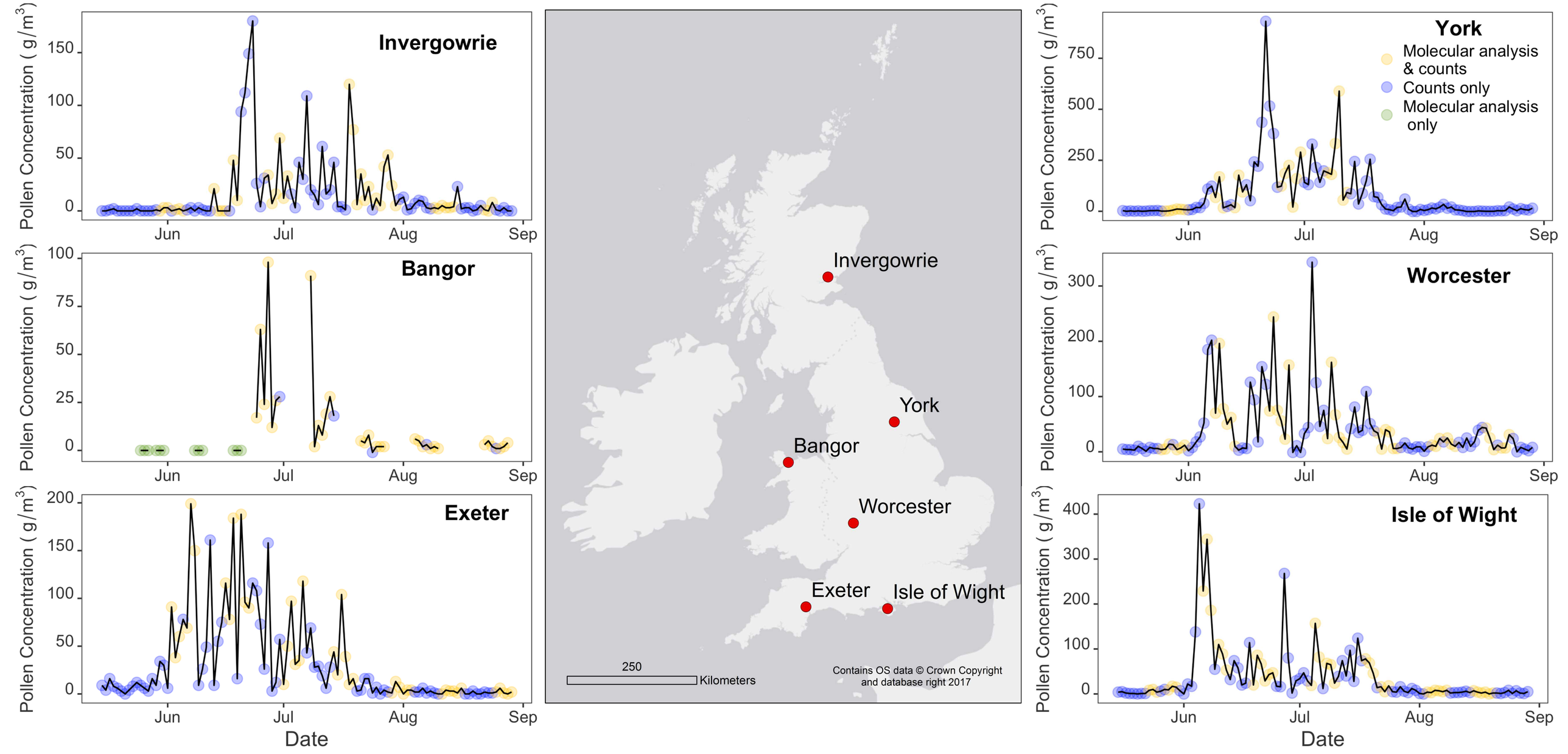
European Regional Development Fund (ERDF) via Welsh Government. We thank the Botanic Gardens Conservation International (BGCI) for access to the list of plant collections in the National Gardens in the UK and Ireland. We thank the Met Office network for providing additional observational grass pollen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

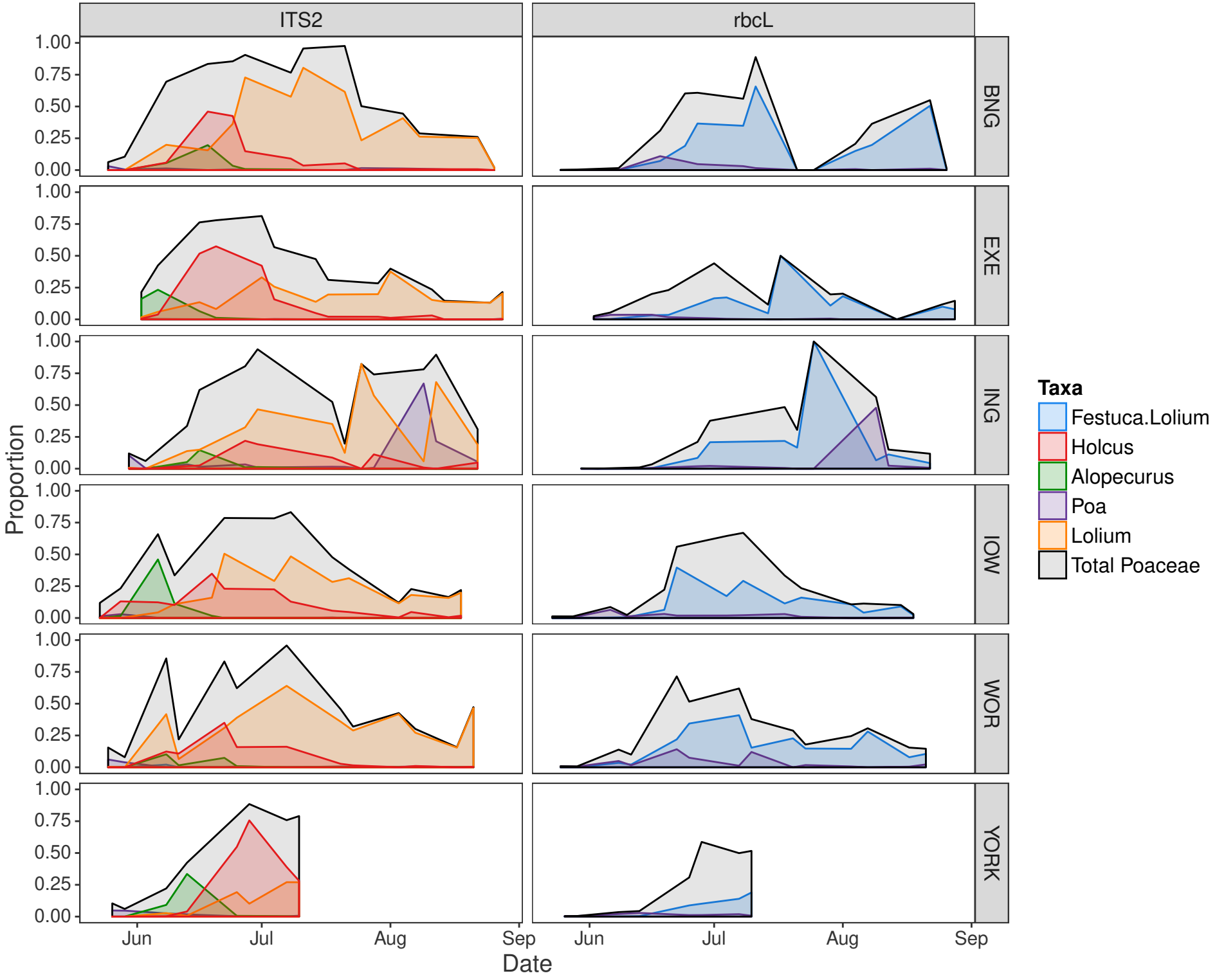
Author Contributions: S.C., N.dV., G.W.G., R.N.M., N.J.O., C.A.S., Y.C., B.W.W. and G.L.B. conceived and designed the study; B.A-G., G.L.B., G.P., A.E., R.N., S.P., K.S., and N.S. collected samples and counted pollen; G.L.B. performed laboratory work, supported by S.C.; N.dV., C.P., C.R.F., L.J., S.C. contributed methods, and 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.

Funding: This work was supported by the Natural Environment Research Council (https://nerc.ukri.org/), awarded to SC (NE/N003756/1), CS (NE/N002431/1), NO (NE/N002105/1) and GWG/NdV/MH (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.

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

426


Invergowrie
Bangor
Exeter
York
Worcester
Isle of Wight
Molecular analysis & counts
Counts only
Molecular analysis only
Pollen Concentration (g/m3)
Jun
Jul
Aug
Sep
Date
250
Kilometers
Contains OS data © Crown Copyright and database right 2017


ITS2
rbcL
BNG
EXE
ING
IOW
WOR
YORK
Proportion
1.00
0.75
0.50
0.25
0.00
Jun
Jul
Aug
Sep
Date
Taxa
Festuca.Lolium
Holcus
Alopecurus
Poa
Lolium
Total Poaceae