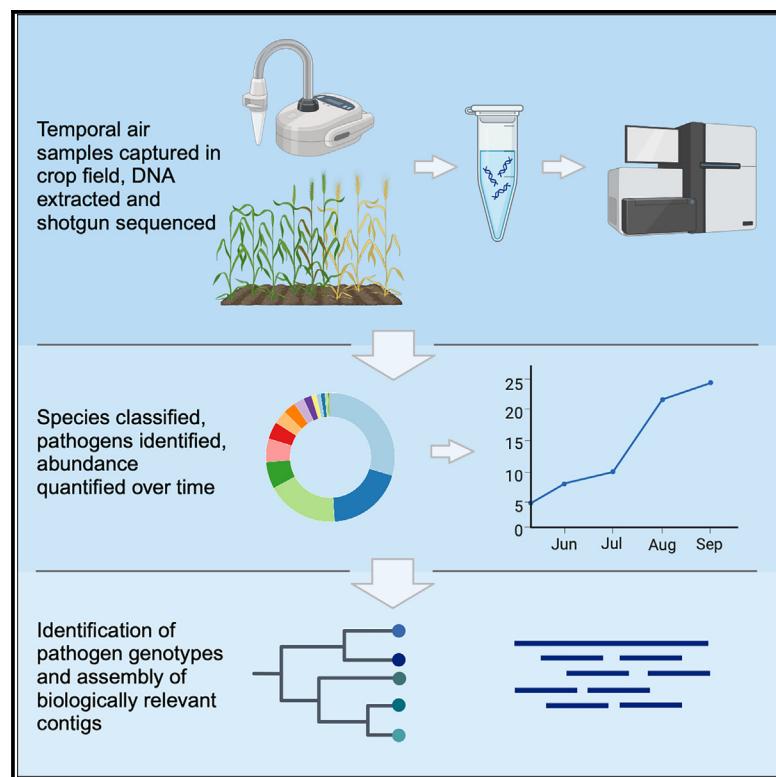


Current Biology

Measuring air metagenomic diversity in an agricultural ecosystem

Graphical abstract



Authors

Michael Giolai, Walter Verweij,
Samuel Martin, Neil Pearson,
Paul Nicholson, Richard M. Leggett,
Matthew D. Clark

Correspondence

richard.leggett@earlham.ac.uk (R.M.L.),
matt.clark@nhm.ac.uk (M.D.C.)

In brief

Giolai et al. demonstrate how airborne DNA in a crop field can be recovered, shotgun sequenced, and taxonomically classified to the species level. Giolai et al. find dynamically changing pathogen (spore) levels across time, and these often correlate with the weather. Pathogens can be classified to their closest strain using known genetic markers.

Highlights

- Airborne DNA can be recovered, shotgun sequenced, and classified to the species level
- Crop pathogen levels in the air are dynamic across a growing season
- Weather conditions correlate with changing pathogen spore levels
- Airborne pathogen spores' DNA can be matched to their closest strain



Article

Measuring air metagenomic diversity in an agricultural ecosystem

Michael Giolai,^{1,5} Walter Verweij,^{2,4} Samuel Martin,² Neil Pearson,² Paul Nicholson,³ Richard M. Leggett,^{2,*} and Matthew D. Clark^{1,6,*}

¹Natural History Museum, London SW7 5BD, UK

²Earlham Institute, Norwich Research Park, Norwich NR4 7UZ, UK

³Crop Genetics Department, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

⁴Enza Zaden, Enkhuizen 1602 DB, the Netherlands

⁵Research Centre for Ecological Change, Organismal and Evolutionary Biology Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki 00014, Finland

⁶Lead contact

*Correspondence: richard.leggett@earlham.ac.uk (R.M.L.), matt.clark@nhm.ac.uk (M.D.C.)

<https://doi.org/10.1016/j.cub.2024.07.030>

SUMMARY

All species shed DNA during life or in death, providing an opportunity to monitor biodiversity via environmental DNA (eDNA). In recent years, combining eDNA, high-throughput sequencing technologies, bioinformatics, and increasingly complete sequence databases has promised a non-invasive and non-destructive environmental monitoring tool. Modern agricultural systems are often large monocultures and so are highly vulnerable to disease outbreaks. Pest and pathogen monitoring in agricultural ecosystems is key for efficient and early disease prevention, lower pesticide use, and better food security. Although the air is rich in biodiversity, it has the lowest DNA concentration of all environmental media and yet is the route for wind-borne spread of many damaging crop pathogens. Our work suggests that ecosystems can be monitored efficiently using airborne nucleic acid information. Here, we show that the airborne DNA of microbes can be recovered, shotgun sequenced, and taxonomically classified, including down to the species level. We show that by monitoring a field growing key crops we can identify the presence of agriculturally significant pathogens and quantify their changing abundance over a period of 1.5 months, often correlating with weather variables. We add to the evidence that aerial eDNA can be used as a source for biomonitoring in terrestrial ecosystems, specifically highlighting agriculturally relevant species and how pathogen levels correlate with weather conditions. Our ability to detect dynamically changing levels of species and strains highlights the value of airborne eDNA in agriculture, monitoring biodiversity changes, and tracking taxa of interest.

INTRODUCTION

Air is rich in biodiversity, containing particles of prokaryotic and eukaryotic organisms such as pollen, matter from animals, and a variety of bacterial and fungal species.^{1–5} Wind patterns are important in influencing climate,⁶ species dispersal, species biogeography,^{7–9} and plant as well as animal health.^{10,11} Especially when anthropogenic actions cause dramatic effects on the environment, measures to understand processes and species fluxes influencing ecosystems become of increasing importance.¹² One means to help characterize such changes is the utilization of nucleic acids found as environmental DNA (eDNA).¹³ eDNA studies have become increasingly valuable in combination with high-throughput sequencing technologies, where they are helpful in increasing sample throughput at a reduced cost and facilitate profiling of environments in a non-destructive, non-invasive manner.^{13–15}

Plant agricultural ecosystems are, as a source of food, essential for human needs.¹⁶ Modern farming is mostly

practiced in large monocultures that, due to their size, are difficult to monitor efficiently.^{17,18} Disease in these monocultures (if not managed effectively) can easily spread and have devastating effects on yield and food security.^{18–20} Disease propagation and outbreaks are not solely based on close-distance transmission of pathogens to neighboring plants, but also on long-distance aerial dispersal of pathogenic material, such as fungal spores between fields or larger areas.⁷ Disease detection in large agricultural patches remains challenging, as large areas of farmland need to be surveyed with methods sensitive enough to detect disease at the earliest stages (e.g., at the time point of arrival or at an asymptomatic stage) in order to avoid further systemic diffusion or the occurrence of substantial disease symptoms that lead to agricultural losses.¹⁹ Early disease monitoring is therefore crucial to prevent and mediate pathogen damage.⁵ Nowadays, such challenges are addressed with a multitude of technologies, ranging from chemical and molecular to internet of things (IoT) (i.e., systems of interconnected digital devices) technologies.²⁰ However,



many current pathogen detection technologies rely on the detection and analysis of visual plant damage. Although highly specific molecular tools (e.g., PCR²¹ or antibody-based assays²²) can test for single pathogens, next-generation sequencing-based approaches can be of great potential to overcome these limitations in throughput, providing the means for detection of many possible pathogens before disease establishment and at a resolution that can distinguish between pathogen strains.

One promising approach to monitor disease onset in agroecosystems is to monitor the species composition of air,⁵ as many diseases propagate through close- or long-distance aerial transmission^{7,23} and airborne particles such as spores can be collected with commercial filtration devices. Approaches for automated sequencing library preparation do exist (e.g., Tecan MagicPrep). The combination of such technologies with a simple data analysis platform—e.g., Oxford Nanopore Technologies' cloud-based "What's in my pot?"—could provide government bodies, farmers, and other users with the possibility to monitor pathogen levels throughout a growing season and to prevent diseases establishing or spreading.

Air has already been a regular subject of study in the past as a means of understanding the species dynamics and composition of ecosystems.^{1–4,10,11,24–28} Already, targeted air eDNA sequencing methods have proven their value in detecting plant pathogens: Aguayo et al.¹⁰ reported that the ash dieback causing fungus *Hymenoscyphus fraxineus* is present at higher frequencies in air samples of forests with confirmed ash dieback disease. Nicolaisen et al.²⁹ reported finding fungal crop pathogens in air samples taken close to fields. Van der Heyden et al. review the application of spore sampling systems in agriculture for air monitoring and discuss the diagnostic potential of air-sampling technologies.⁵

Recent technological advances have led to an increase in activity in the field of air metagenomics: Gusareva et al. described diurnal cycles of metagenomes from bioaerosols,^{2,3} Drautz-Moses et al.⁴ investigated the vertical stratification of bioaerosols, Luhung et al.³⁰ report an ultra-low biomass pipeline for bioaerosol sequencing, and Qin et al.²⁶ and Leung et al.²⁵ study air metagenomes in great detail in densely populated urban environments and large city microbial transport systems. Altogether, the reported studies show that metagenomics approaches in combination with air sampling can be highly useful in profiling ecosystems.^{2–4,25,26,30}

We therefore hypothesized that monitoring air metagenomes in agricultural ecosystems will show a rich dynamic of pests and pathogens associated with the cultivated crops. With precision monitoring, such data could contribute to strategies for minimizing pesticide use and more targeted deployment of existing measures. To test our hypothesis, we established a custom workflow, hereafter termed "AirSeq," utilizing a field-work-compatible, small, battery-powered and portable air-sampling device (i.e., the Coriolis Micro Microbial Air sampler) and a next-generation sequencing workflow preparing aerial eDNA shotgun metagenome sequencing libraries (Figure 1A). Using AirSeq, we show the metagenome dynamics in the months of June and July in a field growing wheat barley and pea close to the Earlham Institute at the Norwich Research Park in Norwich, United Kingdom (UK) (Figure S1).

RESULTS

Detection of *Bacillus thuringiensis* spores in a wind tunnel experiment

To establish AirSeq, we first tested our method in a wind tunnel experiment. A wind tunnel provided us with an environment that was semi-controlled (with some air exchange with surroundings) that could be used to deliver biological particles toward the air-sampling device by putting them into the upstream airflow. As commercially available mock communities can contain organisms that are allergenic or harmful to humans when sprayed as a fine aerosol into air, and mindful of experimental safety, we decided to release *Bacillus thuringiensis* spores, which are regarded as safe even at high levels and are often used by organic farmers.³¹ We hypothesized that releasing *B. thuringiensis* at various concentrations during air collection in the wind tunnel would allow us to test for accuracy and semi-quantitative abilities. As *Bacillus* spores are particularly resistant to extraction methods, successful lysis of *Bacillus* spores will allow us to also detect other microorganisms.^{32–34} Bioinformatic classification methods in calling taxa from shotgun metagenomics data are well described, do not differ between taxa, and are not influenced by the number of species.^{35,36} Hence, semi-quantitative detection of a single taxon will allow us to extrapolate the workflow to an entire metagenome. We continuously sprayed *B. thuringiensis* spores at a total number of 0 (i.e., water as a background control), 0.3, 30, and 300 million spores over a time period of 10 min at a distance of 5 m from the air sampler (total liquid volume used per experiment: 3 L, continuous spray rate: 300 mL/min, wind speed: 2 m/s). For each spore concentration, we collected two replicates with 10 min sampling time. For each replicate, we prepared an Illumina Nextera sequencing library and sequenced the libraries to an average of 3.60 ± 1.65 million 150 base pair (bp) single-end reads (Figure 1B). In the background control, we could not observe the genus *Bacillus* (Figure 1C; Data S1). Releasing increasing amounts of spores, we observed an increase in the relative abundance of the genus *Bacillus* in the samples compared with other detected genera (Data S1). For the comparison of spore numbers and *Bacillus* loads, we calculated a Pearson correlation coefficient of 0.97 and R-squared value of 0.94: 0% of reads were assigned as *Bacillus* with the release of 0 million spores, 0.88% of reads with 0.3 million spores, 3.59% of reads with 3 million spores, 22.02% of reads with 30 million spores, and 70.04% of reads with 300 million spores (Figure 1D). Despite obtaining a Pearson correlation coefficient of 0.97 when comparing the released spore concentration with the relative abundance of *Bacillus* calls, we observed that the two 0.3-million spore replicates and a single 3-million spore sample did not produce many reads in sequencing. This was due to the low final library concentration for these samples. Whereas we obtained an average yield of 66 ng per library, one of the 0.3-million sample yielded just 2.5 ng and two others were not quantifiable. This was likely due to poor DNA extraction efficiency from low biomass samples. Nevertheless, as we still obtained sequencing reads from the samples, we decided to keep them in our analysis.

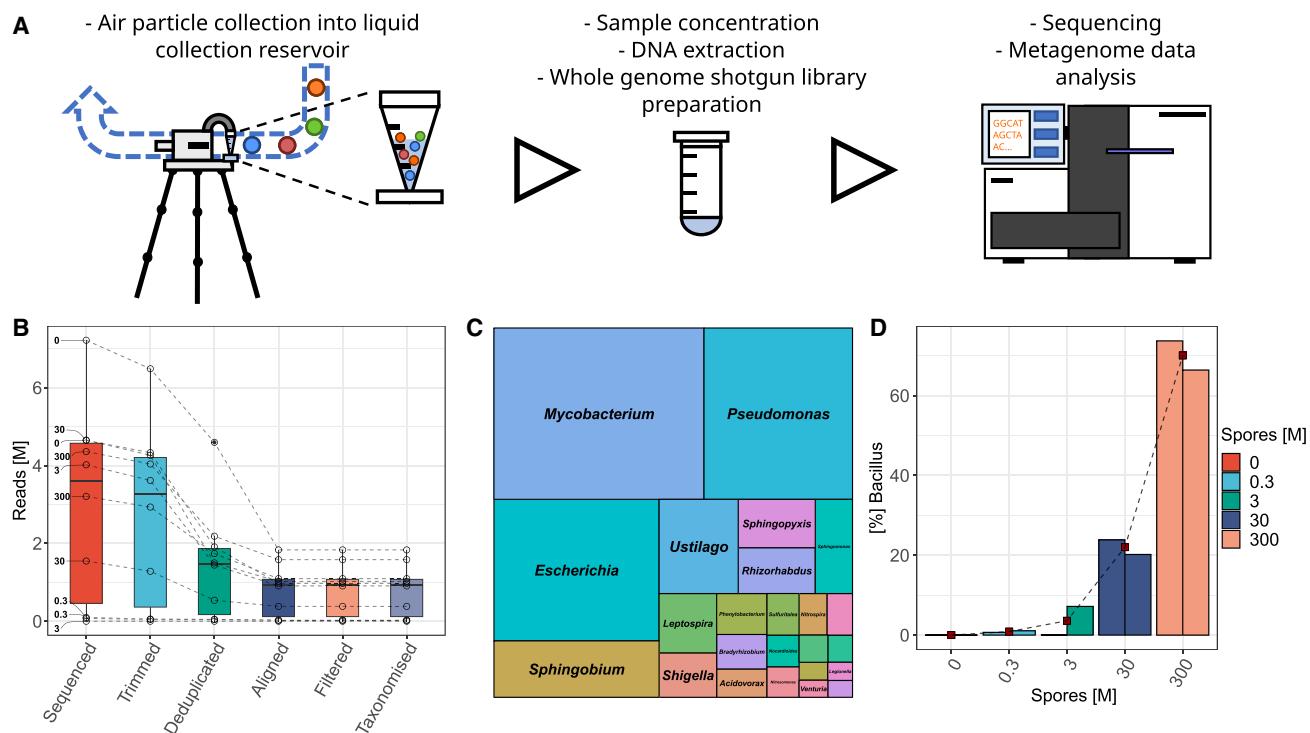


Figure 1. AirSeq performance in a wind tunnel experiment while releasing *B. thuringiensis* spores

(A) AirSeq workflow: air is filtered for a defined time period through a liquid collection buffer reservoir. After sampling the liquid collection buffer is concentrated, the eDNA extracted from the sample and a whole-genome shotgun library prepared using a low-input DNA compatible Illumina Nextera protocol. After library preparation the sample is ready for sequencing and data analysis.

(B) Read numbers (million) over all analysis steps, from number of sequenced, trimmed, deduplicated, diamond aligned, quality filtered, and taxonomically assigned reads. The spore concentration in million released spores is shown as the bold number to the left of the “sequenced” boxplot (sequenced: number of sequenced reads, trimmed: number of quality \geq Q20 and length \geq 75 bp filtered reads, deduplicated: number of reads after deduplication, aligned: number of reads aligning to the nr database, filtered: number of reads aligning to NCBI nr database references with an e value $\leq 1e-10$, taxonomized: reads that were successfully taxonomized based on the NCBI nr database taxon).

(C) The background metagenome of the wind tunnel; the larger the box, the more normalized reads of a specific genus were detected.

(D) Detection curve of percent *Bacillus* in the classified reads of each spore concentration. For each spore concentration two replicates were sampled. When releasing only water (0 million spores) we could not detect any *Bacillus* calls. With an increase in the released spore number the abundance of *Bacillus* in the libraries rose rapidly.

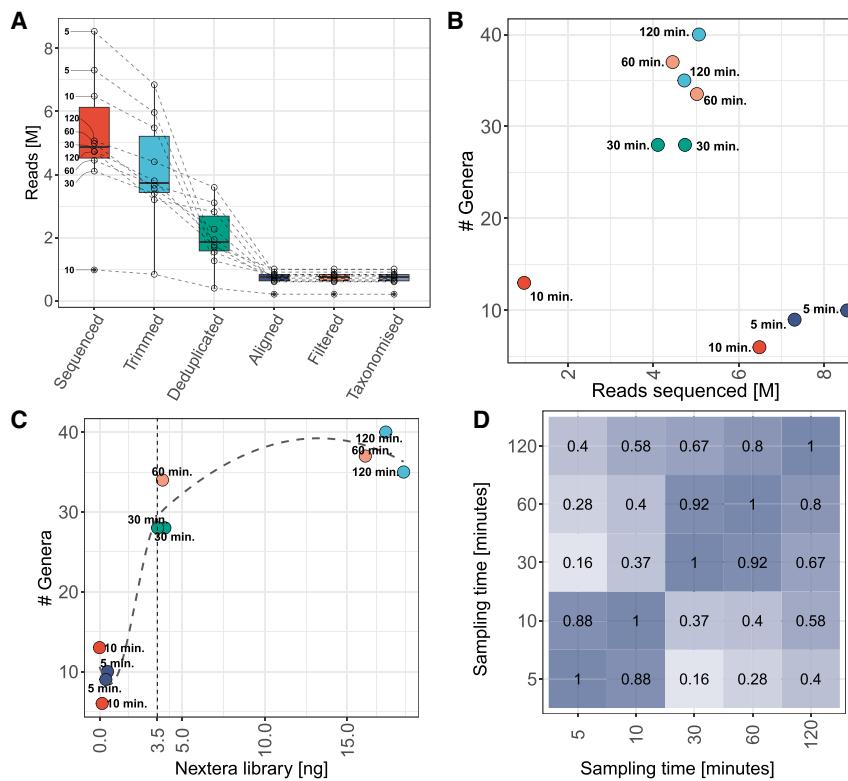
See also [Data S1](#).

Determination of a suitable air-sampling time window for in-field metagenome measurements

As we observed in the wind tunnel control experiment that our AirSeq method can be utilized to detect released organisms in a semi-quantitative manner, we moved onto testing air collected in the field. To determine which sampling time window would deliver best library preparation results and most taxa, we prepared a collection time-series sampling air for 5, 10, 30, 60, and 120 min, consecutively, on the same day at a Norwich Research Park field site close to the Earlham Institute ($52^{\circ}37'21.1''N$, $1^{\circ}13'02.6''E$). After commencing each sampling round, we walked away from the sampling site to minimize human interference. For each time point, we collected two samples. We prepared Illumina Nextera sequencing libraries from the samples and sequenced the libraries to 5.14 ± 1.94 million 150-bp single-end reads (Figure 2A).

To determine a suitable time period for future outdoor collections, we counted the number of classified genera for each time point. We found the fewest genera at 5- and 10-min collection

time (10 and 8, as well as 13 and 6 for each sample and time point, respectively) and observed a steep increase in genus numbers from 10 to 30 min sampling, with 28 detected genera in both samples at 30 min. This number increased even further to 34 and 37 genera in the 60-min samples and 35 and 39 genera in the 120-min samples (Figures 2B and 2C; [Data S1](#)). To determine whether the final Illumina library concentration or the sequencing depth contribute to the library complexity (i.e., a higher number of genera in a library), we analyzed the number of detected genera per library, along with the number of sequenced reads (Figure 2B) and the obtained amount of final Illumina library (Figure 2C). We found that the nanograms of the constructed Illumina library correlated with the number of observed genera and that libraries containing ≥ 3.5 ng showed the highest number of genera (to obtain a library yield of ≥ 3.5 ng, we had to sample air for ≥ 30 min at 200 L per min) (Figure 2C). Although we sequenced some libraries with less than 3.5 ng total yield to a much greater depth than all other samples (Figures 2B and 2C), these samples still contained fewer



(D) Correlation analysis comparing the RPM per detected genus of all time points (Pearson correlation coefficients are shown in the boxes). In particular the 30- and 60-min libraries compared well. One of the 120-min samples was collected during heavy afternoon rain (but on the same day as all others) potentially affecting correlation analysis to 30- and 60-min sampling times.

See also [Data S1](#).

genera than higher-yielding libraries with less sequencing. This indicates that the more metagenomic DNA is present as starting material for the Nextera fragmentation reaction, the higher the likely complexity of a library will be ([Figure 2C](#)).

We compared the samples using Pearson correlation analysis, using the sum of the normalized reads per million (RPM) per detected genus of both samples for each collection time point. The samples are biological and not technical replicates, as they were collected subsequently on the same day. We started by collecting the 5 replicates (one after another) first, and then proceeded with the 10-, 30-, 60-, and 120-min replicates. We observed that the 30- and 60-min sampling time points correlated best with a Pearson correlation coefficient of 0.92 ([Figure 2D](#)). For the 30 and 60- to 120-min samples, we observed a drop in the correlation coefficient to 0.67 and 0.80, respectively. One of the 120-min samples, however, was affected by heavy afternoon rain, potentially inducing a bias in sample composition and so reducing the correlation coefficient.

For future sampling, we decided to collect all air samples for 60 min, considering that in the 60-min time window we found almost as many genera as in the 120 min sampling, the Coriolis air sampler can be powered for 60 min with one battery load, and the observation that the short time (5 and 10 min) air collection samples suffered from low library concentrations ([Figure 2C](#)) and higher read duplication rates ([Figure 2A](#)).

Figure 2. AirSeq performance in a field experiment studying various sampling time

(A) Number of reads (million) along the bioinformatic analysis workflow from sequenced to classified reads. The black, bold number next to the “sequenced” boxplot indicates the sampling time in minutes of a library. Libraries where air was collected for ≤ 10 min and that were sequenced to a high number of reads show higher read duplication rates (sequenced: number of sequenced reads, trimmed: number of quality $\geq Q20$ and length ≥ 75 bp filtered reads, deduplicated: number of reads after deduplication, aligned: number of reads aligning to the nr database, filtered: number of reads aligning to NCBI nr database references with an e value $\leq 1e-10$, taxonomized: reads that were successfully taxonomized based on the NCBI nr database taxon).

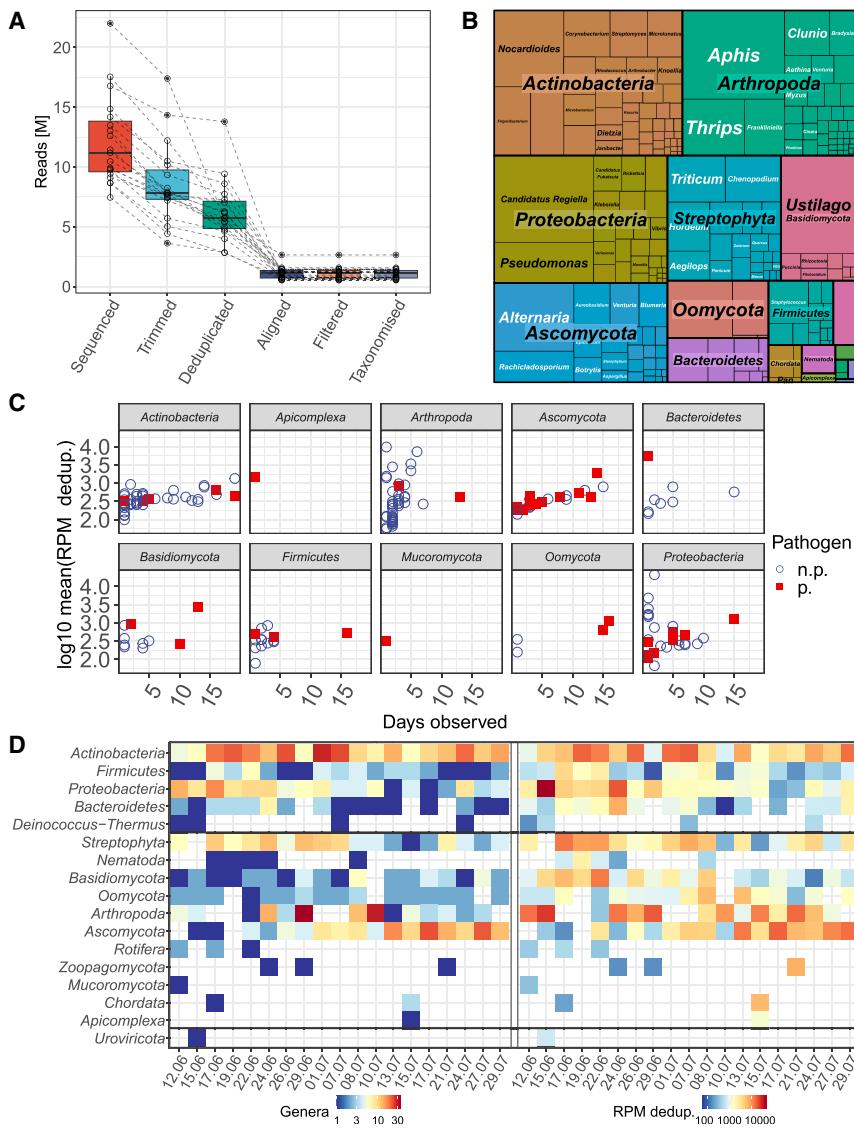
(B) Number of detected genera per million sequenced reads. We observed that a high number of reads in libraries sampled for ≤ 10 min does not increase the number of detected genera, whereas in libraries with air-collection times ≥ 30 min we observe most genera.

(C) Number of detected genera compared with the initial library concentration. The longer the sampling time (for our analyzed time points), the higher the final concentration of an Illumina library. The higher the final Illumina library concentration, the more genera we found. We observed the steepest increase in detected genera for Illumina Nextera libraries between 0 and 3.5 ng total yield (smoothed conditional means fitted in R).

Analysis of agricultural field metagenome temporal dynamics

To widely characterize air metagenome dynamics over multiple weeks in an in-field scenario, we collected air for 60 min, 3 times a week, in duplicates (except for one day, when we took four samples throughout the day), from June 12 to July 29, 2015, at the same coordinates ($52^{\circ}37'21.1''N$, $1^{\circ}13'02.6''E$) as used for the collection time-series experiment. Samples were collected on a Monday, Wednesday, and Friday, with the few exceptions described in [Table S1](#), at the same coordinates as described above. Air metagenome composition in tropical ecosystems has been reported to follow a diurnal cycle.² To consider potential metagenome alterations throughout the day, we maintained a fixed sampling window between 9 and 12 a.m., with few exceptions ([Table S1](#)). All samples were prepared as Illumina Nextera libraries and sequenced to 5.76 ± 1.88 million 150-bp single-end reads ([Figure 3A](#)). We classified all samples ([Data S1](#)) and compared all replicates using their relative abundance in an all vs. all replicate correlation analysis ([Figure S4](#)). We found that most replicates sampled on the same day compared well with a mean and a median Pearson correlation coefficient of 0.67 and 0.84, respectively, for all samples ([Table S2](#)).

The Coriolis air-sampling device was placed in a field immediately surrounded by wheat and barley, as well as peas at a wider distance ([Figure S1](#)). We measured a distance of 15 m from our



sampling spot to the edge of the barley plot, 18 m to the closest wheat field and 61 m to the pea field (Figure S1). We searched the dataset for the presence of the genera *Triticum*, *Hordeum*, and *Pisum*. We did not observe *Pisum*, which is self-fertilizing, closed-flowering, does not release pollen,³⁷ and so was expected not to be present among the identified taxa. We found *Triticum* to be the most highly abundant plant genus, with 2.24% of all normalized counts, and *Hordeum* to be the third-most abundant plant genus, with 1.87% of all normalized reads, immediately after *Chenopodium* (2.04% of normalized reads) (Figure 3B; Table S5). *Chenopodium* is an outcrossing species, a major producer of pollen and a major source of allergens from late summer to early autumn.³⁸ *Chenopodium album* (the fat-hen) was also independently identified by the Norfolk Flora group to be present in our sampling area (occurrence IDs can be controlled under 2cd4p9h.h9mcev and 2cd4p9h.h8c49e on <https://nbnatlas.org/>).

We detected reads from the eukaryote and bacterial domains, as well as a single double-stranded DNA virus (Figures 3B and

Figure 3. AirSeq experiment at a field site growing wheat, barley, and pea crops with air metagenome testing over approximately 1.5 months

(A) Number of reads (million) of each sequencing library along the bioinformatic analysis steps undertaken from the number of sequenced reads, to the number of classified reads (sequenced: number of sequenced reads, trimmed: number of quality \geq Q20 and length \geq 75 bp filtered reads, deduplicated: number of reads after deduplication, aligned: number of reads aligning to the nr database, filtered: number of reads aligning to NCBI nr database references with an e value \leq 1e-10, taxonomized: reads that were successfully taxonomized based on the NCBI nr database taxon).

(B) Abundance (RPM) of all detected phyla over the studied time period; the larger the box, the higher the abundance of a phylum among all classified reads. The boxes within the colored phyla boxes show the abundance of the detected genera within a phylum.

(C) Phyla with described pathogenic genera filtered using the PHI-base pathogen and pest database.³¹ Genera are shown as \log_{10} of the RPM normalized counts over how many days they have been observed by AirSeq. Red squares indicate a genus containing at least one pathogen listed in PHI-base (p.), blue circles a genus that does not have any pathogenic species listed in PHI-base (n.p.). As a particular pathogen rich phylum, we identified the Ascomycota.

(D) Number of observed genera (left) and RPM counts (right) for the observed phyla. From top to bottom the phyla are grouped in bacteria, eukaryotes, and viruses.

See also Data S1.

3C). In detail, we observed one viral (taxon *Lederbergvirus*—a phage³⁹), 5 bacterial, and 11 eukaryotic phyla. As phylogenetic groups with the highest relative abundance over our entire time series, we found

Actinobacteria (20.5% of the total relative abundance, 44 genera), *Arthropoda* (18.8% of total relative abundance, 45 genera), *Proteobacteria* (16.6% of total relative abundance, 30 genera), *Ascomycota* (12.0% of total relative abundance, 26 genera), *Streptophyta* (10.6% of total relative abundance, 23 genera), *Basidiomycota* (7.0% of total relative abundance, 10 genera), and *Oomycota* (4.3% of total relative abundance, 4 genera) (Figure 3B; Data S1). Although these groups counted multiple genera, we found that within each group a few genera were assigned the majority of normalized counts: for *Actinobacteria* this was *Nocardioides*. In *Proteobacteria*, we found *Candidatus regiella* and *Pseudomonas*, the two genera with highest levels of normalized counts. For *Arthropoda*, these were *Aphis* and *Thrips*. For *Ascomycota*, we found *Alternaria* as the genus with the highest relative abundance. *Streptophyta* counted the aforementioned *Triticum*, *Chenopodium*, and *Hordeum*, and *Basidiomycota* contained *Ustilago* as the most abundant genus. For *Oomycota*, we found *Peronospora* as the highest represented genus (Figure 3B; Data S1). This indicates the ability of AirSeq

to successfully capture relevant pathogens to its environment across bacterial, insect, fungal, and oomycete clades; e.g., *Pseudomonas*⁴⁰ (bacteria), *Aphis*⁴¹ (insects), *Thrips*⁴² (insects), *Alternaria*⁴³ (fungi), *Ustilago*⁴⁴ (fungi), and *Peronospora*⁴⁵ (oomycete) are all described to contain plant pathogenic members. We therefore analyzed our data for the presence of pathogenic genera using the information of the 279 pathogen species listed in the Pathogen Host Interaction database (PHI-base) at <http://phi-base.org/>.⁴⁶ We scored a genus as potentially pathogenic if the PHI-base contained at least one species described as a pathogen (Data S1). According to these filtering criteria, we found the phylum Ascomycota to be a taxon of many hits (Figure 3C). Among the overall 26 detected ascomycete genera, we found *Alternaria*, *Aspergillus*, *Bipolaris*, *Blumeria*, *Botrytis*, *Cercospora*, *Claviceps*, *Fusarium*, *Microdochium*, *Pseudocercospora*, *Pyrenophora*, *Venturia*, and *Zymoseptoria* to contain at least a single species listed as pathogenic in the PHI-base.⁴⁶ Studying counts over time (Data S1) in our data, we observed that the ascomycete relative abundance oscillated with elevated counts and genus numbers in the second half of our sampling time series. This potentially indicates the detection of a fungal pathogenic episode by AirSeq (Figures 3C and 3D; Data S1).

Identification and analysis of pathogen species

To characterize our data in greater detail and identify pathogen species instead of genera, we constructed a smaller, custom database containing a single NCBI RefSeq or GenBank genome per pathogen species of the PHI-base database.⁴⁶ Querying large databases (e.g., NCBI nr) with metagenomic reads provides the advantage of querying a wide range of taxa. However, the breadth of available accessions in large-sequence databases in combination with short-read lengths makes bioinformatic classification of reads to the species level challenging.^{35,47} One means to overcome this is to use custom databases containing selected species of interest combined with adequate filtering. We searched our custom database using BWA-MEM,⁴⁸ considering hits mapping with $\geq 95\%$ identity and $\geq 95\%$ matched sequence exclusively to a single species. We further removed all ambiguous mapping reads by filtering for MAPQ > 0. We defined that species will have to be described by mapping to ≥ 10 loci to consider coverage breadth and an abundance threshold of 0.05% (i.e., removing species not supported by 0.05% of the classified reads within a sample³⁵).

Using this approach, we detected 58 of 271 PHI-base pathogen species and *Puccinia triticina*, which we manually added to the database, in our data (Data S1). We found that most detected pathogens (48 of 59) colonize plant hosts (Figures 4A–4C; Data S1). Plant pathogens outweighed non-plant pathogens when comparing total relative abundances. We found that 98.44% of all relative counts were assigned to potential plant pathogens and 49.63% of all RPM were assigned to pathogens colonizing wheat, barley, or pea. At highest abundances, we found the barley pathogenic basidiomycete fungus *Ustilago hordei*,⁴⁴ the wheat and barley pathogenic ascomycete fungus *Blumeria graminis*,⁴⁴ the aphid species *Myzus persicae* (a wide-ranging generalist insect that can colonize wheat and barley),³¹ and *Botrytis cinerea*, a necrotrophic fungus with a host range expanding to more than 200 plant species⁴⁹ (Figures 4A and 4B;

Data S1). The remaining RPM classified the few non-plant pathogenic species or species with a host range that, according to the PHI-base, extended beyond plants, and showed *Aspergillus fumigatus*, *Pseudomonas syringae*, *Staphylococcus epidermidis*, and *Staphylococcus saprophyticus* as most abundant taxa of their category (Figure 4C; Data S1). *A. fumigatus* has been reported to commonly reside in compost,⁵⁰ *P. syringae* is also described as a plant pathogen,⁴⁰ *Staphylococcus saprophyticus* can be isolated from the environment,⁵¹ and *Staphylococcus epidermidis*, while of most interest in combination with human hosts, can also occur on various mammals, such as pigs,⁵² which are actively farmed in Norfolk and account for 29% of the English pig population.⁵³

We counted the observed pathogen species by plant host genus. Most detected plant pathogens we found associated with Solanaceae (including, e.g., potatoes and tomatoes, 22 pathogen species in total) and Poaceae pathogens (including wheat and barley, 21 pathogen species in total). Fabaceae (including peas), we found to be the 4th-ranked genus with the most detected pathogen species (8), right after Brassicaceae (e.g., containing cabbages and kale plants) (Figure 4C). Specifically, we found 10 wheat, 4 barley, and 2 pea pathogen species (Figure 4D; Data S1). This indicates that AirSeq found pathogens agreeing with the crop genera that were grown at our sampling site. Other hosts than wheat, barley, and pea are actively farmed in agriculture-rich Norfolk, e.g., potatoes, tomatoes, cabbages, and kale.^{54,55} Thus, the detection of pathogens colonizing these genera cannot be deemed to be false positives, as these species could have been transported by wind to our sampling area.

Studying relative abundances over time (Figure 4F; Data S1), we found *B. graminis* and *U. hordei* to be the most present during our sampling time period. In agreement with our previous finding indicating oscillating ascomycete presence, which was elevated at the end of our sampling period, we called the ascomycetes *Pyrenophora tritici-repentis*, *Zymoseptoria tritici*, *Pyrenophora teres*, *Parastagonospora nodorum*, and *Fusarium culmorum* with highest respective values between July 13 and 29 (Figure 4B).

As weather and, in particular, humidity and temperature, dramatically influences disease levels,^{56,57} we searched for pathogen profiles over time that associate with the weather data. For this, we measured the temperature, humidity, rainfall, wind speed, and wind direction between June 17 and July 29 at our sampling site. Additionally, we documented the weather and wind direction observed while sampling (Figure 5A; Table S1). Although we observed mostly dry weather, with a few exceptions of light rain when sampling (to avoid interference of rain with sampling), our weather data showed three periods of rain and elevated humidity in mid-June, mid-July, and at the end of July (Figure 5A). Humidity and temperature (elevated and moderate, respectively) are important variables influencing infection success of plant pathogens.^{56,57} To match the weather data with pathogen occurrence profiles over time we correlated wheat, barley, and pea pathogen RPM values with weather variables (Figure 5A) for pathogens occurring more than 5 times in our sampling series. The occurrence profiles of pathogens correlated well with the average measured humidity and the average 24-h rainfall (both important for disease progression⁵⁷)

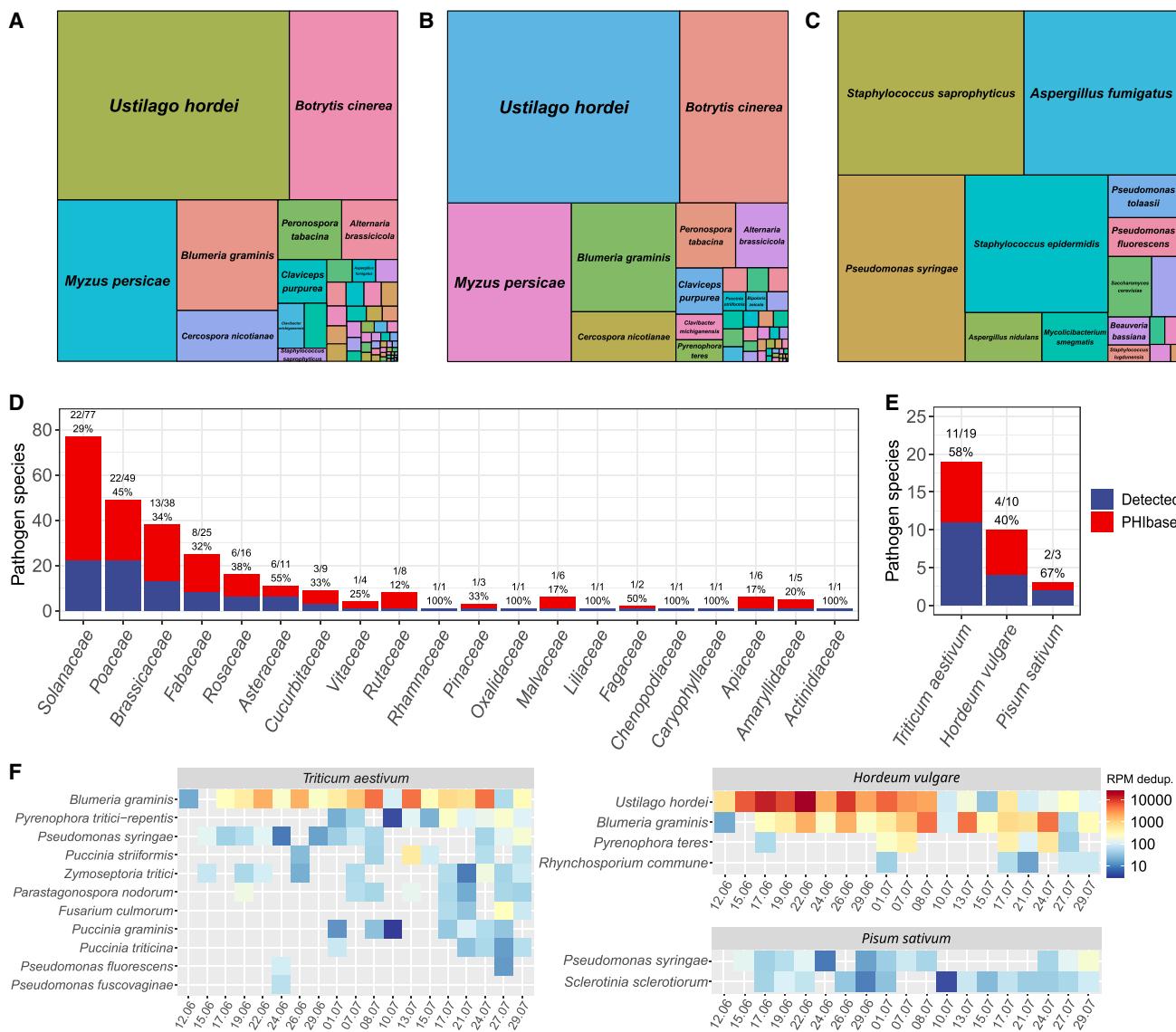


Figure 4. Pests and pathogens detected for various plant genera, wheat, barley, and pea

- (A) Treemap showing all detected pathogens scaled by relative abundance, i.e., the larger an area, the more RPM we found for a species over the entire time series. Each area shows and is colored by a species.
- (B) Treemap showing the subset of detected plant pathogens scaled by the relative abundance of a species. The figure largely resembles the treemap shown in (A) as plant pathogens amount to the majority of detected data.
- (C) Treemap showing the subset of pathogens with non-plant, or extending beyond plant, hosts (according to PHI-base). The areas are scaled to the relative abundance of non-plant pathogens and colored per species.
- (D) The number of detected pathogens and pests for the various plant host families of the PHI-base database. The red bar shows the number of pathogens described in the PHI-base database per host family, the blue bar shows the number of observed pathogens in our AirSeq data.
- (E) The number of pathogens and pests described in the PHI-base database (red) for wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), and pea (*Pisum sativum*) and the number of observed species in our AirSeq data (blue).
- (F) The RPM normalized counts over time of all observed pests and pathogens described to colonize wheat, barley, and pea. The species are sorted by the sum of the RPM counts over time from top to bottom.

See also Data S1.

(Figures 5B and 5C); e.g., *Zymoseptoria tritici* spore release has been described to positively correlate with rainfall and humidity,⁵⁸ the same has been reported for *B. graminis*⁵⁹ (Figure 5C). Both correlated with a Pearson correlation coefficient of 0.68 and 0.53, respectively, with mean daily humidity and Pearson correlation coefficients of 0.74 and 0.37 over the total 24-h

rainfall. For both, we found a negative correlation with the average daily temperature (Pearson correlation coefficients of -0.53 for *Z. tritici* and -0.29 for *B. graminis*). Both species are reported to show reduced disease severity and spore release efficiency with increasing temperatures.^{58,59} We found similar humidity-dependent patterns for, e.g., *Fusarium culmorum*,

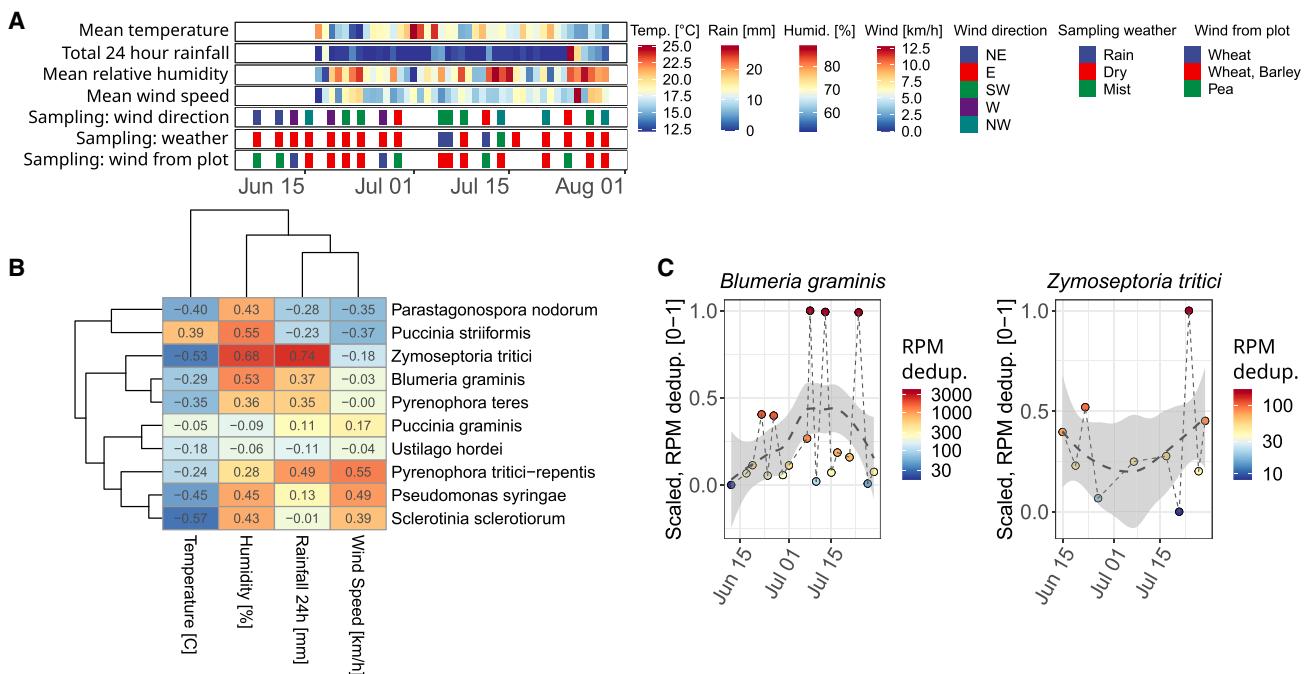


Figure 5. AirSeq pathogen profiles following weather variables

(A) Measured weather variables at our sampling site for the air-sampling time series (blank values indicate non-available weather variables, these data are not available due to a technical fault of the weather station).
(B) Heatmap showing Pearson correlation coefficients of wheat, barley, and pea pathogen RPM values correlated with weather variables (for pathogens occurring >5 times in our sampling series). For humidity, we observed higher positive correlation values for some pathogens than for other measured weather variables.
(C) Manually selected species that follow measured weather variables well: *B. graminis* correlated well with humidity and rainfall. When comparing peaks of *B. graminis* detection with humidity levels in (A), higher levels of *B. graminis* agree well with periods of elevated humidity. We observed a similar pattern for *Z. tritici*, which occurred at a higher relative abundance toward the end of the sampling time period where we observed elevated levels after a rain event and a period of elevated humidity.

Rhynchosporium commune, *Puccinia striiformis*, and *Sclerotinia sclerotiorum*.^{60–63}

Assessment of air metagenome pathogen strain information

As whole-genome sequencing in the past has been successfully applied to distinguish between races and strains of plant pathogens,⁴⁰ we hypothesized that identification of single-nucleotide polymorphisms (SNPs) for selected species in our AirSeq data, and comparing the variants to a reference database of described, strain-specific nucleotide variants, would enable us to analyze pathogen dynamics at the highest resolution. We tested our hypothesis in a proof-of-concept experiment by matching AirSeq SNP data with SNPs obtained from 39 strains of the obligate fungus and yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* (PST).⁶⁴ The 39 PST strains were characterized by Hubbard et al. in 2013,⁶⁴ in and close to East Anglia, a geographical area that includes our sampling site. *Puccinia* spores can also passively travel across large distances and thus are likely to be captured by air sequencing technologies.^{23,65,66} We also visibly detected PST infections in our field site (Figure S3).

To compare our AirSeq data for overlaps with SNPs of the 39 field strains, we mapped the AirSeq sequencing reads to the *PST-130* EnsemblFungi genome version *PST-130_1.0*,⁶⁷ filtered

reads for a % identity and % matched-sequence score ≥ 95 , an MAPQ value > 0 , and called SNPs. We excluded all variants with a quality score less than 20 and a read depth less than 20. We also mapped the data of the 39 field strains to the *PST-130_1.0* genome using BWA-MEM⁴⁸ with default parameters (as described in Hubbard et al.) and called SNPs with the same filtering thresholds as for the AirSeq data. We analyzed the transition to transversion ratios of the AirSeq data with *Puccinia striiformis* calls and the 39 field isolates and found that the values of both datasets agreed well and were slightly larger than 2 (2.68 ± 0.42 for AirSeq and 2.06 ± 0.49 for the PST isolates). To find which PST strains were captured by AirSeq, we searched for sequence changes that were uniquely assigned to single PST strains among the 329 variants from AirSeq. We found 15 unique variants shared with the 13/123 strain. Other PST strains shared 1–4 SNPs with the AirSeq data (Data S1), potentially indicating that the 13/123 strain or an isolate related to the PST 13/123 strain was captured by AirSeq. The 13/123 strain was collected by Hubbard et al. in Lincolnshire, a county close to our sampling area (~40 miles to closest point).

Classification of metagenome assemblies

Assembling short reads into contigs has the advantage that longer sequences of DNA contain more information. Long contigs span larger genomic areas, and thus taxonomic

assignments can be performed with higher precision.^{47,68} Using short-read classification, we previously detected organisms we deemed as important in our AirSeq experiment due to their biological relevance in an agricultural field as well as their abundance in our sequencing data. These are the plants *T. aestivum*, *H. vulgare*, and *C. album*, and the fungal plant pathogens *B. graminis*, *U. hordei*, and *P. striiformis*. We hypothesized that the presence of these species in our sequencing dataset should allow contigs to be assembled that can be aligned to the respective reference genomes of these organisms. We therefore performed a metagenome assembly of each collected sample using MEGAHIT.⁶⁹ On average, we obtained 15,985 [33–78,742] contigs per sample, assembling an average total amount of 2.88 Mbp [39.40 kbp–16.20 Mbp]. The contigs we obtained had a mean size of 7,256 bp [925 bp–35.17 kbp] and an average N50 of 5,163 bp [615 bp–48.62 kbp]. We aligned the contigs to the reference genomes of interest using BLAST+.⁷⁰ The obtained results were visualized using Alvis,⁷¹ showing mapping positions and coverage along each reference genome. For all references we observed genome coverage at multiple loci across the genome, further indicating the presence of the selected species in AirSeq, demonstrating that the technology can also be used to capture large genetic fragments of pathogens from air. All genome coverage maps of the described species are shown as Alvis output in Data S2.

Controlling for human contaminations across all samples

For all air-sampling experiments, the Coriolis Micro Microbial Air sampler had to be started manually. Although we immediately left the sampling site after turning on the device, we tested for potential human loads in our samples to determine human interference. In a few samples (5 of 60), we observed an average human load of 1.70% (0.78%–3.08%), four of which (nos. 8, 30, 40, and 47) were associated with our field sampling series and one (no. 11) with the wind tunnel series.

DISCUSSION

The discipline of air biome studies has existed for decades, but in recent years there has been considerable technical activity, often using metagenomic approaches.^{2–4,25,26,30} This includes studies characterizing metagenomes of the atmosphere in the tropics,² airborne microbiome communities in West Siberia,³ the microbiome composition of the lower troposphere,⁴ and urban areas.^{25,26} In contrast to species-specific markers or amplification of targeted regions for taxonomic profiling,⁷² recent shotgun metagenome sequencing strategies in combination with air biome testing allow the sampling of all genomic regions of all organisms present in a sample.^{2–4,25,26,30} Air biome studies of ecosystems can improve our understanding of the spatial, temporal, and environmental determinants of species distributions and, in detail, disease epidemiology, ultimately allowing us to model and forecast disease development for threat evaluation.⁵

Especially in agriculture, a rapid, non-invasive, and comprehensive air sequencing technology could provide advantages in disease control.^{5,10,29} Plants are simultaneously exposed to a wide variety of pests and disease-causing organisms, ranging

from viruses, bacteria, fungi, and oomycetes to insects, gastropods, and nematode pests.⁷³ Many plant pathogens and pests are not restricted to local, small areas but spread autonomously (e.g., migrating insects) or passively (e.g., insect vectors, wind, transport, and trade) over considerable distances, even trans-continental.^{7,23,74,75} As modern farming is mostly practiced in large monocultures of genetically similar hosts, the spread of novel or damaging pathogens can, and has, led to devastating yield losses threatening food security.⁷⁶

The value of marker-based disease detection in agricultural air biomes has already been shown for specific diseases.^{10,29} Here, we show that airborne shotgun metagenomes of agroecosystems are a source of diverse biotic data that can be utilized for efficient disease monitoring of pests and pathogens at a species- and strain-level resolution. We do so by collecting air using a battery-powered, mobile air-sampling device, followed by DNA extraction, low DNA input, and whole-metagenome Illumina sequencing library preparation and analysis.

We present a custom workflow enabling researchers to reproduce all steps, from collecting air samples to preparing sequencing-ready, shotgun metagenome Illumina libraries. We benchmarked our AirSeq workflow assessing its quantitative ability and accuracy. When releasing *B. thuringiensis* spores in a wind tunnel (i.e., a confined environment), we found that AirSeq could detect the relative abundance of *Bacillus* as we increased spore releases into the airstream, and detected no *Bacillus* in the no-spore controls.

We found air-sampling time to be an important variable determining the genus richness of a sample. In an in-field experiment, we collected air for increasing amounts of time, from 5 to 120 min, consecutively, on the same day, observing that longer collections increase the number of observed genera in samples. We successfully prepared and sequenced libraries from 5- and 10-min collection time, detecting at least 9 and 6 genera in the samples, but found that sampling times ≥ 30 min are important to increase the observed genus numbers to 28 and more. The final library concentrations also increased with the collection time, reflecting the increased DNA inputs. The results therefore indicate that the longer the air sampling, the more DNA for library preparation is collected. For our experiments, we decided to collect for a consecutive time of 60 min, i.e., filtering 12,000 L of air at a flow rate of 200 L per min using the Coriolis Micro air sampler. We found that this is a convenient time period, using one battery charge of our device, with which we were able to detect as many genera as the 120-min collection window allowed. The necessary collection time, however, can vary depending on the environmental conditions, the desired application (e.g., rapid air testing or continuous sampling), and the air-sampling device used. Collection systems from other manufacturers filter air using higher rates than 200 L per min and so could reduce collection times (e.g., SASS 4100 Dry Air Sampler with over 4,000 L per min; Research International, Monroe, WA, USA).

To study air metagenome dynamics over time, we collected air 3 times a week (Monday, Wednesday, and Friday) between June 12 and July 29 in an agricultural environment growing wheat, barley, and pea. For this time period, we observed one viral (taxon *Lederbergvirus*—a phage³⁹), 5 bacterial, and 11 eukaryotic phyla. As the detected *Lederbergvirus* genome is composed of double-stranded DNA that our fragmentase-based library

preparation approach uses for reaction,⁷⁷ the organism was able to be sequenced. Library preparation technologies optimized for RNA or single-stranded DNA could potentially also capture single-stranded DNA or even RNA viruses.

As the phyla with highest relative abundance, we found the prokaryotes *Actino-* and *Proteobacteria*, as well as the eukaryotic *Arthropoda*, *Ascomycota*, *Basidiomycota*, and *Streptophyta*. For some eukaryotic phyla, we observed temporal dynamics such as a higher number of genera during specific time periods (e.g., *Streptophyta* at the beginning of our time series, oscillating *Ascomycota* abundances elevated at the end of our time series) and varying counts over time (e.g., *Arthropoda* counts, in relation to their time series, peaking on specific days). Among the characterized genera, we also found the crops growing at our study site in high abundance. We observed *Hordeum* (includes barley) and *Triticum* (includes wheat) as the first- and third-most abundant next to the genus *Chenopodium*. We also show that it is possible to perform metagenome assemblies and identify *T. aestivum*, *H. vulgare*, and *C. album* from contig data.

We analyzed our AirSeq data for pathogen species using the host-pathogen information contained in the PHI-base⁴⁶ database. We observed 58 of 271 PHI-base pathogens and *P. tritici*. Representative for the studied environment, we counted 48 of 59 plant pathogens (the 48 plant pathogens contributed to 98.44% of all counts). Most pathogens that we detected were described to colonize main crops (e.g., potatoes, tomatoes, wheat, barley, and pea) and relatives that are reported to be present in the agricultural area around Norwich.^{55,78} In detail, we found 11 wheat, 4 barley, and 2 pea pathogen species.

Especially for the grass crops, we observed important and devastating species, such as the wheat and barley powdery mildew causing *B. graminis*⁷⁹ and the wheat *Septoria tritici* blotch disease causing *Z. tritici*.⁸⁰ Powdery mildew (*B. graminis*) and *Septoria tritici* blotch (*Z. tritici*) are among the most common diseases affecting wheat in the UK. We also detected *U. hordei* as the most abundant barley pathogen. *U. hordei* is the cause of covered smut of barley, and common in the UK. Similar to *B. graminis* and *Z. tritici*, it was found at high levels in AirSeq samples, therefore demonstrating how crops are threatened for a prolonged period, spanning the duration of flowering, when they are most susceptible to infection. Among the wheat pathogens we further found *P. tritici-repentis* and *Parastagonospora nodorum*. *P. tritici-repentis* and *P. nodorum*, respectively, cause tan spot and *Septoria nodorum* blotch of wheat.⁸¹ The former disease is becoming more common, particularly in the warmer, southern parts of the country. The latter disease was once common in the UK but has been very largely replaced by *Septoria tritici* blotch caused by *Z. tritici*. A PCR-based analysis of leaves of wheat in the UK found up to 100% of samples to contain DNA of *Z. tritici*, with 20% and 30% of samples, respectively, containing pathogens *P. tritici-repentis* or *P. nodorum*.⁸² The detection of *P. tritici-repentis* and *P. nodorum* in the AirSeq samples indicates that these pathogens are present on alternative hosts from which spores are released and that could pose a threat to UK wheat crops if environmental conditions are conducive to infection.⁷⁵ We also detect *F. culmorum*, which is one of the two commonly found species (next to *F. graminearum*, which we do not detect) on flowering wheat spikes in the UK.⁸³

We could relate pathogen detection to weather events. We observed multiple fungal crop pathogen (in particular ascomycetes) species that coincide with elevated humidity and rainfall. This is interesting, as the release of ascospores from perithecia is associated with periods of high humidity.⁸⁴ We also found some fungal pathogens to coincide well with the measured wind speed. As fungal spores can spread by wind,^{7,23} this could indicate the dispersal of spores either from or to our measurement location. We did not observe similar trends for the measured average daily temperature. As changing weather conditions can impact on air metagenome composition and also influence air DNA sampling itself,^{56,85,86} some of the variation we observed in our correlation experiment can likely be attributed to technical constraints, posing limits on our data.

Among the most devastating wheat diseases threatening food production globally are three rusts: *Puccinia graminis* f. sp. *tritici* (stem rust), *Puccinia triticina* (leaf rust), and PST (stripe rust).⁸⁷ Indicating the efficacy of AirSeq in detecting diseases, we observed low abundances of *P. triticina*, *P. striiformis*, and *P. graminis* on selected days of our sampling time series. *P. striiformis* was the most abundant of the three rusts we found growing on leaves in the field.

The detection of *P. graminis* in the AirSeq samples is of particular significance. Stem rust, caused by this pathogen, has been absent from the UK for almost 60 years, but was again observed on a single plant in 2013. Since this time, there have been additional reports of stem rust caused by this pathogen.⁸⁸ This indicates that the pathogen may not be uncommon and suggests that the relative rarity of stem rust may have been due to environmental factors not being conducive to widespread infection.

Strain-level epidemiology is important to monitor and measure disease threats,^{64,89} and as strains vary in their ability to overcome different cultivars' resistances, is useful information. *De novo* identification of strains using metagenomics remains challenging. Assays require good coverage rates⁸⁹ and sufficient available genomic diversity of a pathogen to link SNPs to proper genotypes. Although this information is not available for many plant pathogens to date, we hypothesized that comparison of nucleotide variants to a database containing strains collected in England would allow insights into the presence of pathogen strains at a shallow sequencing depth. As a reference dataset for a proof-of-concept experiment to characterize PST strains, we used the RNA sequencing dataset established by Hubbard et al., containing 39 PST UK isolates.⁶⁴ The study was performed close to our sampling area. PST is widespread across the western hemisphere, and in the past, new strains have emerged with expanded virulence profiles and higher aggressiveness able to overcome resistance genes in, e.g., European germplasm.^{90,91} Hubbard et al. report a shift in UK PST strains with the introduction of a diverse set of new exotic lineages in the last two decades.⁶⁴

In our analysis, we compared SNPs obtained from mapping our AirSeq reads and the data of Hubbard et al. to the PST-130 reference genome. After quality filtering (% identity and % matched-sequence cutoff ≥ 95 for mapped reads, MAPQ > 0, quality threshold > 20 for variant filtering, >20 depth), we found 15 of 154 variants in our AirSeq data that could be uniquely assigned to the 13/123 PST strain sampled in Lincolnshire, a county close to our sampling site, by Hubbard et al.⁶⁴ As our sampling was performed on a single site and over a limited

time period, making definitive conclusions about the origin of the detected strain (i.e., locally or from neighboring or farther away areas) remains a challenge. PST spores, however, have been reported to wind disperse across large distances—up to hundreds of kilometers.^{7,23}

With the global demand to reduce pesticide application limiting the means to control diseases, disease monitoring becomes increasingly important.⁵ Early disease monitoring, especially before the disease spreads, is crucial to prevent and mediate pathogen damage, and to reduce financial losses to a minimum.⁷⁶ This can occur if a damaging, hypervirulent strain is dispersed by wind across large distances to a susceptible, uniform host population (e.g., PST).^{76,92} Knowledge about the spatial distribution of such a strain therefore would provide a useful advantage to control disease early (e.g., by small-scale local pesticide application or early uprooting of mildly infected plants). In many countries farmland sizes can reach hundreds to thousands of hectares.^{17,18,93} This makes in-person crop inspections difficult and creates the need for new technologies.⁹⁴ Data gathered by a network of air-sampling devices characterizing pathogen and pest dynamics could therefore support farmers to evaluate present threats and so help to make best-informed decisions in challenging disease outbreaks. A combination of environmental and metagenome thresholds, however, would have to be developed from empirically determined data for sound disease forecasts. Much less sophisticated but proven methods are already well established and could be widely expanded by combining them with air biome testing: farmers rely on Hutton Criteria that consider temperature and relative humidity for late blight risk estimation⁹⁵—risk criteria that could easily be expanded using AirSeq.

Altogether, our study adds to the evidence that metagenomic eDNA extracted from air can be used as a source for bio-monitoring in agricultural and terrestrial ecosystems.^{2–4,10,25–30} The detection of multiple taxa in air samples tractable to agricultural environments and, more importantly, the observation of agricultural pathogens colonizing the plant species growing at our field site underlines the value of airborne eDNA methods. This indicates that AirSeq technologies are not exclusively restricted to agricultural environments but harbor the potential to profile metagenomes in any scenario where air typing is of importance.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
 - Generic air sampling device setup
 - Air sampling in the wind tunnel
 - Air sampling in the field
 - DNA extraction
 - Illumina sequencing library preparation
 - Negative controls and laboratory background control
 - Air metagenome NCBI nr database searching
 - Air metagenome PHI-base database searching
 - Air metagenome *Puccinia striiformis* f. sp. *Triticis* PST-130 analysis

- Air metagenome assemblies and analysis with ALVIS
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2024.07.030>.

ACKNOWLEDGMENTS

The authors acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC), part of UK Research and Innovation; this research was funded by the BBSRC Institute Strategic Programme grant BB/J004669/1, Core Capability grant BB/CCG1720/1, and Core Strategic Programme grant (Genomes to Food Security) BB/CSP1720/1 and its constituent work package BBS/E/T/000PR9817. The authors acknowledge the Research/Scientific Computing teams at The James Hutton Institute and The National Institute of Agricultural Botany (NIAB) for providing computational resources and technical support for the “UK’s Crop Diversity Bioinformatics HPC” (BBSRC grant BB/S019669/1), use of which has contributed to the results reported within this paper. We thank the NIAB Silsoe Spray Application Unit for the *B. thuringiensis* sample and assistance with wind tunnel experimentation. We thank Dr. Darren Heavens for providing the figure showing the Coriolis Micro dimensions. Part of this work was delivered via the BBSRC National Capability in Genomics (BB/J010375/1) at Earlham Institute (formerly The Genome Analysis Centre, Norwich), by members of the Platforms and Pipelines Group and the Research Computing Group who manage and deliver High Performance Computing at EI. The graphical abstract was created with BioRender.com.

AUTHOR CONTRIBUTIONS

M.G. analyzed the data and wrote the manuscript. W.V. conducted the experiments. R.M.L., S.M., and N.P. performed additional data analysis. R.M.L., P.N., and M.D.C. consulted on data analysis. W.V., M.D.C., and R.M.L. designed the study. R.M.L. and M.D.C. edited the manuscript.

DECLARATION OF INTERESTS

M.G., R.M.L., and M.D.C. have patents pending on AirSeq technology.

Received: February 3, 2023

Revised: April 26, 2024

Accepted: July 4, 2024

Published: August 2, 2024

REFERENCES

1. Banchi, E., Ametrano, C.G., Tordoni, E., Stanković, D., Ongaro, S., Tretiach, M., Pallavicini, A., and Muggia, L.; ARPA Working Group (2020). Environmental DNA assessment of airborne plant and fungal seasonal diversity. *Sci. Total Environ.* 738, 140249. <https://doi.org/10.1016/j.scitotenv.2020.140249>.
2. Gusareva, E.S., Acerbi, E., Lau, K.J.X., Luhung, I., Premkrishnan, B.N.V., Kolundžija, S., Purbojati, R.W., Wong, A., Houghton, J.N.I., Miller, D., et al. (2019). Microbial communities in the tropical air ecosystem follow a precise diel cycle. *Proc. Natl. Acad. Sci. USA* 116, 23299–23308. <https://doi.org/10.1073/pnas.1908493116>.
3. Gusareva, E.S., Gaultier, N.P.E., Premkrishnan, B.N.V., Kee, C., Lim, S.B.Y., Heinle, C.E., Purbojati, R.W., Nee, A.P., Lohar, S.R., Yanqing, K., et al. (2020). Taxonomic composition and seasonal dynamics of the air microbiome in West Siberia. *Sci. Rep.* 10, 21515. <https://doi.org/10.1038/s41598-020-78604-8>.
4. Drautz-Moses, D.I., Luhung, I., Gusareva, E.S., Kee, C., Gaultier, N.E., Premkrishnan, B.N.V., Lee, C.F., Leong, S.T., Park, C., Yap, Z.H., et al. (2022). Vertical stratification of the air microbiome in the lower

- troposphere. Proc. Natl. Acad. Sci. USA 119, e2117293119. <https://doi.org/10.1073/pnas.2117293119>.
5. Van der Heyden, H., Dutilleul, P., Charron, J.-B., Bilodeau, G.J., and Carisse, O. (2021). Monitoring airborne inoculum for improved plant disease management: A review. Agron. Sustain. Dev. 41, 40. <https://doi.org/10.1007/s13593-021-00694-z>.
 6. Cronin, M.F., Gentemann, C.L., Edson, J., Ueki, I., Bourassa, M., Brown, S., Clayson, C.A., Fairall, C.W., Farrar, J.T., Gille, S.T., et al. (2019). Air-sea fluxes with a focus on heat and momentum. Front. Mar. Sci. 6, 430.
 7. Brown, J.K.M., and Hovmöller, M.S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. Science 297, 537–541. <https://doi.org/10.1126/science.1072678>.
 8. Kuparinen, A., Katul, G., Nathan, R., and Schurr, F.M. (2009). Increases in air temperature can promote wind-driven dispersal and spread of plants. Proc. Biol. Sci. 276, 3081–3087. <https://doi.org/10.1098/rspb.2009.0693>.
 9. Kling, M.M., and Ackerly, D.D. (2021). Global wind patterns shape genetic differentiation, asymmetric gene flow, and genetic diversity in trees. Proc. Natl. Acad. Sci. USA 118, e2017317118. <https://doi.org/10.1073/pnas.2017317118>.
 10. Aguayo, J., Husson, C., Chancerel, E., Fabreguettes, O., Chandelier, A., Fourrier-Jeandel, C., Dupuy, N., Dutech, C., Ios, R., Robin, C., et al. (2021). Combining permanent aerobiological networks and molecular analyses for large-scale surveillance of forest fungal pathogens: A proof-of-concept. Plant Pathol. 70, 181–194. <https://doi.org/10.1111/ppa.13265>.
 11. Richardson, M., Gottel, N., Gilbert, J.A., Gordon, J., Gandhi, P., Reboulet, R., and Hampton-Marcell, J.T. (2019). Concurrent measurement of microbiome and allergens in the air of bedrooms of allergy disease patients in the Chicago area. Microbiome 7, 82. <https://doi.org/10.1186/s40168-019-0695-5>.
 12. Malhi, Y., Franklin, J., Seddon, N., Solan, M., Turner, M.G., Field, C.B., and Knowlton, N. (2020). Climate change and ecosystems: threats, opportunities and solutions. Philos. Trans. R. Soc. Lond. B Biol. Sci. 375, 20190104. <https://doi.org/10.1098/rstb.2019.0104>.
 13. Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., et al. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Mol. Ecol. 26, 5872–5895. <https://doi.org/10.1111/mec.14350>.
 14. Thomsen, P.F., and Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. Biol. Conserv. 183, 4–18. <https://doi.org/10.1016/j.biocon.2014.11.019>.
 15. Ficetola, G.F., Miaud, C., Pompanon, F., and Taberlet, P. (2008). Species detection using environmental DNA from water samples. Biol. Lett. 4, 423–425. <https://doi.org/10.1098/rsbl.2008.0118>.
 16. Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., et al. (2005). Global consequences of land use. Science 309, 570–574. <https://doi.org/10.1126/science.1111772>.
 17. Lacy, K., Orazem, P.F., and Schneekloth, S. (2023). Measuring the American farm size distribution. Am. J. Agric. Econ. 105, 219–242. <https://doi.org/10.1111/ajae.12318>.
 18. Lowder, S.K., Skoet, J., and Raney, T. (2016). The number, size, and distribution of farms, smallholder farms, and family farms worldwide. World Dev. 87, 16–29. <https://doi.org/10.1016/j.worlddev.2015.10.041>.
 19. Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., Villa, P., Stroppiana, D., Boschetti, M., Goulart, L.R., et al. (2015). Advanced methods of plant disease detection: A review. Agron. Sustain. Dev. 35, 1–25. <https://doi.org/10.1007/s13593-014-0246-1>.
 20. Buja, I., Sabella, E., Monteduro, A.G., Chiriacò, M.S., De Bellis, L., Luvisi, A., and Maruccio, G. (2021). Advances in plant disease detection and monitoring: from traditional assays to in-field diagnostics. Sensors (Basel) 21, 2129. <https://doi.org/10.3390/s21062129>.
 21. Zheng, Y., Luo, Y., Zhou, Y., Zeng, X., Duan, X., Cao, X., Song, Y., and Wang, B. (2013). Real-time PCR quantification of latent infection of wheat powdery mildew in the field. Eur. J. Plant Pathol. 136, 565–575. <https://doi.org/10.1007/s10658-013-0188-5>.
 22. Gao, L., Chen, W., Liu, T., and Liu, B. (2013). An immunofluorescence assay for the detection of wheat rust species using monoclonal antibody against urediniospores of *Puccinia triticina*. J. Appl. Microbiol. 115, 1023–1028. <https://doi.org/10.1111/jam.12295>.
 23. Rieux, A., Soubeiran, S., Bonnot, F., Klein, E.K., Ngando, J.E., Mehl, A., Ravigne, V., Carlier, J., and de Lapeyre de Bellaire, L. (2014). Long-distance wind-dispersal of spores in a fungal plant pathogen: estimation of anisotropic dispersal kernels from an extensive field experiment. PLoS One 9, e103225. <https://doi.org/10.1371/journal.pone.0103225>.
 24. Ferguson, R.M.W., Garcia-Alcea, S., Coulon, F., Dumbrell, A.J., Whitby, C., and Colbeck, I. (2019). Bioaerosol biomonitoring: Sampling optimization for molecular microbial ecology. Mol. Ecol. Resour. 19, 672–690. <https://doi.org/10.1111/1755-0998.13002>.
 25. Leung, M.H.Y., Tong, X., Boifot, K.O., Bezdan, D., Butler, D.J., Danko, D.C., Gohli, J., Green, D.C., Hernandez, M.T., Kelly, F.J., et al. (2021). Characterization of the public transit air microbiome and resistome reveals geographical specificity. Microbiome 9, 112. <https://doi.org/10.1186/s40168-021-01044-7>.
 26. Qin, N., Liang, P., Wu, C., Wang, G., Xu, Q., Xiong, X., Wang, T., Zolfo, M., Segata, N., Qin, H., et al. (2020). Longitudinal survey of microbiome associated with particulate matter in a megacity. Genome Biol. 21, 55. <https://doi.org/10.1186/s13059-020-01964-x>.
 27. Lynggaard, C., Bertelsen, M.F., Jensen, C.V., Johnson, M.S., Frøslev, T.G., Olsen, M.T., and Bohmann, K. (2022). Airborne environmental DNA for terrestrial vertebrate community monitoring. Curr. Biol. 32, 701–707.e5. <https://doi.org/10.1016/j.cub.2021.12.014>.
 28. Clare, E.L., Economou, C.K., Bennett, F.J., Dyer, C.E., Adams, K., McRobie, B., Drinkwater, R., and Littlefair, J.E. (2022). Measuring biodiversity from DNA in the air. Curr. Biol. 32, 693–700.e5. <https://doi.org/10.1016/j.cub.2021.11.064>.
 29. Nicolaisen, M., West, J.S., Sapkota, R., Canning, G.G.M., Schoen, C., and Justesen, A.F. (2017). Fungal communities including plant pathogens in near surface air are similar across Northwestern Europe. Front. Microbiol. 8, 1729. <https://doi.org/10.3389/fmicb.2017.01729>.
 30. Luhung, I., Uchida, A., Lim, S.B.Y., Gaultier, N.E., Kee, C., Lau, K.J.X., Gusareva, E.S., Heinle, C.E., Wong, A., Premkrishnan, B.N.V., et al. (2021). Experimental parameters defining ultra-low biomass bioaerosol analysis. npj Biofilms Microbiomes 7, 37. <https://doi.org/10.1038/s41522-021-00209-4>.
 31. Gervais, J., Cocks, M., Cross, A., and Jenkins, J. (2022). *Bacillus thuringiensis* (Bt) Fact Sheet (National Pesticide Information Center, Oregon State University Extension Services). <http://npic.orst.edu/factsheets/btgen.html>.
 32. Li, Q., Korza, G., and Setlow, P. (2017). Killing the spores of *Bacillus* species by molecular iodine. J. Appl. Microbiol. 122, 54–64. <https://doi.org/10.1111/jam.13310>.
 33. Knüpfer, M., Braun, P., Baumann, K., Rehn, A., Antwerpen, M., Grass, G., and Wölfel, A.R. (2020). Evaluation of a highly efficient DNA extraction method for *Bacillus anthracis* endospores. Microorganisms 8, 763. <https://doi.org/10.3390/microorganisms8050763>.
 34. Nakano, M. (2023). An improved DNA extraction method for detecting *Bacillus subtilis* spores in spiked foods and beverages. Int. J. Food Microbiol. 401, 110280. <https://doi.org/10.1016/j.ijfoodmicro.2023.110280>.
 35. Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007). MEGAN analysis of metagenomic data. Genome Res. 17, 377–386. <https://doi.org/10.1101/gr.596910.7>.
 36. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60. <https://doi.org/10.1038/nmeth.3176>.
 37. Dostálová, R., Seidenglanz, M., and Griga, M. (2005). Simulation and assessment of possible environmental risks associated with release of

- genetically modified peas (*Pisum sativum* L.) into environment in Central Europe. *Czech J. Genet. Plant Breed.* 41, 51–63. <https://doi.org/10.17221/3672-CJGPB>.
38. Guedes, A., Ribeiro, N., Ribeiro, H., Oliveira, M., Noronha, F., and Abreu, I. (2009). Comparison between urban and rural pollen of *Chenopodium alba* and characterization of adhered pollutant aerosol particles. *J. Aerosol Sci.* 40, 81–86. <https://doi.org/10.1016/j.jaerosci.2008.07.012>.
 39. ViralZone (2011). Lederbergvirus. https://viralzone.expasy.org/5197_outline=all_by_species.
 40. Karasov, T.L., Almario, J., Friedemann, C., Ding, W., Giolai, M., Heavens, D., Kersten, S., Lundberg, D.S., Neumann, M., Regalado, J., et al. (2018). *Arabidopsis thaliana* and *Pseudomonas* pathogens exhibit stable associations over evolutionary timescales. *Cell Host Microbe* 24, 168–179.e4. <https://doi.org/10.1016/j.chom.2018.06.011>.
 41. Coeur D'Acier, A., Jousselin, E., Martin, J.F., and Rasplus, J.Y. (2007). Phylogeny of the Genus *Aphis* Linnaeus, 1758 (Homoptera: Aphididae) inferred from mitochondrial DNA sequences. *Mol. Phylogenetic Evol.* 42, 598–611. <https://doi.org/10.1016/j.ympev.2006.10.006>.
 42. Mound, L.A., Wang, Z., Lima, É.F.B., and Marullo, R. (2022). Problems with the concept of “pest” among the diversity of pestiferous thrips. *Insects* 13, 61. <https://doi.org/10.3390/insects13010061>.
 43. Pinto, V.E.F., and Patriarca, A. (2017). Alternaria species and their associated mycotoxins. In *Mycotoxicogenic Fungi Methods in Molecular Biology*, A. Moretti, and A. Susca, eds. (Springer), pp. 13–32. https://doi.org/10.1007/978-1-4939-6707-0_2.
 44. Dean, R., Van Kan, J.A.L., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D., Rudd, J.J., Dickman, M., Kahmann, R., Ellis, J., et al. (2012). The top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* 13, 414–430. <https://doi.org/10.1111/j.1364-3703.2011.00783.x>.
 45. Thines, M., and Choi, Y.-J. (2016). Evolution, diversity, and taxonomy of the Peronosporaceae, with focus on the genus *Peronospora*. *Phytopathology* 106, 6–18. <https://doi.org/10.1094/PHYTO-05-15-0127-RWV>.
 46. Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S.Y., De Silva, N., Martinez, M.C., Pedro, H., Yates, A.D., et al. (2020). PHIL-base: the pathogen-host interactions database. *Nucleic Acids Res.* 48, D613–D620. <https://doi.org/10.1093/nar/gkz904>.
 47. Ayling, M., Clark, M.D., and Leggett, R.M. (2020). New approaches for metagenome assembly with short reads. *Brief. Bioinform.* 21, 584–594. <https://doi.org/10.1093/bib/bbz020>.
 48. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1303.3997>.
 49. Williamson, B., Tudzynski, B., Tudzynski, P., and Van Kan, J.A.L. (2007). *Botrytis cinerea*: the cause of grey mould disease. *Mol. Plant Pathol.* 8, 561–580. <https://doi.org/10.1111/j.1364-3703.2007.00417.x>.
 50. Mousavi, B., Hedayati, M.T., Hedayati, N., Ilkit, M., and Syedmousavi, S. (2016). Aspergillus species in indoor environments and their possible occupational and public health hazards. *Curr. Med. Mycol.* 2, 36–42. <https://doi.org/10.18869/acadpub.cmm.2.1.36>.
 51. De Paiva-Santos, W., De Sousa, V.S., and Giambiagi-deMarval, M. (2018). Occurrence of virulence-associated genes among *Staphylococcus saprophyticus* isolated from different sources. *Microb. Pathog.* 119, 9–11. <https://doi.org/10.1016/j.micpath.2018.03.054>.
 52. Argudín, M.A., Vanderhaeghen, W., and Butaye, P. (2015). Antimicrobial resistance and population structure of *Staphylococcus epidermidis* recovered from pig farms in Belgium. *Vet. J.* 203, 302–308. <https://doi.org/10.1016/j.tvjl.2015.01.008>.
 53. Department for Environment, Food & Rural Affairs (2023). Agricultural facts: east of England region. <https://www.gov.uk/government/statistics/agricultural-facts-england-regional-profiles/agricultural-facts-east-of-england-region>.
 54. Roger, J.G. (1963). The Atlas of the British Flora. *Trans. Bot. Soc. Edinb.* 39, 433–434. <https://doi.org/10.1080/13594866309441728>.
 55. Stroh, P.A., Walker, K.J., Humphrey, T.A., and Pescott, O.L. (2020). Plant Atlas 2020: Online Atlas of the British and Irish Flora. <https://www.brc.ac.uk/plantatlas/>.
 56. Miller, P.R. (1953). The effect of weather on diseases. *Yearbook of Agriculture*, 83–93.
 57. Romero, F., Cazzato, S., Walder, F., Vogelsgang, S., Bender, S.F., and Van Der Heijden, M.G.A. (2022). Humidity and high temperature are important for predicting fungal disease outbreaks worldwide. *New Phytol.* 234, 1553–1556. <https://doi.org/10.1111/nph.17340>.
 58. Cordo, C.A., Mónaco, C.I., Altamirano, R., Perelló, A.E., Larrán, S., Kripelz, N.I., and Simón, M.R. (2017). Weather conditions associated with the release and dispersal of *Zymoseptoria tritici* spores in the Argentine Pampas region. *Int. J. Agron.* 2017, 1–13. <https://doi.org/10.1155/2017/1468580>.
 59. Te Beest, D.E., Paveley, N.D., Shaw, M.W., and Van Den Bosch, F. (2008). Disease–weather relationships for powdery mildew and yellow rust on winter wheat. *Phytopathology* 98, 609–617. <https://doi.org/10.1094/PHYTO-98-5-0609>.
 60. Doohan, F.M., Brennan, J., and Cooke, B.M. (2003). Influence of climatic factors on *Fusarium* species pathogenic to cereals. *Eur. J. Plant Pathol.* 109, 755–768. <https://doi.org/10.1023/A:1026090626994>.
 61. Zhang, X., Ovenden, B., and Milgate, A. (2020). Recent insights into barley and *Rhynchosporium* commune interactions. *Mol. Plant Pathol.* 21, 1111–1128. <https://doi.org/10.1111/mpp.12945>.
 62. Clarkson, J.P., Fawcett, L., Anthony, S.G., and Young, C. (2014). A model for *Sclerotinia sclerotiorum* infection and disease development in lettuce, based on the effects of temperature, relative humidity and ascospore density. *PLoS One* 9, e94049. <https://doi.org/10.1371/journal.pone.0094049>.
 63. Huang, S., Zuo, S., Zheng, D., Liu, Y., Du, Z., Kang, Z., and Zhao, J. (2019). Three formae speciales of *Puccinia striiformis* were identified as heteroecious rusts based on completion of sexual cycle on *Berberis* spp. under artificial inoculation. *Phytopathol. Res.* 1, 14. <https://doi.org/10.1186/s42483-019-0021-y>.
 64. Hubbard, A., Lewis, C.M., Yoshida, K., Ramirez-Gonzalez, R.H., de Vallavieille-Pope, C., Thomas, J., Kamoun, S., Bayles, R., Uauy, C., and Saunders, D.G.O. (2015). Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome Biol.* 16, 23. <https://doi.org/10.1186/s13059-015-0590-8>.
 65. Farber, D.H., Medlock, J., and Mundt, C.C. (2017). Local dispersal of *Puccinia striiformis* f. sp. *tritici* from isolated source lesions. *Plant Pathol.* 66, 28–37. <https://doi.org/10.1111/ppa.12554>.
 66. Chen, W., Wellings, C., Chen, X., Kang, Z., and Liu, T. (2014). Wheat stripe (yellow) rust caused by *Puccinia striiformis* f. sp. *tritici*. *Mol. Plant Pathol.* 15, 433–446. <https://doi.org/10.1111/mpp.12116>.
 67. Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., et al. (2002). The Ensembl genome database project. *Nucleic Acids Res.* 30, 38–41. <https://doi.org/10.1093/nar/30.1.38>.
 68. Pearman, W.S., Freed, N.E., and Silander, O.K. (2020). Testing the advantages and disadvantages of short- and long- read eukaryotic metagenomics using simulated reads. *BMC Bioinformatics* 21, 220. <https://doi.org/10.1186/s12859-020-3528-4>.
 69. Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics* 31, 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
 70. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. <https://doi.org/10.1186/1471-2105-10-421>.
 71. Martin, S., and Leggett, R.M. (2021). Alvis: a tool for contig and read ALignment VISeualisation and chimera detection. *BMC Bioinformatics* 22, 124. <https://doi.org/10.1186/s12859-021-04056-0>.

72. Ruppert, K.M., Kline, R.J., and Rahman, M.S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* 17, e00547. <https://doi.org/10.1016/j.gecco.2019.e00547>.
73. Bentham, A.R., De la Concepcion, J.C., Mukhi, N., Zdrzalek, R., Draeger, M., Gorenkin, D., Hughes, R.K., and Banfield, M.J. (2020). A molecular roadmap to the plant immune system. *J. Biol. Chem.* 295, 14916–14935. <https://doi.org/10.1074/jbc.REV120.010852>.
74. Sappington, T.W. (2018). Migratory flight of insect pests within a year-round distribution: European corn borer as a case study. *J. Integr. Agric.* 17, 1485–1505. [https://doi.org/10.1016/S2095-3119\(18\)61969-0](https://doi.org/10.1016/S2095-3119(18)61969-0).
75. Bebb, D.P., Holmes, T., and Gurr, S.J. (2014). The global spread of crop pests and pathogens. *Glob. Ecol. Biogeogr.* 23, 1398–1407. <https://doi.org/10.1111/geb.12214>.
76. McDonald, B.A., and Stukenbrock, E.H. (2016). Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371, 20160026. <https://doi.org/10.1098/rstb.2016.0026>.
77. Picelli, S., Björklund, A.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* 24, 2033–2040. <https://doi.org/10.1101/gr.177881.114>.
78. NBN Trust (2023). The National Biodiversity Network (NBN) Atlas. <https://ror.org/00mcxye41> <https://www.brc.ac.uk/plantatlas/>.
79. Wicker, T., Oberhaensli, S., Parlange, F., Buchmann, J.P., Shatalina, M., Roffler, S., Ben-David, R., Doležel, J., Šimková, H., Schulze-Lefert, P., et al. (2013). The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. *Nat. Genet.* 45, 1092–1096. <https://doi.org/10.1038/ng.2704>.
80. Fones, H., and Gurr, S. (2015). The impact of *Septoria tritici* Blotch disease on wheat: An EU perspective. *Fungal Genet. Biol.* 79, 3–7. <https://doi.org/10.1016/j.fgb.2015.04.004>.
81. Duba, A., Goriewa-Duba, K., and Wachowska, U. (2018). A review of the interactions between wheat and wheat pathogens: *Zymoseptoria tritici*, *Fusarium* spp. and *Parastagonospora nodorum*. *Int. J. Mol. Sci.* 19, 1138. <https://doi.org/10.3390/ijms19041138>.
82. Justesen, A.F., Corsi, B., Ficke, A., Hartl, L., Holdgate, S., Jørgensen, L.N., Lillemo, M., Lin, M., Mackay, I.J., Mohler, V., et al. (2021). Hidden in plain sight: a molecular field survey of three wheat leaf blotch fungal diseases in North-Western Europe shows co-infection is widespread. *Eur. J. Plant Pathol.* 160, 949–962.
83. Urban, M., King, R., Andongabo, A., Maheswari, U., Pedro, H., Kersey, P., and Hammond-Kosack, K. (2016). First draft genome sequence of a UK Strain (UK99) of *Fusarium culmorum*. *Genome Announc.* 4, e00771–16. <https://doi.org/10.1128/genomeA.00771-16>.
84. David, R.F., Marr, L.C., and Schmale, D.G. (2016). Ascospore release and discharge distances of *Fusarium graminearum* under controlled temperature and relative humidity. *Eur. J. Plant Pathol.* 146, 59–69. <https://doi.org/10.1007/s10658-016-0891-0>.
85. James, G.L., Latif, M.T., Isa, M.N.M., Bakar, M.F.A., Yusuf, N.Y.M., Broughton, W., Murad, A.M., and Abu Bakar, F.D.A. (2021). Metagenomic datasets of air samples collected during episodes of severe smoke-haze in Malaysia. *Data Brief* 36, 107124. <https://doi.org/10.1016/j.dib.2021.107124>.
86. Constantinidou, H.-I., Κωνσταντινίδου, E.-I., Hirano, S., Baker, L., and Upper, C. (1990). Atmospheric dispersal of ice nucleation-active bacteria: The role of rain. *Phytopathology* 80, 934–937.
87. Aboukhaddour, R., Fetch, T., McCallum, B.D., Harding, M.W., Beres, B.L., and Graf, R.J. (2020). Wheat diseases on the prairies: A Canadian story. *Plant Pathol.* 69, 418–432. <https://doi.org/10.1111/ppa.13147>.
88. Lewis, C.M., Persoons, A., Bebb, D.P., Kigathi, R.N., Maintz, J., Findlay, K., Bueno-Sancho, V., Corredor-Moreno, P., Harrington, S.A., Kangara, N., et al. (2018). Potential for re-emergence of wheat stem rust in the United Kingdom. *Commun. Biol.* 1, 13. <https://doi.org/10.1038/s42003-018-0013-y>.
89. Albanese, D., and Donati, C. (2017). Strain profiling and epidemiology of bacterial species from metagenomic sequencing. *Nat. Commun.* 8, 2260. <https://doi.org/10.1038/s41467-017-02209-5>.
90. Dimmock, J., and Gooding, M. (2002). The influence of foliar diseases, and their control by fungicides, on the protein concentration in wheat grain: a review. *J. Agric. Sci.* 138, 349–366. <https://doi.org/10.1017/S0021859602002058>.
91. Hovmöller, M.S., Walter, S., and Justesen, A.F. (2010). Escalating threat of wheat rusts. *Science* 329, 369.
92. Huang, L., Yang, H., Xia, C., Li, H., Wang, J., Wang, A., Zhang, M., Kang, X., Gao, L., Zhou, Y., et al. (2022). Long-distance transport of *Puccinia striiformis* f. sp. *tritici* by upper airflow on the Yunnan–Guizhou Plateau disrupts the balance of agricultural ecology in Central China. *Plant Dis.* 106, 2940–2947. <https://doi.org/10.1094/PDIS-01-22-0038-RE>.
93. Loughrey, J., Donnellan, T., Lennon, J., Loughrey, J., Donnellan, T., and Lennon, J. (2016). The inequality of farmland size in Western Europe. In 90th Annual Conference of the Agricultural Economics Society. <https://doi.org/10.22004/AG.ECON.236341>.
94. Walter, A., Finger, R., Huber, R., and Buchmann, N. (2017). Opinion: Smart farming is key to developing sustainable agriculture. *Proc. Natl. Acad. Sci. USA* 114, 6148–6150. <https://doi.org/10.1073/pnas.1707462114>.
95. Dancey, S. (2018). Development and implementation of a new national warning system for potato late blight in Great Britain. PhD thesis (University of Dundee).
96. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
97. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>.
98. Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* 33, 1635–1638. <https://doi.org/10.1093/molbev/msw046>.
99. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
100. Estévez, S.R. (2022). Robaina/filterSAM: v0.0.10. Version v0.0.10 (Zenodo). <https://doi.org/10.5281/ZENODO.7056278>.
101. Sherrill-Mix, S. (2021). taxonomizr: Functions to Work with NCBI Accessions and Taxonomy. Version 0.8.0. <https://github.com/sherrillmix/taxonomizr>.
102. Hahn, M.W. (2004). Broad diversity of viable bacteria in 'sterile' (0.2 µm) filtered water. *Res. Microbiol.* 155, 688–691. <https://doi.org/10.1016/j.resmic.2004.05.003>.
103. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. <https://doi.org/10.1093/nar/gkv1189>.
104. Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013). GenBank. *Nucleic Acids Res.* 41, D36–D42. <https://doi.org/10.1093/nar/gks1195>.
105. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFTools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Field captured air samples in water	This study	NA
Chemicals, peptides, and recombinant proteins		
Triton X-100	Merck Sigma Aldrich	T8787-50ML
UltraPure Dnase/Rnase free, distilled water	Invitrogen	10977023
<i>B. thuringiensis</i> spores	Silsoe Spray Application Unit	N/A
GenFind V2 kit	Beckman Coulter Life Sciences	A41497
Qubit 2.0 High Sensitivity reagents	Thermo Fisher	Q32581
Buffer PB	QIAGEN	19066
AMPure XP beads	Beckman Coulter	A63882
KAPA 2G Robust polymerase kit	Sigma Aaldrich	KK5023
Nextera DNA Library Prep Kit	Illumina	FC-121-1030
Critical commercial assays		
Agilent Bioanalyzer High Sensitivity DNA Analysis reagents	Agilent	5067-4626
Deposited data		
FASTQ sequence data deposited at ENA	This paper	PRJEB58191
Software and algorithms		
Data analysis scripts	This paper	https://github.com/mgiolai/crop_airseq
Fastp	Chen et al. ⁹⁶	https://github.com/OpenGene/fastp
MEGAHIT	Li et al. ⁶⁹	https://github.com/voutcn/megahit
BLAST+	Camacho et al. ⁷⁰	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
ALVIS	Martin and Leggett ⁷¹	https://github.com/SR-Martin/alvis
CD-HIT	Fu et al. ⁹⁷	https://sites.google.com/view/cd-hit
DIAMOND	Buchfink et al. ³⁶	https://github.com/bbuchfink/diamond
ETE Toolkit	Huerta-Cepas et al. ⁹⁸	http://etetoolkit.org/download/
BWA	Li ⁴⁸	https://github.com/lh3/bwa
Samtools	Li et al. ⁹⁹	http://samtools.sourceforge.net/
Filtersam	Estévez ¹⁰⁰	https://pypi.org/project/filtersam/
Taxonomizr	Sherrill-Mix ¹⁰¹	https://github.com/sherrillmix/taxonomizr
Other		
0.22 µm pore size, hydrophilic PVDF, 13 mm diameter filter membrane	Merck Sigma Aldrich	GVWP01300
13 mm diameter stainless steel Swinny filter holder	Merck Sigma Aldrich	XX3001200
0.7 mm Garnet particle containing tube	MoBio – now QIAGEN PowerBead	13123-50

RESOURCE AVAILABILITY

Lead contact

Further information or request for resources should be forwarded to the lead contact, Matthew Clark (matt.clark@nhm.ac.uk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- FASTQ sequence data have been deposited at the European Nucleotide Archive (ENA) under accession number PRJEB58191 and are publicly available as of the date of publication.
- All original code used for analysis has been deposited in the GitHub repository https://github.com/mgiolai/crop_airseq and is publicly available as of the date of publication.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Wind tunnel air samples were collected at the Silsoe Spray Applications Unit (National Institute of Agricultural Botany – NIAB) on 3rd July 2015, in a 10 m long, 3 m wide and 2 m high wind tunnel.

Field air samples were collected at a Norwich Research Park field site in between the Earlham Institute and a minor road, both at the farthest possible distance of 100 m from our sampling site (Coordinates: 52°37'21.1"N, 1°13'02.6"E, [Figure S1](#)). Samples were collected 3 times a week from 8th June to 29th July 2015.

METHOD DETAILS

Generic air sampling device setup

All air samples in all experiments were collected using the Coriolis Micro Microbial Air sampler (Bertin Instruments, Montigny le Bretonneux, France) with its tripod fully extended ([Figure S2](#)) and according to the following settings:

The sampling device is placed on a tripod stand 115 cm above ground; air intake is performed 130 – 132 cm above ground. During sampling we set a flow rate of 200 l/minute with 10 ml 0.05 % (v/v) Triton X-100 (T8787-50ML, Merck Sigma Aldrich) as collection buffer prepared with UltraPure Dnase/Rnase free, distilled water (10977023, Invitrogen). After collection we transferred the buffer to a 15 ml conical tube and stored the collected sample at –20 °C until processing.

Air sampling in the wind tunnel

Wind tunnel air samples were collected at the Silsoe Spray Applications Unit (National Institute of Agricultural Botany – NIAB) in a 10 m long, 3 m wide and 2 m high wind tunnel. We placed the Coriolis Micro sampler (tripod fully extended) at a 5 m distance from the facilities liquid sprayer releasing a dilution series of 0.3, 30 and 300 million *B. thuringiensis* spores (a kind gift from the Silsoe Spray Application Unit) or a water control for 10 minutes at a wind speed of 2 m/s. The device had to be started manually; however the wind-tunnel was vacant while sampling air. During this experiment we kept the spore release rate constant at 300 ml/minute, releasing a total of 3 l spore dilution or water control. The solutions were released at a height of 1.25 m, air was collected at a height of 0.80 m. The wind tunnel setup was performed in an enclosed room from which air exchange with surroundings occurred, hence the room does not provide an entirely controlled environment.

Air sampling in the field

Field samples were collected at a Norwich Research Park field site in between the Earlham Institute and a minor road, both at the farthest possible distance of 100 m from our sampling site (Coordinates: 52°37'21.1"N, 1°13'02.6"E, [Figure S1](#)). The shortest distance to the barley, wheat and pea fields we measured as 15 m, 18 m and 61 m respectively. Wheat was grown most abundantly at our test site. We selected this environment as an accessible field that is regularly used for plant and plant disease studies (see [Figure S3](#)), but with limited access of people and no human movement while sampling. The Coriolis Micro device (tripod fully extended) was started manually for sampling. After commencing sampling, we left the site and returned after sampling had finished. To keep our sampling times comparable over the selected time-period, we commenced with sample collection mostly between 9 am to 12 am with few exceptions ([Table S1](#)).

In the field we fully extended the tripod of the Coriolis Micro sampler. On warm days in the field, we observed that the collection buffer evaporated during collection to a volume range of 3.5 – 8 ml. To avoid drying of the collection buffer while sampling on especially hot days we topped-up the solution during sampling by briefly stopping air collection, adding 10 ml fresh collection buffer and then continuing with sampling until the 60-minute collection mark was reached (buffer top-up takes approximately one minute).

DNA extraction

To standardise the genomic DNA extraction, we filtered the collected sample through a 0.22 µm pore size, hydrophilic PVDF, 13 mm diameter filter membrane (GVWP01300, Merck Sigma Aldrich) using a 13 mm diameter stainless steel Swinny filter holder (XX3001200, Merck Sigma Aldrich). We used a 0.22 µm pore size filter as this pore size is regularly used to filter sterilise non-autoclavable solutions in laboratories and so should retain a large fraction of the air metagenome, however some microbes do pass through such filters, possibly in a size dependent manner.¹⁰² We also found that 0.22 µm pore size membranes did not clog on days with high pollen levels in the air (determined by collecting air in the field and filtering of the collection buffer without further processing). We discarded the flowthrough of the PVDF membrane and transferred the membrane immediately after filtration to a 0.7 mm Garnet particle containing tube (13123-50, MoBio – now QIAGEN PowerBead) using sterile tweezers. We immediately added 500 µl UltraPure Dnase/Rnase free,

distilled water to the membrane and ground the reaction for 10 minutes on the TissueLyser II (Qiagen) at 32 Hz. We transferred the supernatant of the grinding step (approximately 300 µl) to a 1.5 ml conical tube and proceeded with DNA purification. DNA purification was performed using the GenFind V2 kit (A41497, Beckman Coulter Life Sciences) according to the standard protocol but with two minor modifications: Our starting volume was 300 µl instead of the 200 µl as in the standard protocol. We therefore adjusted (i.e., increased) the volumes of the buffer solutions as described in the GenFind V2 Blood & Serum protocol (PN B66719AB) for 300 µl starting volume and we reduced the elution buffer volume to 10 µl UltraPure Dnase/Rnase free, distilled water – in brief: we prepared 450 µl binding buffer (1.5 x sample volume) by combining 440 µl magnetic particle free GenFind V2 Binding Buffer with 10 µl of GenFind V2 Binding Buffer containing magnetic particles. We added the binding buffer to the 300 µl lysate and mixed well. We incubated the reaction for 5 minutes at room temperature, briefly centrifuged the reaction using a table-top centrifuge and pelleted the beads on a magnetic stand. We washed the reaction using 500 µl GenFind V2 Wash Buffer 1, re-pelleted the beads on a magnetic stand and washed the reaction using 500 µl GenFind V2 Wash Buffer 2. We pelleted the beads on a magnetic stand, discarded the supernatant and eluted the extracted DNA in 10 µl water. The extracted DNA was not quantified as DNA amounts were below the detection limit of Qubit 2.0 High Sensitivity reagents (Q32581, Thermo Fisher). All extractions were prepared in a PCR workstation with designated equipment for the workstation. The workstation was UV-light cleaned after each usage.

Illumina sequencing library preparation

To prepare an Illumina sequencing library from the extracted air DNA we concentrated the 10 µl previously extracted sample to 1.5 µl in an Eppendorf SpeedVac Concentrator 5301. To the 1.5 µl concentrated air DNA extraction we added 2.5 µl Illumina Nextera reaction buffer (FC-121-1030, Illumina), 1 µl 1 pg/µl Lambda DNA (SD0011, ThermoFisher Scientific) as fragmentation reaction control and 0.2 µl Nextera enzyme in a total reaction volume of 5 µl. We incubated the reaction for 5 minutes at 55 °C in a G-Storm GS1 (G-Storm) thermal cycler and after incubation added 5 µl Buffer PB (19066, QIAGEN) to inactivate and strip the transposase from the fragmented DNA. We cleaned the reaction using 10 µl (1 x ratio) AMPure XP beads (Beckman Coulter) and eluted the reaction in 20 µl UltraPure Dnase/Rnase free, distilled water. We immediately proceeded with library amplification adding 10 µl 5X KAPA 2G Robust Buffer, 1 µl 10 mM dNTPs, 5 µl 2.5 µM P5 and 5 µl P7 oligonucleotide, 8.9 µl water and 0.1 µl KAPA 2G Robust polymerase to the eluted DNA. We incubated the reaction in a G-Storm GS1 thermal cycler with the programme: 3 min at 72 °C, 1 min at 95 °C, [10 s at 95 °C, 30 s at 65 °C, 2 min 30 s at 72 °C] for 18 cycles and a final elongation step for 2 min 30 s at 72 °C. After PCR amplification we cleaned the reaction using 50 µl (1 x ratio) AMPure XP beads eluting in 50 µl 1 x TE buffer. After clean-up we quantified the libraries using Qubit 2.0 High Sensitivity reagents and analysed the size using Agilent Bioanalyzer High Sensitivity DNA Analysis reagents (5067-4626, Agilent) (Figure S5). As for the DNA extraction, all library preparation reactions were prepared in a designated PCR workstation. The workstation was UV-light cleaned after each usage. All samples were submitted for Illumina HiSeq 2500 150 single-end chemistry sequencing at the Earlham Institute.

Negative controls and laboratory background control

DNA extractions and library preparations were performed in a PCR workstation equipped with pipettes, consumables and reagents, which all, including the workstation were exclusively in use for AirSeq work. Human traffic in the laboratory for AirSeq experimentation was minimal. The workstation was cleaned after each use with UV-light. We performed negative controls for each batch of library preparations and analysed the final Nextera negative control libraries using Qubit 2.0 High Sensitivity reagents as well as Agilent Bioanalyzer High Sensitivity DNA Analysis reagents (Figure S6). The constructed Nextera libraries did not have a measurable DNA concentration using Qubit 2.0 High Sensitivity reagents nor showed presence of an amplicon smear on an Agilent Bioanalyzer High Sensitivity DNA Chip. We did however observe strong adapter dimer amplification in an experiment when we loaded uncleaned Nextera library on an Agilent Bioanalyzer High Sensitivity Chip (Figure S6). This indicates the absence of contamination in our negative control libraries. Considering that the negative controls did not have a measurable DNA concentration and so would have been outcompeted by other library molecules in Illumina flow cell clustering or would have severely under-clustered if sequenced on an additional flow cell, we decided to collect air in the laboratory next to the PCR workbench as this was the most likely source of contamination. We used this sample as background control instead of our negative controls to consider influences of the lab air environment on our samples. Genera found in this lab control sample and the genus *Homo* are removed from the data (https://github.com/mgiolai/crop_airseq).

Air metagenome NCBI nr database searching

Reads were adaptor and quality trimmed using fastp 0.20.1⁹⁶ and deduplicated using CD-HIT-auxtools 4.6.8.⁹⁷ To analyze the composition of the air samples, we aligned the filtered and deduplicated reads to the entire NCBI nr protein database (downloaded on the 29.10.2021) using DIAMOND 2.0.13³⁶ with the settings –eval 1e-10 –mid-sensitive –max-hsp 1 –top 10 (https://github.com/mgiolai/crop_airseq).

We filtered and classified the diamond output files using custom scripts (https://github.com/mgiolai/crop_airseq) maintaining the DIAMOND e-value threshold $\leq 1e-10$ for filtering alignments. Taxonomic binning was performed as described for the MEGAN lowest common ancestor algorithm³⁵ considering taxonomic assignments of a bitscore within 10 % of the best bitscore; in brief: We firstly removed all alignments with an e-value $\leq 1e-10$ from the DIAMOND output file. In a second step we calculated the 90 % ratio of the highest bitscore for each read (i.e., bitscore-threshold = highest-bitscore * 0.9). We then removed all alignments below the calculated bitscore-threshold for the read. We then classified the remaining reads according to the lowest common ancestor algorithm by assigning a taxon to a read if all alignments shared the same taxonomy (i.e., if all alignments for a specific read showed the same genus,

the read was classified as that genus, otherwise the same analysis was repeated for the next higher taxonomic level). After taxonomic classification we required taxonomic assignments to be supported by 0.1 % of all classified reads to be valid. Read classification was performed querying the NCBI taxonomy database with the ETE Toolkit 3.1.1⁹⁸ (https://github.com/mgiolai/crop_airseq). The resulting count tables were analysed using custom R-4.1.0 scripts (https://github.com/mgiolai/crop_airseq). Reads were normalised to reads per million by dividing the read number of a taxon with the number of quality filtered and deduplicated reads divided by one million. An all versus all sample comparison using Pearson correlation of detected genera gave an overview of the pipeline success (see [Figure S4](#)), with particular attention paid to replicate samples (see [Table S2](#)).

Air metagenome PHI-base database searching

We downloaded the PHI-base v4.12⁴⁶ pathogen species genomes (one genome per species, 271 of 279 genomes were available, the genomes were queried on 09.12.2022) from NCBI RefSeq¹⁰³ and GenBank.¹⁰⁴ We also added the reference genome for *P. tritici* (GCA_019358815.1) and the T2T Consortium *Homo sapiens* genome assembly (T2T-CHM13v2.0, GCF_009914755.1) to the database. The human reference was added to consider potential biases originating from human reads. Single genomes were selected as the newest version of the ‘reference genome’ or, if a reference genome was not available, as the ‘representative genome’ for a species. For multiple available genomes we selected the longest genome by the primary assembly length listed in the NCBI assembly stats file. If neither reference nor representative genomes were available, we selected the longest ‘Complete genome’ or if no complete genomes were available the longest genome per pathogen species (https://github.com/mgiolai/crop_airseq). The genomes were combined in a single fasta file. We mapped the fastp 0.20.1 adapter and quality trimmed, as well as CD-HIT-auxtools 4.6.8 deduplicated reads to a database containing a single genome of each pathogen using BWA-MEM 0.7.17 standard settings (https://github.com/mgiolai/crop_airseq).

We filtered the sam files for a MAPQ > 0 using samtools-1.16.1,⁹⁹ %-identity \geq 95 and %-matched sequence score \geq 95 using filtersam-0.0.11¹⁰⁰ and analyzed the alignments using custom R-4.1.0 scripts (https://github.com/mgiolai/crop_airseq) only considering reads that mapped to a single pathogen species, requiring that a species was covered at \geq 10 distinct loci and supported by 0.05 % of all taxonomically assigned reads. In brief: We imported the MAPQ, %-identity and %-match filtered sam file into R 4.1.0. Based on the taxonomy ID which we had integrated in the FASTA sequence description of each species we determined to which species a read mapped. We removed all reads which mapped to multiple species. We further required that a species was supported by hits to multiple distinct loci in a genome to exclude species calls originating from few regions. We removed all species with reads assigned to less than 10 different genomic loci (empirically determined to consider coverage breadth). We applied an abundance threshold of 0.05 % only considering species supported by 0.05 % of all taxonomically assigned reads. Taxonomic assignment was performed querying the NCBI taxonomy database with the R package taxonomizr 0.8.0.¹⁰¹ Reads were normalized to reads per million by dividing the read number of a taxon with the number of quality filtered and deduplicated reads divided by one million (https://github.com/mgiolai/crop_airseq).

Air metagenome *Puccinia striiformis f. sp. Tritici* PST-130 analysis

We mapped the fastp 0.20.1 adapter and quality trimmed, as well as CD-HIT-auxtools 4.6.8 deduplicated reads to a database containing a single genome of each pathogen using BWA-MEM 0.7.17 standard settings to the ensemble PST-130_1.0 genome assembly (https://ftp.ensemblgenomes.org/pub/fungi/release-55/fasta/puccinia_striiformis). We filtered the reads for MAPQ > 0 using samtools-1.16.1 and a %-identity and %-matched sequence score \geq 95 using filtersam-0.0.11 and generated a text pileup from the mapping file using bcftools-1.16 mpileup.¹⁰⁵ We called and filtered variants using bcftools-1.16 with a quality filter setting of -I “QUAL > 20 && DP > 2”. Transition to transversion ratios were calculated using the bcftools-1.16 stats command. The variant calling files were analysed using custom R-4.1.0 scripts (https://github.com/mgiolai/crop_airseq).

Air metagenome assemblies and analysis with ALVIS

We assembled fastp 0.20.1 adapter and quality trimmed reads using MEGAHIT-1.2.9⁶⁹ specifying the --presets meta-sensitive option. We mapped the contigs obtained from MEGAHIT to selected reference genomes using BLAST+ 2.9.0,⁷⁰ the output was passed to Alvis-1.2⁷¹ with default settings to create heatmaps showing alignments of contigs to reference genomes ([Data S2](#)). The genome accessions we selected for this analysis are: *T. aestivum* (Ensembl release 57), *C. album* (GCA_948465745.1), *H. vulgare* (GCF_904849725.1), *B. graminis* (GCA_905067625.1), *P. striiformis* (GCF_021901695.1), *U. hordei* (GCA_022749175.1).

QUANTIFICATION AND STATISTICAL ANALYSIS

Quality control, trimming and deduplicating of raw reads, followed by the searching of the large NCBI nr sequence database, with result filtering and taxonomic classification of reads is described in the section “air metagenome NCBI nr database searching”. The selection of reference sequences for a custom pathogen database, along with the method used for filtering alignment results is described in the section “air metagenome PHI-base database searching”. Identification of polymorphisms in wheat yellow rust derived reads, including the method used for filtering alignments, is described in “air metagenome *Puccinia striiformis f. sp. Tritici* PST-130 analysis”. The filtering of reads, the subsequent metagenomic assembly and the analysis of the resulting contigs is described in “air metagenome assemblies and analysis with ALVIS”.