

## ARTICLE OPEN



# Experimental parameters defining ultra-low biomass bioaerosol analysis

Irvan Luhung<sup>1,3</sup>, Akira Uchida<sup>1,3</sup>, Serene B. Y. Lim<sup>1</sup>, Nicolas E. Gaultier<sup>1</sup>, Carmon Kee<sup>1</sup>, Kenny J. X. Lau<sup>1</sup>, Elena S. Gusareva<sup>1</sup>, Cassie E. Heinle<sup>1</sup>, Anthony Wong<sup>1</sup>, Balakrishnan N. V. Premkrishnan<sup>1</sup>, Rikky W. Purbojati<sup>1</sup>, Enzo Acerbi<sup>1</sup>, Hie Lim Kim<sup>1</sup>, Ana C. M. Junqueira<sup>1,2</sup>, Sharon Longford<sup>1</sup>, Sachin R. Lohar<sup>1</sup>, Zhei Hwee Yap<sup>1</sup>, Deepa Panicker<sup>1</sup>, Yanqing Koh<sup>1</sup>, Kavita K. Kushwaha<sup>1</sup>, Poh Nee Ang<sup>1</sup>, Alexander Putra<sup>1</sup>, Daniela I. Drautz-Moses<sup>1</sup> and Stephan C. Schuster<sup>1</sup>✉

Investigation of the microbial ecology of terrestrial, aquatic and atmospheric ecosystems requires specific sampling and analytical technologies, owing to vastly different biomass densities typically encountered. In particular, the ultra-low biomass nature of air presents an inherent analytical challenge that is confounded by temporal fluctuations in community structure. Our ultra-low biomass pipeline advances the field of bioaerosol research by significantly reducing sampling times from days/weeks/months to minutes/hours, while maintaining the ability to perform species-level identification through direct metagenomic sequencing. The study further addresses all experimental factors contributing to analysis outcome, such as amassment, storage and extraction, as well as factors that impact on nucleic acid analysis. Quantity and quality of nucleic acid extracts from each optimisation step are evaluated using fluorometry, qPCR and sequencing. Both metagenomics and marker gene amplification-based (16S and ITS) sequencing are assessed with regard to their taxonomic resolution and inter-comparability. The pipeline is robust across a wide range of climatic settings, ranging from arctic to desert to tropical environments. Ultimately, the pipeline can be adapted to environmental settings, such as dust and surfaces, which also require ultra-low biomass analytics.

*npj Biofilms and Microbiomes* (2021)7:37; <https://doi.org/10.1038/s41522-021-00209-4>

## INTRODUCTION

Great naturalists of centuries-past have catalogued planetary ecosystems at the macroscopic level, primarily for terrestrial and aquatic environments, where organisms were most accessible<sup>1,2</sup>. Microscopic life was subsequently given the same attention, again initially focusing on terrestrial and aquatic systems<sup>3,4</sup>. Microbial inhabitants of the third ecosystem of planetary scale, the atmosphere, proved much more difficult to assess due to technological challenges in regard to accessibility. These challenges are largely associated with the low-density gaseous state and resulting ultra-low biomass of air<sup>5–7</sup>. As a consequence, atmospheric research first described the physicochemical nature of the atmosphere, thereby generating a comprehensive understanding of inanimate components of the troposphere and stratosphere<sup>8</sup>. The origin of these components of air is typically categorised as either inorganic gases or volatile organic compounds (VOCs), the latter of which serve as proxies for the biological activity of organisms<sup>9,10</sup>. The following progression in the field involved the identification of airborne organisms via cultivation and microscopy<sup>11,12</sup>. This provided a foundation for understanding the composition of airborne microbial organisms via nucleic acid taxonomic identification. A large increase of the taxonomic resolution was subsequently achieved by the use of ITS and 16S rRNA gene markers. The ultra-low biomass nature of air posed major technical obstacles to using these molecular techniques, with inherent requirements such as long sampling duration and high amounts of gene marker amplification<sup>13–18</sup>.

The nascent field of bioaerosol studies was further progressed by employing metagenomics, which enabled direct nucleic acid analysis without the biases associated with gene amplification.

However, to overcome issues associated with limited biomass, long sampling duration times (days to weeks) were unavoidable, which in turn impeded the temporal resolution and the number of required samples analysed<sup>19–21</sup>.

Advances in temporal and taxonomic resolution only became possible with the onset of new technologies involving high volumetric flow rate air samplers coupled with metagenomic data generated by next-generation sequencing platforms that had low biomass requirements<sup>22</sup>. This approach, which analyses the accessible spectrum of airborne community DNA, therefore enables assessment of the functional complement of airborne microorganisms.

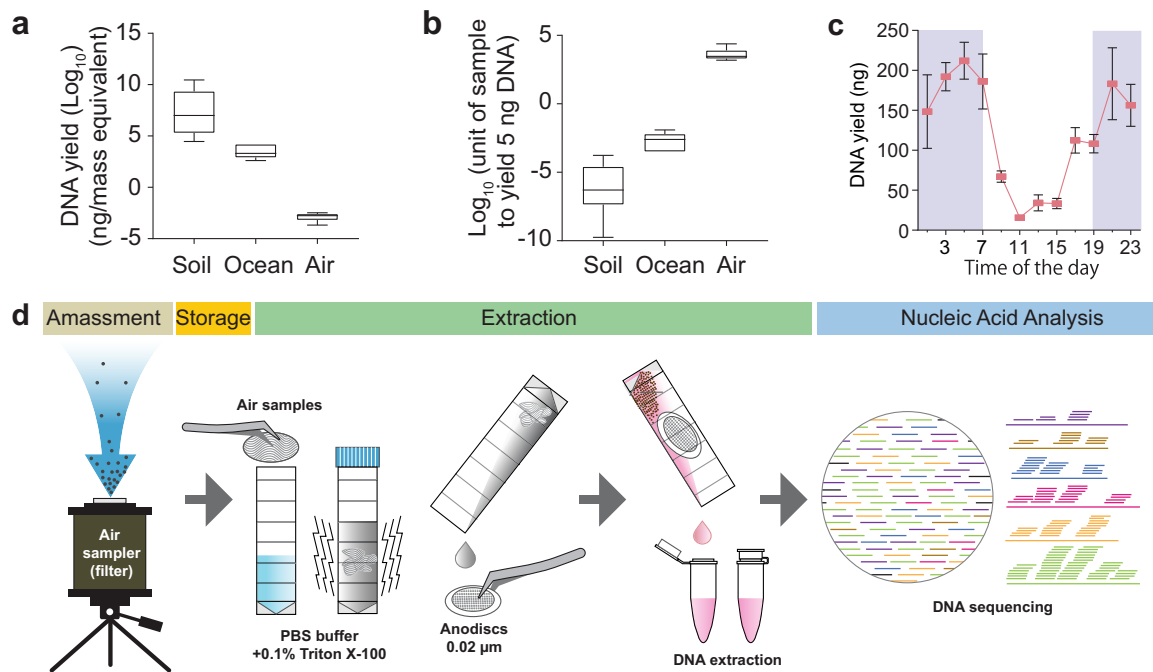
Here, we detail optimisation of multiple stages of an ultra-low biomass analysis pipeline for air samples, which can also be tailored to studies of similarly ultra-low biomass environments such as dust and surfaces. The versatility and robustness of the presented pipeline enable analysis of a wide range of environmental settings, both indoor and outdoor, encompassing a wide scope of climatic settings including tropical, temperate, desert and arctic regions.

## RESULTS

### Environmental samples: soil, water, air

Ecosystems and habitats are highly variable and complex, and hence a universal approach is not always applicable. Using DNA concentration as a proxy, terrestrial, aquatic and atmospheric ecosystems can harbour up to a six-log difference in microbial biomass (Fig. 1a). This results in vastly different sampling requirements and volumes for molecular analysis (Fig. 1b). In addition, biomass concentrations might follow cyclic processes resulting in density fluctuations, as shown in marine environments<sup>23</sup> as well as

<sup>1</sup>Singapore Centre for Environmental Life Sciences Engineering (SCELE), Nanyang Technological University, Singapore, Singapore. <sup>2</sup>Present address: Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-590, Brazil. <sup>3</sup>These authors contributed equally: Irvan Luhung, Akira Uchida. ✉email: scschuster@ntu.edu.sg



**Fig. 1 Challenges in air microbiome analysis.** **a** Total DNA yield (ng/mass equivalent) for soil, ocean water and air sample collected from the same proximity and processed with the same method. **b** estimated sample volume required to yield 5 ng of DNA. For box plots, the centre line, bound of box and whiskers represent median, 25th–75th percentile and min-to-max values, respectively. **c** Fluctuation of airborne biomass (ng) at different times of the day. The red dots and error bars are mean and standard deviation among the replicates. **d** Developed sampling and analysis pipeline for metagenomic analysis of ultra-low biomass environmental samples.

atmospheric environments where higher bioaerosol concentrations are typically observed at night<sup>22</sup> (Fig. 1c) or during haze events<sup>24</sup>.

To address the challenges in analysing a wide range of biomass concentrations at different spatial and temporal settings, we developed a robust ultra-low biomass pipeline, comprising the four-stages of amassment, storage, extraction and nucleic acid analysis. Parameters that impact upon the pipeline's efficacy were investigated, with the aim of enabling customisation (Fig. 1d). The summarised results are displayed in Fig. 2. The subsequent sections detail each investigated parameter individually.

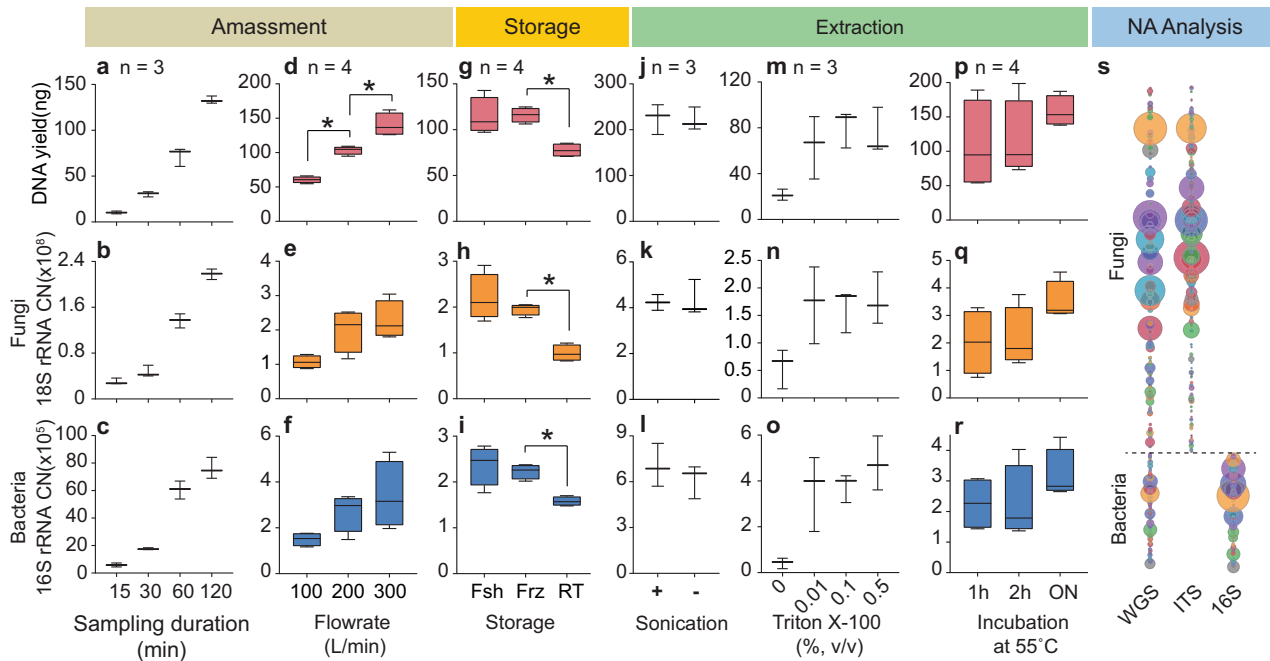
### Amassment

The ultimate success of sequencing and PCR-based analyses rests on sufficient quantities of nucleic acids being amassed, which for air sampling is a trade-off between sampling flow rate and sampling duration. While this study uses a filter-based sampler, other types of air samplers, such as liquid impingers serve a similar function and produce comparable results (Supplementary Fig. 8)<sup>25</sup>. For our purpose, ideal air samplers should be portable, battery-powered and have an acceptable noise emission (~50 dB).

The air sampling flow rate and duration were optimised to improve the temporal resolution of each sample from days, weeks or months to hours or even minutes, while still maintaining maximal taxonomic resolution. This was achieved by evaluating how these two factors directly impact the DNA quantity and metagenomic profile of the sample. Using 300 L/min flow rate, the minimal required sampling duration was investigated using different time-based sampling regimes (Fig. 3a). Sampling duration was segmented into sequentially doubling time intervals. For example, the first and second 15-min intervals (5:00–5:15 am and 5:15 to 5:30 am) were individually analysed and compared to a 30-min sample (5:00–5:30 am) taken in parallel. This process was undertaken for 15, 30 and 60-min intervals with a final sampling duration up to 180 min. Quantitative analysis showed consistently increasing DNA yields as a function of sampling duration (Fig. 2a–c). No notable loss of DNA yield was observed

within the tested range of duration (15 min–3 h). Within this range, combining two successive time segments resulted in similar DNA quantities as a single time segment of the combined duration, as quantified using Qubit and qPCR (Fig. 3b, Supplementary Fig. 1). Within the three investigated intervals (three duration groups each for Qubit, bacterial and fungal qPCR), the differences averaged 25%, with a median of 18%. Importantly, the microbial taxonomic profiles from comparable time intervals were not affected. This is demonstrated by the shift in relative abundances of taxa, such as *Kocuria palustris*, and *Leifsonia xyli*, between the two subsequent 15-min samples (Fig. 3c). Averaging these species compositions from the two subsequent 15-min filter samples resulted in abundances that mirror that of the 30 min time interval sample collected in parallel (Fig. 3d). This was consistent across all sampling duration regimes with three replicates each (BrayCurtis and Jaccard  $p > 0.05$ ).

The second experiment examined the impact of the air flow rate and the total volume of air sampled. With the sampling duration set at 2 h, airflow was varied between 100, 200 and 300 L/min, resulting in total air volumes of 12, 24 and 36 m<sup>3</sup>, respectively. The DNA yield and copy number of marker genes (16S and 18S rRNAs) increased as a function of air volume sampled (Fig. 2d–f). However, DNA concentration normalised per air volume diminished by up to 20% when the flow rate was increased from 100 to 300 L/min (Supplementary Fig. 2a). The diminishing return of amassment is likely due to decreasing particle retention efficiency at higher flow rates for extended periods of time<sup>26</sup>. For the purpose of this study, optimal sampling efficiency is forfeited in favour of higher flow rates (300 L/min) because the total amount of biomass collected per unit of time still out-performs the decrease in amassment efficiency. This enables measurements with higher time resolution within a day for environmental time-series studies. The biological significance of this was demonstrated by the discovery of diel dynamics of outdoor airborne microbial communities<sup>22</sup>.



**Fig. 2 Summary of quantitative analysis with DNA yield, 18S copy number (CN) and 16S copy number (CN).** **a–c** Assessment of air sampling duration from 15 min to 3 h. **d–f** Assessment of air sampling flow rate from 100 L/min to 300 L/min. **g–i** The integrity of sampled biomass when processed fresh (Fsh), stored in freezer for 5 d (Frz) or stored at room temperature for 5 d (RT). **j–l** Impact of sonication on DNA yield. **(m–o)** Impact of detergent addition at different concentrations (0.01–0.5% v/v) during filter sample wash. **p–r** Impact of extended pre-incubation (1 h to overnight) at 55 °C during DNA extraction. The centre line, bound of box and whiskers represent median, 25th–75th percentile and min-to-max values, respectively. **s** Whole-genome shotgun (WGS) and amplicon (ITS/16S) sequencing approaches. \* denotes statistical significance ( $p < 0.05$ ) tested with Mann–Whitney tests.

Further analysis demonstrated that flow rate does not impact the qualitative and quantitative assessment of metagenomic data (Supplementary Fig. 2b). The community structure (BrayCurtis,  $p > 0.05$ ) and richness (Jaccard,  $p > 0.05$ ) were not significantly different for samples collected with different flow rates.

### Storage

Analysis of the storage component in this pipeline evaluated the integrity (biomass quality and composition) of air filter samples stored under different conditions. The three conditions investigated were (i) instant processing (Fsh), (ii) 5-day storage at  $-20$  °C (Frz), and (iii) 5-day storage at room temperature (RT, average 23 °C, RH 65%).

No significant differences were observed between fresh and freezer samples in terms of both absolute (Qubit, qPCR) and relative (metagenomic) abundances. This suggests temporary freezer storage is a viable alternative to immediate filter processing. However, RT samples were significantly different from the fresh and freezer regimes in regard to DNA quantities (20–30% loss) (Fig. 2g–i). Also, a minor decrease of relative abundance of certain taxa was observed (BrayCurtis,  $p < 0.05$ ) (Fig. 4a); however, there was no loss in the number of species detected (Jaccard,  $p > 0.05$ ) (Supplementary Fig. 3). This outcome implies that microbial growth on the filter substrate is impeded within the course of several days, thus enabling sample collection for field surveys without the need for refrigeration<sup>27</sup>.

### Filter processing and DNA extraction

As library construction for DNA sequencing requires the removal of particle/biomass from the air filter substrate (referred to as filter processing), the protocol was optimised for efficient biomass retrieval. Importantly, the ultra-low biomass nature of the sample renders filter processing the most limiting, and hence, the most critical step across the entire pipeline for maximising yield.

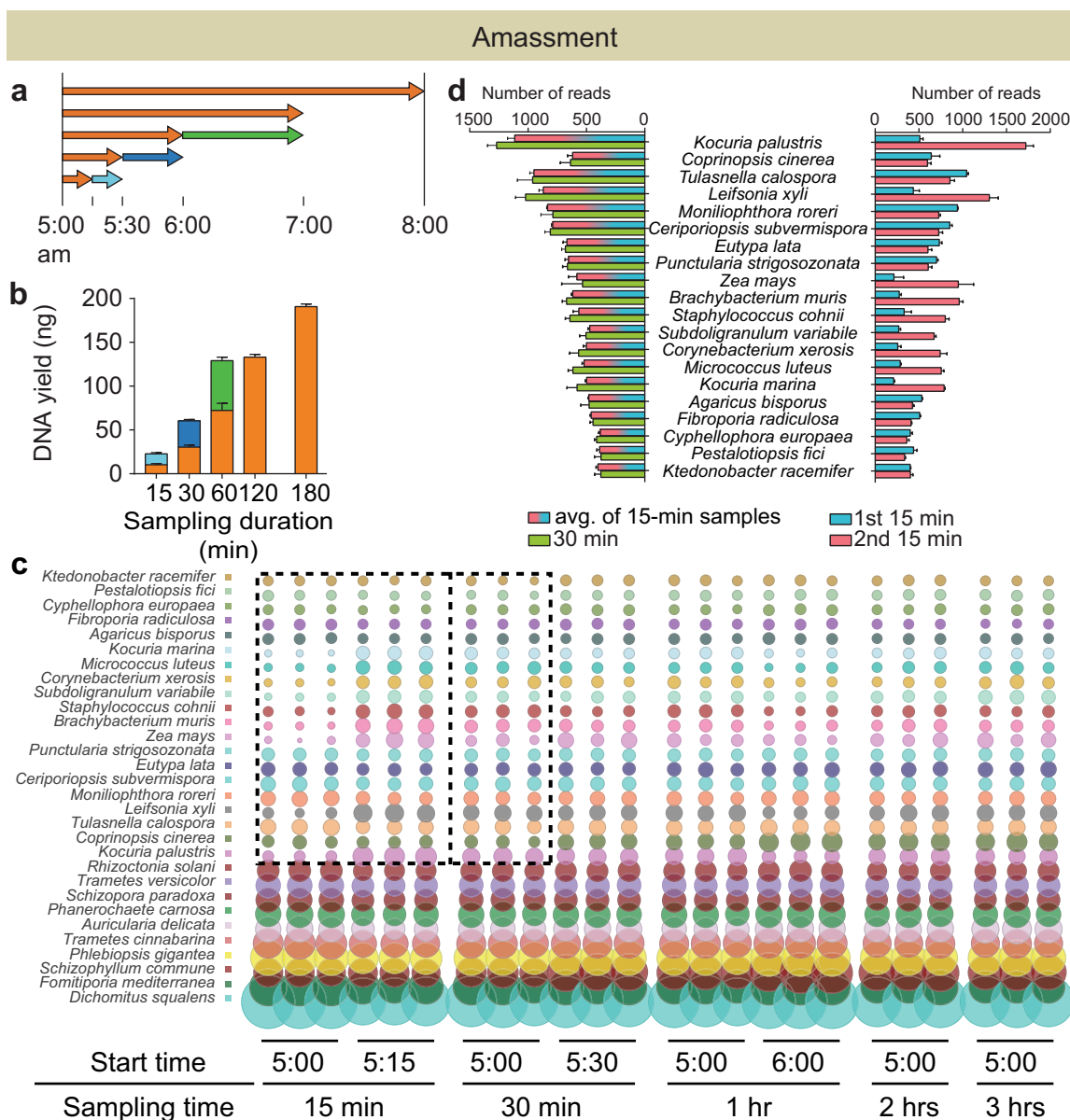
In general, filter samples can be processed in one of two ways, either direct DNA extraction on the filter, or by first removing the

biomass from the filter prior to DNA extraction. Direct DNA extraction was deemed inefficient as the filter absorbs most of the lysis buffer, which consequently inhibits cell lysis. In contrast, first removing the biomass by washing the filter in a buffer (PBS) and then concentrating on a thinner membrane with smaller mesh-size (0.2  $\mu$ m PES or Anodisc membrane)<sup>28</sup>, resulted in significantly higher DNA recovery (Fig. 4b).

To further improve biomass recovery, additional steps such as water-bath sonication (RT, 1 min)<sup>29,30</sup> and the use of detergent (Triton-X 100) during filter wash were tested. For comparison of samples processed with and without sonication, no significant difference in either quantitative or metagenomic analyses was found (Fig. 2j–l and Supplementary Fig. 4). In contrast, adding detergent during the filter wash significantly improved DNA yield (Fig. 2m–o). The hydrophobic nature of the air sampling filter impeded wetting by the wash buffer. Hence, particles were not effectively suspended in the wash buffer when mechanically agitated. The addition of non-ionic detergent, Triton X-100, at varying concentrations (%v/v) (PBS-T) to the initial PBS buffer was effective in overcoming this challenge.

The detergent wash resulted in significant differences in absolute and relative abundance analyses, especially in the instance of bacteria. DNA yield, as well as copy number of bacterial 16S and fungal 18S rRNA genes, increased 2.4, 8.6 and 2.0-fold, respectively (Fig. 2m–o). The metagenomic analysis confirmed this finding (BrayCurtis and Jaccard,  $p < 0.05$ , Fig. 4c, Supplementary Fig. 5). The number of detected bacterial taxa increased eight-fold compared to a 1.3-fold increase in fungal taxa. Expectedly, PBS-T treated samples also showed greater taxonomic diversity (Fig. 4c).

Varying concentrations of Triton X-100 (0.01, 0.1 and 0.5% (v/v)) in PBS were investigated, with no significant difference between the three concentrations for quantitative analyses (Fig. 2m–o). However, metagenomic analysis identified notable differences in microbiome composition (BrayCurtis  $p < 0.05$ , Supplementary Fig.



**Fig. 3 Sampling duration assessment.** **a** Illustration of different time-based sampling regimes. **b** Comparison of DNA yield (ng) between the corresponding sampling regimes, e.g. first 15-min yield (orange) + second 15-min yield (light blue) compared to first 30-min yield (orange). The bars represent mean values and the error bars were standard deviation among the replicates. **c** Taxonomic compositions of the top 30 species, the highlighted portion focuses on species which shifted in abundance between the first and second 15-min samples. **d** Comparison of relative abundances of the selected species, the first and second 15-min samples were averaged and compared to the taxonomic composition of first 30-min sample. The bars represent mean values and the error bars were standard deviation among the replicates.

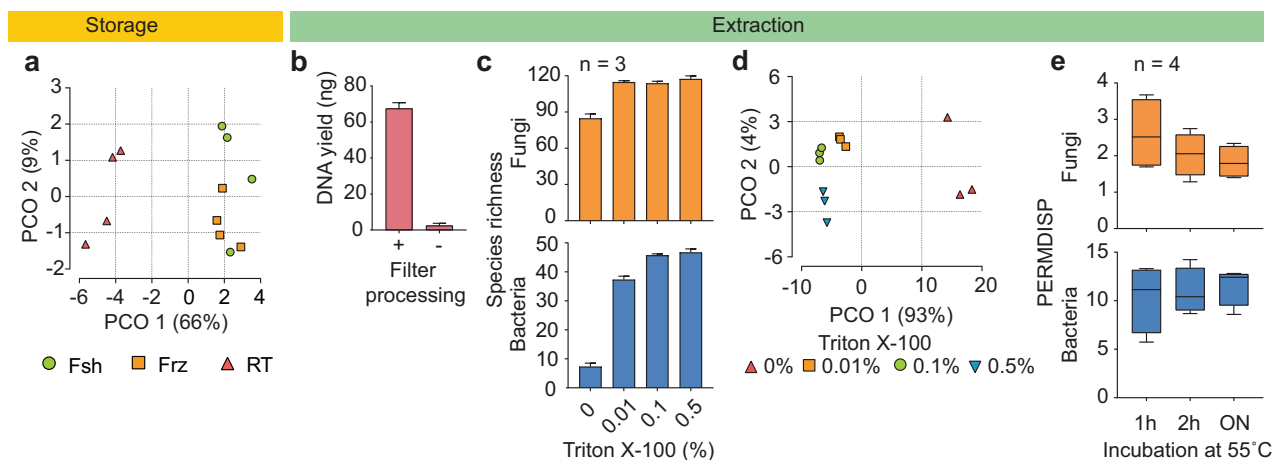
5a–b) driven by an increase in bacterial taxa. Increasing Triton X-100 beyond 0.1% concentration, yielded no significant further gains (Fig. 4d). Hence, Triton X-100 at 0.1% was deemed sufficient for wetting the filter and releasing attached bioaerosol particles into the buffer medium. Despite the 0.1% concentration of Triton X-100 being above the critical micelle concentration<sup>31</sup>, Triton X-100 did not trigger unwanted premature lysis of microbial cells, as there were no significant differences in DNA yield between the three concentrations. If premature lysis occurred, extracellular DNA would not have been retained on the subsequent Anodisc membrane, resulting in lower DNA recovery.

Following filter processing, the recovered biomass was filtered through a 0.02  $\mu\text{m}$  pore-sized Anodisc membrane (Whatmann, USA) mounted on a vacuum manifold (Fig. 1d), with the Anodisc

directly fitting into the DNA extraction kit bead tube. DNA extraction used the standard protocol of the extraction kit with slight modification to improve lysis<sup>26</sup>. In this regard, the addition of overnight pre-incubation of the samples at 55 °C is recommended as it improves evenness among the samples, especially for the representation of fungal taxa, as shown by the quantitative and PERMDISP analysis (Fig. 2p–r, Fig. 4e).

#### Nucleic acid analysis of ultra low biomass samples

The outcome of the above sample processing pipeline results in double-stranded DNA samples (in the range of 0.1–7.1 ng DNA/m<sup>3</sup> of air sampled). These can subsequently be analysed, not only by amplification-based techniques (16S/ITS), but also via direct DNA



**Fig. 4 Storage and biomass extraction.** **a** Principal coordinate analysis (Bray-Curtis) on genus level for samples processed fresh (Fsh), stored in freezer (Frz) and room temperature (RT). **b** Comparison of DNA yield (ng) with (+) and without (–) the filter wash step. **c** Total identified species for fungi (orange) and bacteria (blue) for samples processed with different concentration of detergent (0–0.5% v/v) during the wash step. The bars represent mean values and the error bars were standard deviation among the replicates. **d** Principal coordinate analysis (Bray-Curtis) on genus level for samples processed with different concentration of detergent (0–0.5% v/v) during the wash step. **e** PERMDISP analysis for samples processed with extended incubation at 55 °C prior to cell lysis. The centre line, bound of box and whiskers represent median, 25th–75th percentile and min-to-max values, respectively.

sequencing (shotgun), resulting in either gene-based or metagenomic profiles of airborne environmental communities. Both approaches, 16S/ITS amplicon and whole-genome shotgun metagenomic (WGS), produce sequence data that may be compared against publicly available data archives. For the remainder of this manuscript, the advantages and disadvantages of both techniques will be discussed in relation to ultra-low biomass analysis.

Amplicon-based sequencing approaches have been the method of choice in the majority of past bioaerosol studies<sup>13–18</sup>. This was due to the assumption that the low amount of amassable biomass from air was insufficient for shotgun metagenomics<sup>32</sup>. Our study shows that the above-described ultra-low biomass air sampling and processing pipeline is capable of robustly producing metagenomic datasets, as demonstrated i) for a range of DNA input amounts, ii) by the reproducibility among replicates, iii) by the robustness of air samples analysis from various climatic conditions and iv) contamination control.

**Required input amount:** from the same DNA sample, a range of DNA input amounts for shotgun metagenomic sequencing and analysis (0.5–10 ng) were tested. Using our pipeline, taxonomic representation for each DNA input condition was visualised at the species level using bubble charts (Fig. 5a). For the tested range of 0.5–10 ng, no significant change of species-level composition was observed. The species-level metagenomic profile for each sample was consistent even when the PCR cycles required during DNA library construction were increased from 6 to 15 (Fig. 5a).

**Reproducibility:** An experimental time series of outdoor air in a tropical setting<sup>22</sup> was used to assess sample-to-sample variability. Over 24 h, air samples were collected at 2-h time intervals (12 time points) in triplicate. The metagenomic profiles of samples within the same replicate group were highly consistent, with an average similarity of 91% (87–95%, SIMPER analysis). The taxonomic profiles, however, were distinct between sampling time points (Fig. 5b). The higher variability observed for day-time samples can be attributed to increased atmospheric turbulence due to convection, while a narrower range was observed during night-time hours.

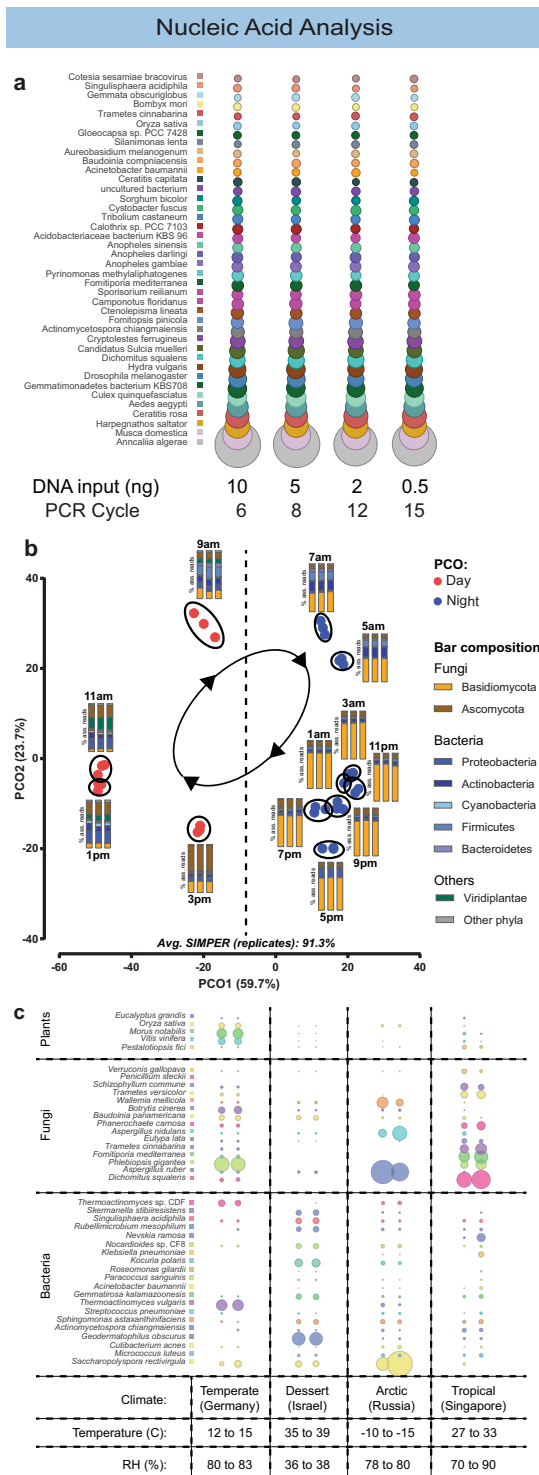
**Robustness:** The above-tested range of 0.5–10 ng of DNA templates, with their respective PCR cycles (15 to 6 cycles), was suitable for a global air microbiome survey that involved a wide range of environmental conditions. The pipeline presented here robustly produced metagenomic datasets from air samples

collected in locations with a diverse range of temperature (–10 to 39 °C) and humidity (36–90%), within the four climatic zones (temperate, desert, sub-arctic and tropical) (Fig. 5c).

**Contamination control:** The negative controls consisted of filter blanks (clean, unused filter) mounted on the air samplers for 1 min without airflow, which were then transported and processed in an identical manner to air samples. The DNA yield from negative controls was not detectable (Supplementary Fig. 7a). The number of reads generated by Illumina sequencing were on average 1000-fold less for the negative controls compared to the air samples (Supplementary Fig. 7b), with taxonomic analysis indicating human contamination as the most likely source (Supplementary Fig. 7c). The number of reads from our air samples which could be mapped back to the filter blanks were very low and they were removed by our statistical analysis threshold (<0.05% of assigned reads). It can be deemed that despite the ultra-low biomass nature of our analytical pipeline, contamination is not a concern (Supplementary Discussion 7).

In a final step, extracted genomic DNA from the pipeline was analysed by both metagenomic and 16S/ITS amplicon sequencing, resulting in sets of distinct taxonomic profiles based on their respective databases (Fig. 6a). For fungi, results from both sequencing analysis methods concur with the observed trends for the specific abundances of microbial taxa during day/night at higher taxonomic resolution, e.g., Ascomycota being prevalent during day-time and Basidiomycota during night-time. The 16S amplicon analysis, however, was less robust as three out of four samples resulted in no detectable PCR product, even with higher DNA input (4–46 ng) and additional PCR cycles (Fig. 6a). This was caused by low amounts of 16S rDNA gene template in tropical air samples (Supplementary Fig. 9). The only successfully analysed 16S sample resulted in a similar taxonomic profile to that of the WGS pipeline at the phylum level, with Firmicutes dominating over Actinobacteria and Proteobacteria.

The above analysis highlights biases in the success rate of fungal ITS and/or bacterial 16S amplification for air samples from a diverse range of environmental conditions. Numerous studies have reported similar challenges<sup>13,16</sup>. In contrast, regardless of potential inhibitor content and/or taxonomic composition of the air samples, the WGS pipeline consistently captured the biological diversity of airborne microbial communities in various climatic



**Fig. 5 Whole genome shotgun (WGS) sequencing of air samples.** **a** Comparison of taxonomic profile at species level for the same air sample that was subjected to WGS sequencing with different DNA input amounts (10–0.5 ng). **b** Reproducibility of samples collected at the same time and location (triplicates) illustrated in principal coordinate analysis (Bray–Curtis) at species level. The bars show the microbial community composition of the triplicates in % of assigned reads. **c** Robustness of air sampling and processing pipeline tested at locations with temperate, dessert, sub-arctic and tropical climates.

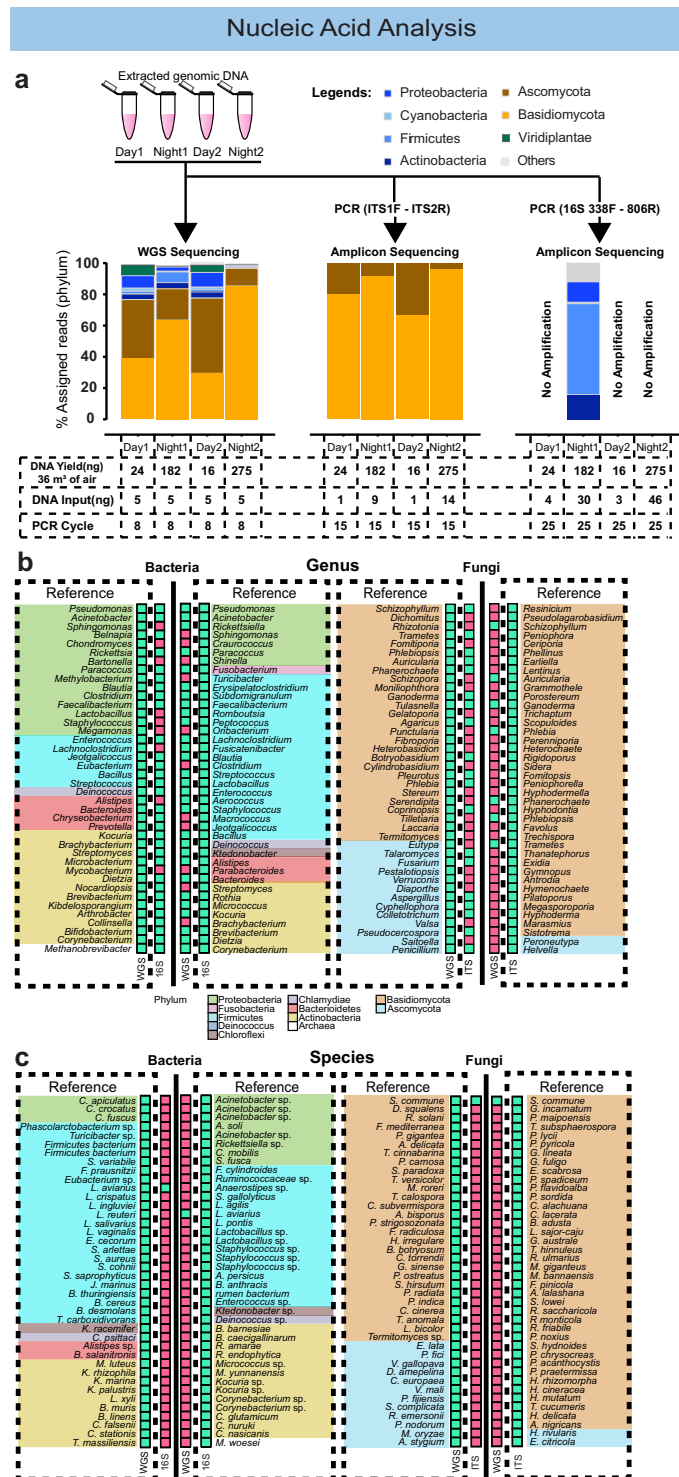
conditions (Figs. 5c, 6a). Moreover, unlike the single gene amplicon approach, the WGS pipeline directly compared DNA read abundances from a diverse set of taxa (bacteria, fungi, plants and others) at a single quantitative scale.

In contrast to phylum level analysis, WGS and amplicon analytical pipelines are substantially less congruent at the genus or species level, due to the respective database sizes. Our metagenomic reads were aligned to the non-redundant (nr) database and assigned to taxa using the MEGAN software<sup>33</sup>, while the amplicon reads were aligned to the 16S SILVA database for bacteria<sup>34</sup> and ITS UNITE<sup>35</sup> database for fungi using blastn<sup>36</sup>. The resulting taxonomic classifications from the two analysis approaches show significant agreement at higher taxonomic levels (e.g., up to phylum level). At genus and species levels, taxonomic concordance is diminished, as shown for the top 40 most abundant taxa for both analysis types (Fig. 6b, c). In this regard, the metagenomic and 16S amplicon approach agree in 72–78% of instances on the genus level. However, only one out of 40 taxa (2.5%) was in agreement on the species level. This concordance is even less for fungal taxonomy. On the genus level, 19 out of 40 taxa (47.5%) were in agreement when the WGS was used as a reference. When the ITS was chosen as a reference, seven out of 40 (17.5%) assignments were in agreement. As observed for bacteria, only 1 out of 40 taxonomic assignments was shared on a species level. In general, the amplicon databases possess a much larger representation of fungal and bacterial taxa. The higher overlap for bacteria was likely due to higher representation of bacterial genomes in the nr sequence database due to increasing accessibility for generating genome-wide data for small microbial genomes. In contrast, the accessibility does not extend to genome sizes exceeding 100 MB for some fungal organisms. In this regard, the sequencing, assembly and annotation of fungal genomes are still challenging.

rDNA sequences generated by both sequencing methods concur when analysed for marker gene content. In this regard, the metagenomic datasets analysed in this study contain about 1% rDNA genes (ITS and 16S), which can be aligned to 16S SILVA and ITS UNITE databases. The metagenomics rDNA read analysis and the amplicon sequencing results produced highly overlapping taxonomic profiles for the top 40 most abundant taxa for fungi and the top 10 most abundant taxa for bacteria (Supplementary Fig. 6). With metagenomic sequencing becoming more accessible, it is therefore possible to combine the benefits of 16S- and ITS-based taxonomy to investigate understudied ultra-low biomass environments, while simultaneously enabling taxonomic and functional analyses<sup>37</sup>.

## DISCUSSION

The here-presented air sampling and analysis pipeline enable qualitative and quantitative assessment of microbial diversity in an ultra-low biomass ecosystem. The 5–7 log difference in biomass concentration of air samples, compared to seawater or soil, requires sufficiently large volumes of air to be sampled. Based on our optimisation results, we propose default sampling parameters of 300 L/min for 2 h. This enables DNA accumulation rate which is ~8–170-fold higher than reported in recent studies<sup>26,28,38</sup> (Supplementary Table 1). This large improvement allows for shorter sampling time ( $\geq 15$  min), while still enabling WGS metagenomic analysis with species-level taxonomic classification. Such high temporal and taxonomic resolution are crucial for ecological studies of air microbiomes, which rapidly respond to diel dynamics or sudden environmental changes. It should be noted that factors such as time of sampling within a day, sampling duration and climatic settings of the sampling location impact the



**Fig. 6 Comparison of taxonomic profiles between WGS and amplicon sequencing pipelines.** **a** Taxonomic profile of WGS, ITS amplicon and 16S amplicon pipeline at phylum level of four independently collected air samples (two days and two nights). **b** Presence–absence comparison of the top 40 most abundant genus for bacteria (WGS vs 16S) and fungi (WGS vs ITS). **c** Presence–absence comparison of the top 40 most abundant species for bacteria (WGS vs 16S) and fungi (WGS vs ITS).

analysis outcome, and therefore comparability between the above studies. Our proposed method, however, has also been evaluated for its robustness across a wide range of environmental settings (arctic, desert, temperate and tropical climatic zones) (Fig. 5c).

Further, our results indicate that biomass amassed from air samples using filter-based devices during remote fieldwork may

be stored at room temperature for extended periods of time with tolerable loss of extractable DNA (20% in 5 d) and without compromising microbial community structure. While these effects could potentially be counter-acted by nucleic acid stabilisation methods<sup>39</sup>, this approach is not recommended during sampling campaigns, as it would require additional handling of the air filter

samples in the field. This could result in contamination and complicate transport due to the introduction of liquid materials (e.g., commercial air travel). The advantage of dry storage and transport also does not extend to other types of air samplers, such as liquid impingers.

For nucleic acid extraction, it could be shown that the amassed biomass should not be extracted directly on the filter, but rather first be removed owing to the adherence of the low quantities of nucleic acids to the large surface area of the filter membrane. Therefore, extraction and wash buffer conditions should be optimised to enable the extraction of sub-nanomolar concentrations of DNA/RNA. This optimisation includes the use of detergent and extended incubation times. In particular, the addition of non-ionic detergents, such as Triton X-100, significantly increases the recovered biomass, while extended incubation times improve the evenness of the large sets of samples. This observation is also highly relevant in the context of sampling potentially infectious biological materials, such as airborne retroviruses, which can concurrently be inactivated with Triton X-100<sup>40</sup>.

Both metagenomic and amplicon sequencing methods can be applied to air samples (Fig. 2s). The metagenomic approach is advantageous with regards to enabling simultaneous functional and taxonomic analysis and has the advantage that bacteria and fungi can be analysed within the same quantitative scale. Further, the rapid expansion of the public WGS databases continues to enable species-level taxonomic identification at an increasing rate. In contrast, the content of amplicon sequencing databases (ITS or 16S) are likely to grow at a slower rate, given the increased accessibility of WGS.

While our study demonstrated that the extracted DNA from the ultra-low biomass pipeline was sufficient for WGS and ITS amplicon analyses, 16S amplicons did not perform equally well for tropical air samples (Fig. 6a). This may be due to the fact that the DNA library construction for WGS is less sensitive to inhibitors and the relative ratio of bacterial vs. fungal DNA. Both factors impact on the efficacy of the polymerase chain reaction. Nevertheless, specific gene marker/amplicon analysis can be advantageous for studies that target well characterised, less diverse microbial communities.

Finally, due to database biases, both methods appear to converge on the phylum level, but to a lesser degree at the genus level. On the species level both methods do not produce significant agreement. To harness the advantages from both sequencing technologies, it is beneficial to combine both approaches by also analysing the rDNA sequences from the metagenomic data (Supplementary Fig. 6). The results from this combined analysis enable data interpretation from a single data source (metagenomic data), to inform both WGS and marker genes analysis pipelines.

The here-presented methodology is limited by the size range of the chosen filter medium (0.5–>10 µm, <50% efficiency for particles <0.5 µm). As this study aims to reduce required sampling times, total suspended (biological) particles (TSP) need to be collected and analysed. While this study does not profile particle size range, recent studies have demonstrated that the most relevant airborne bacteria and fungi fall within the size range of the filter medium<sup>13,15</sup>.

In summary, the above-described ultra-low biomass analysis pipeline provides detailed insights into the factors that influence analysis outcomes for low-biomass microbial environments. High volumetric air sampling techniques in combination with applied nucleic acid analysis, results in high temporal and taxonomic resolution of inherent airborne microbial communities. The presented findings are potentially also applicable to other low-biomass environments, such as dust and surfaces.

## METHODS

### Air sampling

Air samples for optimisation purposes were collected in Singapore at a roof-top balcony of a university building (N1.346247, E103.679467). As the study focuses on improving the time resolution of the analysis, a high-flow rate, filter-based air sampler (SASS3100, Research International, USA) was used to collect total suspended particles (TSP) with no size cut-off. The filter medium was SASS Bioaerosol electret filter (6 cm diameter, expected 50% efficiency for 0.5 µm particle size, Research International, USA). For sample collection, air samplers were attached upright on a tripod 1.5 m above the concrete floor of the balcony.

In addition to Singapore, samples from different climatic settings were collected in a consistent manner from sites in Germany, Russia and Israel to test the robustness of the proposed pipeline. These international locations showed contrasting settings for temperature (T) and relative humidity (RH).

After sampling, the filters were returned to their original filter pouches and transported to the laboratory for direct processing or storage at –20 °C. Information on exact sampling time, flow rate, duration and the environmental parameter measurements of all sampling activities used in this study can be found in Table 1.

### Temperature and relative humidity

Temperature (T) and relative humidity (RH) at the sampling site were measured using HOBO Temp/RH 2.5% Data Logger (Onset, USA).

### Filter processing, DNA extraction, quantitation and sequencing

All filter samples were subsequently processed for DNA extraction, quantitation, qPCR, metagenomic sequencing and computational analysis as described in our previous study<sup>22</sup>. In brief, the filter samples were first washed 3 times using 2 mL of phosphate-buffered saline (pH 7.2) with 0.1% (v/v) Triton X-100 assisted with water-bath sonication at room temperature for 1 min. After washing, the suspension liquid was concentrated onto a 0.02 µm Anodisc filter (Whatman, UK) using a vacuum manifold (DHI, Denmark). DNA was then extracted from the Anodisc with the DNeasy PowerWater kit (Qiagen, USA) following the manufacturer's standard protocol with modifications to increase DNA yield<sup>26</sup>.

Final DNA solution was subjected for fluorometer quantification, qPCR and shotgun metagenomic sequencing. Fluorometer quantitation was measured with Qubit 2.0 (Invitrogen, USA) using the High Sensitivity double stranded DNA (HS dsDNA) kit. Taqman qPCR assays with universal bacterial (16S rRNA gene)<sup>41</sup> and fungal (18S rRNA gene)<sup>42</sup> primer set and probes were used to quantify the copy numbers of bacteria and fungi, respectively. The complete list of primers can be found in Table 2.

For direct metagenomic sequencing, libraries were prepared using Swift Biosciences' Accel-NGS 2S Plus DNA Library kit following the standard protocol. All libraries were subsequently dual-barcoded with Swift Biosciences' 2S Dual Indexing kit. PCR amplification selectively enriches for library fragments that have adapters ligated on both ends. The number of cycles were adjusted based on the starting amount of DNA (8–15 cycles). Upon pooling at equal volumes, libraries were sequenced on Illumina HiSeq2500 Rapid runs at a final concentration of 10–11 pM and a read-length of 251 bp paired-end (Illumina V2 Rapid sequencing reagents). Each ultra-low biomass sample was sequenced to a depth of at least two million paired-reads.

Raw reads from the sequencer were first trimmed from adapter sequences, low quality bases (<20 score) and short reads (<30 bp) using Cutadapt (v.1.8.1)<sup>43</sup>. The processed reads were then aligned against the NCBI's NR database (v.25-02-2016) using RAPSearch2 (v.2.15)<sup>44</sup>. Results from the RAPSearch2 alignment were finally converted to read-match archive (rma) to be visualised with MEGAN5 software<sup>33</sup>.

### Experimental parameters optimisation

Important parameters for sampling, extraction and sequencing were tested and optimised based on absolute (fluorometer and qPCR) and relative abundance assessment (DNA sequencing). Importantly, it should be noted, that only samples collected at an identical time and location may be compared. Therefore, it is mandatory as an experimental setup to deploy multiple air samplers for each set of the parameter optimisation experiments. This is due to the high volatility of biomass concentration and composition of air, particularly when sampling at different time points throughout day and night. The replicability and robustness of this study



**Table 1.** Details of sampling activities.

Sample set	Sampling date	Sampling time, duration and flow rate	Temperature (°C)	RH (%)	Rain	Sample size	No. of samples
Singapore							
1	29-Nov-17	01:00–03:00 (2 h, 100 L/min, 200 L/min, 300 L/min)	24.8–25.3	98–100	No	4	12
2	15-Dec-17	05:05–08:05 (15 min, 30 min, 1 h, 2 h, 3 h, 300 L/min)	24.7–25.7	99–100	No	3	27
3	29-Nov-17	06:15–08:15 (2 h, 300 L/min)	24.6–25.4	99–100	No	4	12
4	23-Feb-17 24-Feb-17	17:00–17:00 (2 h, 300 L/min)	24.0–34.0	63–100	Yes	3	36
5	8-May-16 9-May-16	17:00–17:00 (2 h, 300 L/min)	28.0–33.0	59–89	No	3	36
6	29-Nov-17	03:50–05:50 (2 h, 300 L/min)	24.5–24.8	99–100	Yes	3	12
7	28-Nov-17	20:40–22:40 (2 h, 300 L/min)	24.9–25.3	98–99	No	3	12
8	24-Nov-17	05:00–07:00 (2 h, 300 L/min)	23.9–24.5	99–100	No	4	12
9	22-Nov-17	23:00–01:00 (2 h, 300 L/min)	26.0–26.5	97–99	No	3	3
10	23-Nov-17	11:00–13:00 (2 h, 300 L/min)	29.0–30.0	77–80	No	2	2
11	27-Nov-17	23:00–01:00 (2 h, 300 L/min)	24.5–25.5	93–96	No	3	3
12	28-Nov-17	11:00–13:00 (2 h, 300 L/min)	28.5–29.5	75–77	No	2	2
13	29-Aug-16	13:00–15:00 (2 h, 300 L/min)	31.0–32.0	70–80	No	1	1
14	30-Aug-16	13:00–15:00 (2 h, 300 L/min)	31.0–32.0	70–80	No	1	1
15	21-Sep-15	13:30–15:30 (2 h, 300 L/min)	31.0–32.0	63–70	Yes	1	1
Germany							
16	30-Jul-17	12:00–14:00 (2 h, 300 L/min)	12.0–15.0	80–83	No	2	2
Israel							
17	4-Jul-17	08:30–10:30 (2 h, 300 L/min)	35.0–39.0	36–38	No	2	2
Russia							
18	2-Dec-17	09:00–11:00 (2 h, 300 L/min)	–10.0––15.0	78–80	No	1	1
19	3-Dec-17	15:00–17:00 (2 h, 300 L/min)	–10.0––15.0	78–80	No	1	1
<b>Total no. of samples (including blanks):</b>							<b>183</b>

**Table 2.** List of primers and probes applied in the study.

Name	Sequence	Notes
16S 341F	5'-CCTACGGGDDGGCWGCA-3'	Bacteria qPCR
16S 805R	5'-GGACTACHVGGGTMTCTAATC-3'	Bacteria qPCR
Taqman probe	6FAM-5'-CAGCAGCCGCGGTA-3'-BBQ	Bacteria qPCR probe
FungiQuant-F	5'-GGRAAATCACCAGGTCCAG-3'	Fungi qPCR
FungiQuant-R	5'-GSWCTATCCCAKACGA-3'	Fungi qPCR
FungiQuant-PrlLNA	6FAM-5'-TGGTGCATGGCCGTT-3'-BBQ	Fungi qPCR probe
16S 341F Illumina	5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CCT ACG GGN BGC ASC AG -3'	Amplicon for bacteria
16S 805R Illumina	5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGG ACT ACH VGG GTW TCT AAT -3'	Amplicon for bacteria
ITS1F Illumina	5'- TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CTT GGT CAT TTA GAG GAA GTA A -3'	Amplicon for fungi
ITS2R Illumina	5'- GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGC TGC GTT CTT CAT CGA TGC -3'	Amplicon for fungi

was, therefore, enabled through simultaneous deployment of up to 12 air samplers at any given time ( $n = 12$ ).

Comparison to other types of environmental samples: The ultra-low concentration of airborne biomass was investigated relative to other types of environmental samples. To negate possible differences due to sampling location and/or processing method, soil (1 gram per sample extraction), water (1 mL per sample extraction) and air samples (300 L/min, 2 h sampling duration) were collected within the same proximity (in Singapore) and were subsequently processed with identical protocol. Only DNA yield (ng/unit mass or volume of the samples) was assessed for this experiment.

The assessment parameters are sampling duration and sampling flow rate. *Sampling duration experiment:* With a fixed air flow rate (300 L/min,  $n = 3$ ), sampling duration was varied at 15, 30, 60, 120 and 180 min. Further, multiple shorter duration samples were also compared to longer duration samples with matching time segments, i.e. first and second 15 min samples were compared to the matching 30 min sample. *Air flow rate experiment:* With a fixed duration (2 h), three groups of air samplers ( $n = 4$ ) were run at the same time with varying flow rate at 100, 200 and 300 L/min. The experiments were assessed based on the impact of sampling duration and airflow variations on DNA quantity and microbial composition.

**Sample storage experiment.** Three sets of air samples collected simultaneously (300 L/min, 2 h,  $n = 4$ ) were subjected to the following storage regimes; direct processing (fresh),  $-20^{\circ}\text{C}$  storage for 5 days (freezer) and room temperature storage for 5 days (RT) and compared for both DNA quantity and microbial profiles.

Parameters optimised for filter processing and DNA extraction were the use of sonication, detergent and impact of pre-incubation. **Sonication experiment:** Two sets of air samples collected at the same time (300 L/min, 2 h,  $n = 3$ ) were subjected to filter washing with the room temperature water-bath sonication step included and excluded. **Detergent experiment:** Four sets of air samples collected at the same time (300 L/min, 2 h,  $n = 3$ ) were washed with buffer containing four different concentrations of non-ionic detergent Triton-X 100 (%v/v): No detergent (0%), 0.01, 0.1 and 0.5%. **Pre-incubation experiment:** Three sets of air samples collected at the same time (300 L/min, 2 h,  $n = 4$ ) were subjected to three different durations of pre-incubation in  $55^{\circ}\text{C}$  water bath prior to proceeding with the subsequent lysis step of the DNA extraction. The durations were 1 h, 2 h and overnight (14–16 h). These durations were selected to enable the completion of the entire extraction process (filter washing and DNA extraction) within a standard working day (~8 h).

All the above experiments were assessed based on DNA quantity and microbial profiles of the resulting analysis.

The DNA sequencing result was evaluated for the DNA input amount, reproducibility, robustness and taxonomic classification difference between metagenomics and amplicon. **DNA input experiment:** From a given extracted DNA sample, four different DNA input amounts for direct metagenomic sequencing were tested: 10 ng, 5 ng, 2 ng and 0.5 ng. The number of PCR cycles during library construction were adjusted based on the DNA amount. The final result was assessed based on the taxonomic composition of the sequencing analysis. **Reproducibility between replicates:** A set of time series samples was analysed to investigate the similarity of the metagenomic profiles between the replicates. The time-series data contains twelve sets of time points with three replicates each. Each set was collected with 300 L/min flow rate and 2-hour sampling duration, spanning across 24 h. **Robustness across a range of climatic settings:** Air samples collected from locations with different climates (highly variables T and RH) were analysed regarding the success rate of DNA sequencing library construction due to varying amounts and quality of DNA input. 300 L/min flow rate and 2 h sampling duration were used to collect samples in Germany (temperate), Israel (dessert) and Russia (sub-arctic). **Comparison of shotgun metagenomic and amplicon marker gene sequencing:** The two sequencing approaches were evaluated using taxonomic assignments from identical sets of extracted air samples. DNA samples were split for shotgun metagenomic, 16S bacterial amplicon and ITS fungal amplicon sequencing. The sequencing and analysis methods for the bacterial and fungal amplicon sequencing are detailed in the following section.

### PCR-based amplicon sequencing and analysis

A subset of our ultra-low biomass samples were also subjected to amplicon sequencing for direct comparison with the shotgun metagenomic sequencing approach. For these samples, the first stage PCR was performed with the extracted genomic DNA as a template and the ITS1F-ITS2R<sup>45</sup> primers for fungi and 16S 341F-805R<sup>46</sup> primers for bacteria. Details of these primer sequences can be found in Table 2. KAPA HiFi HotStart master mix was used with a total reaction volume of 25  $\mu\text{L}$ . For DNA input amount, 3  $\mu\text{L}$  and 10  $\mu\text{L}$  of DNA templates were used for fungi and bacteria, respectively. The cycling condition was  $95^{\circ}\text{C}$  for 3 min, amplification cycles with  $95^{\circ}\text{C}$  for 30 s,  $65^{\circ}\text{C}$  for 30 s,  $72^{\circ}\text{C}$  for 30 s, and a final extension at  $72^{\circ}\text{C}$  for 5 min. The fungal samples were amplified with 15 cycles and the bacteria samples were amplified with 25 cycles. The PCR products were then purified with AMPure XP beads (Beckman Coulter) before performing the second stage PCR.

The second stage PCR (Indexing PCR) was performed according to the recommendations in Illumina's "16S Metagenomic Sequencing Library Preparation" application note. This step uses a limited cycle PCR to complete the Illumina sequencing adapters and add dual-index barcodes to the amplicon target. Five microliters of the intermediate PCR product from the first stage PCR (Amplicon PCR) were used as template for the indexing PCR and samples were amplified with eight PCR cycles. Nextera XT v2 indices were used for dual-index barcoding to allow pooling of the amplicon targets for sequencing.

Finished amplicon libraries were quantitated using Promega's QuantiFluor dsDNA assay and the average library size was determined on an Agilent TapeStation 4200. Library concentrations were then normalised to 4 nM and

validated by qPCR on a QuantStudio-3 real-time PCR system (Applied Biosystems), using the Kapa library quantification kit for Illumina platforms (Kapa Biosystems). The libraries were then pooled at equimolar concentrations and sequenced on the Illumina MiSeq platform with 20% PhiX spike-in and at a read-length of 300 bp paired-end (MiSeq V3 reagents).

After sequencing, raw reads were first trimmed from adapter sequences, low-quality bases and short reads using Cutadapt (v.1.8.1)<sup>43</sup>. After trimming, the R1 and R2 reads were first paired with minimum overlap of 10 bp and subsequently aligned against UNITE ITS database (v.7.1) for the ITS sequences and SILVA 16S database (release 132) for the 16S sequences using command line blastn<sup>36</sup> (version 2.2.28+). Results from blastn alignments were also converted to read-match archive (rma) format for visualisation with the MEGAN5 software to facilitate direct comparison with the metagenomic sequencing analysis. The default LCA parameters were used.

### Statistical analysis

For quantitative analysis from Qubit 2.0 Fluorometer and qPCR, all statistical tests were conducted with Mann–Whitney test. As mentioned previously, we acknowledge the limitations of these tests due to the relatively low number of observations ( $n = 3$  or  $n = 4$ ) for each set of samples. Due to the volatile nature of air sample, only samples collected at the same time and location can be directly compared. Thus, the number of deployments was limited by the number of samplers which could be deployed at a given time ( $n = 12$ ).

For metagenomic analysis, significant differences between groups of samples were mainly determined by ANOSIM test based on distance matrices between the samples compared. Distance matrices were created through PRIMER7 software based on taxa (genus level, cut-off at 0.05% of total assigned reads) read counts of each sample generated by MEGAN5. The distance matrix calculated based on Bray–Curtis algorithm was used to evaluate proportional difference (community structure) of the microbial communities between samples, while the distance matrix calculated based on Jaccard algorithm was used to determine presence–absence difference (community membership/richness) of different taxa detected in the compared group of samples. For reproducibility assessment among replicates, environmental time series data were used in which air samples with two-hourly time resolution were collected in 24 h with three replicates each. Similarity percentage (SIMPER) analysis was conducted with PRIMER7 software with the samples grouped based on the replicates.

### Blank sample collection and analysis

Five filter blank samples were collected and analysed. Filter blank samples were collected by attaching a clean, unused filter onto the air sampler at the sampling location and collecting them after 1 min without running the sampler. They were subjected to the same extraction methods and metagenomic analysis pipeline as the actual air samples.

### Reporting summary

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

### DATA AVAILABILITY

All raw unprocessed reads have been submitted to NCBI under the bio-project accession number PRJNA638794.

Received: 6 November 2020; Accepted: 19 March 2021;

Published online: 16 April 2021

### REFERENCES

1. Darwin, C. *The Voyage of the Beagle* (Cosimo Inc., 2008).
2. Von Humboldt, A. & Aimé B. *Personal Narrative of Travels to the Equinoctial Regions of America: During the Years 1799-1804* (Cosimo Inc., 2013).
3. Gilbert, J. A., Jansson, J. K. & Knight, R. The Earth Microbiome project: successes and aspirations. *BMC Biol.* **12**, 1–4 (2014).
4. Silvia, C. M. & Stal, J. L. *The Marine Microbiome* (Springer International, 2016).
5. Burrows, S. M., Elbert, W. & Lawrence, M. G. Bacteria in the global atmosphere. *Atmos. Chem. Phys.* **9**, 9263–9280 (2009).
6. Bauer, H. et al. The contribution of bacteria and fungal spores to the organic carbon content of cloud water, precipitation and aerosols. *Atmos. Res.* **64**, 109–119 (2002).

7. Prussin, A. J., Garcia, E. B. & Marr, L. C. Total concentrations of virus and bacteria in indoor and outdoor air. *Environ. Sci. Technol. Lett.* **2**, 84–88 (2015).
8. Jones, A. M. & Harrison, R. M. The effects of meteorological factors on atmospheric bioaerosol concentrations—a review. *Sci. Total Environ.* **326**, 151–180 (2004).
9. Schulz-Bohm, K., Martín-Sánchez, L. & Garbeva, P. Microbial volatiles: small molecules with an important role in intra- and inter-kingdom interactions. *Front Microbiol* **8**, 1–10 (2017).
10. Misztal, P. K. et al. Emission factors of microbial volatile organic compounds from environmental bacteria and fungi. *Environ. Sci. Technol.* **52**, 8272–8282 (2018).
11. Bourdillon, B. Y. R. B., Lidwell, M. & Thomas, J. C. A slit sampler for collecting and counting air-borne bacteria. *Epidemiol. Infect.* **41**, 197–224 (1941).
12. Palmgren, U., Ström, G., Blomquist, G. & Malmberg, P. Collection of airborne micro-organisms on Nuclepore filters, estimation and analysis-CAMNEA method. *J. Appl. Bacteriol.* **61**, 401–406 (1986).
13. Yamamoto, N. et al. Particle-size distributions and seasonal diversity of allergenic and pathogenic fungi in outdoor air. *ISME J.* **6**, 1801–1811 (2012).
14. Lang-Yona, N. et al. Annual distribution of allergenic fungal spores in atmospheric particulate matter in the eastern mediterranean; A comparative study between ergosterol and quantitative PCR analysis. *Atmos. Chem. Phys.* **12**, 2681–2690 (2012).
15. Hospodsky, D. et al. Human occupancy as a source of indoor airborne bacteria. *PLoS ONE* **7**, e34867 (2012).
16. Fu, X. et al. Indoor microbiome, environmental characteristics and asthma among junior high school students in Johor Bahru, Malaysia. *Environ. Int.* **138**, 105664 (2020).
17. Luhung, I. et al. Exploring temporal patterns of bacterial and fungal DNA accumulation on a ventilation system filter for a Singapore university library. *PLoS ONE* **13**, e0200820 (2018).
18. Amend, A. S., Seifert, K. A., Samson, R. & Bruns, T. D. Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proc. Natl Acad. Sci.* **107**, 13748–13753 (2010).
19. Tringe, S. G. et al. The airborne metagenome in an indoor urban environment. *PLoS ONE* **3**, e1862 (2008).
20. Yoosheph, S. et al. A metagenomic framework for the study of airborne microbial communities. *PLoS ONE* **8**, e81862 (2013).
21. Cao, C. et al. Inhalable microorganisms in Beijing's PM<sub>2.5</sub> and PM<sub>10</sub> pollutants during a severe smog event. *Environ. Sci. Technol.* **48**, 1499–1507 (2014).
22. Gusareva, E. S. et al. Microbial communities in the tropical air ecosystem follow a precise diel cycle. *Proc. Natl Acad. Sci.* **116**, 23299–23308 (2019).
23. Ottesen, E. A. et al. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* **345**, 207–212 (2014).
24. Kai, W. et al. Ambient bioaerosol particle dynamics observed during haze and sunny days in Beijing. *Sci. Total Environ.* **550**, 751–759 (2016).
25. Dybwad, M., Skogan, G. & Blatny, J. M. Comparative testing and evaluation of nine different air samplers: end-to-end sampling efficiencies as specific performance measurements for bioaerosol applications. *Aerosol Sci. Technol.* **48**, 282–295 (2014).
26. Luhung, I. et al. Protocol improvements for low concentration DNA-based bioaerosol sampling and analysis. *PLoS ONE* **10**, e0141158 (2015).
27. Spring, A. M. et al. A method for collecting atmospheric microbial samples from set altitudes for use with next-generation sequencing techniques to characterize communities. *Air Soil Water Res.* <https://doi.org/10.1177/1178622118788871> (2018).
28. Jiang, W. et al. Optimized DNA extraction and metagenomic sequencing of airborne microbial communities. *Nat. Protoc.* **10**, 768–779 (2015).
29. Kim, H., Park, K. & Lee, M. Biocompatible dispersion methods for carbon black. *Toxicol. Res.* **28**, 209–216 (2012).
30. Muthukumar, S. et al. The optimisation of ultrasonic cleaning procedures for dairy fouled ultrafiltration membranes. *Ultrasonic Sonochem.* **12**, 29–35 (2005).
31. Cragg, M. S. et al. Complement-mediated lysis by anti-CD20 mAb correlates with segregation into lipid rafts. *Blood* **101**, 1045–1052 (2003).
32. Núñez, A. et al. Monitoring of airborne biological particles in outdoor atmosphere Part 2: metagenomics applied to urban environments. *Int. Microbiol.* **19**, 69–80 (2016).
33. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
34. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
35. Abarenkov, K. et al. The UNITE database for molecular identification of fungi—recent updates and future perspectives. *N. Phytol.* **186**, 281–285 (2010).
36. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
37. Urlich, T. et al. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS ONE* **3**, e2527 (2008).
38. Dommergue, A. et al. Methods to investigate the global atmospheric microbiome. *Front. Microbiol.* **10**, 243 (2019).
39. Spens, J. et al. Comparison of capture and storage methods for aqueous microbial eDNA using an optimized extraction protocol: advantage of enclosed filter. *Methods Ecol. Evolution* **8**, 635–645 (2017).
40. Patterson, E. I. et al. Methods of inactivation of SARS-CoV-2 for downstream biological assays. *J. Infect. Dis.* **222**, 1462–1467 (2020).
41. Liu, C. M. et al. BactQuant: an enhanced broad-coverage bacterial quantitative real-time PCR assay. *BMC Microbiol.* **12**, 56 (2012).
42. Liu, C. M. et al. FungiQuant: a broad-coverage fungal quantitative real-time PCR assay. *BMC Microbiol.* **12**, 1 (2012).
43. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10 (2011).
44. Zhao, Y., Tang, H. & Ye, Y. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* **28**, 125–126 (2012).
45. Bokulich, N. A. & Mills, D. A. Improved selection of Internal Transcribed Spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Appl Environ. Microbiol.* **79**, 2519–2526 (2013).
46. Takahashi, S., Tomita, J., Nishioka, K., Hisada, T. & Nishijima, M. Development of a prokaryotic universal primer for simultaneous analysis of bacteria and archaea using next-generation sequencing. *PLoS ONE* **9**, e105592 (2014).

## ACKNOWLEDGEMENTS

The work was supported by Singapore Ministry of Education Academic Research Fund Tier 3 grant (grant MOE2013-T3-1-013).

## AUTHOR CONTRIBUTIONS

I.L., A.U., S.C.S.: Designed the study, conducted experiments, analysed the data, wrote the manuscript. S.B.Y.L.: Conducted experiments, analysed the data. D.I.D.-M., S.L.: Analysed the data, wrote the manuscript. N.E.G., C.K., K.L.: Conducted experiments. B. N.V.P., R.W.P., E.G., E.A., C.E.H., A.W., H.L.K., A.C.M.J.: Analysed the data. S.R.L., Y.Z.H. D. P., K.Y., K.K.K., A.P.N., A.P.: Conducted experiments. I.L. and A.U. contributed equally to this manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41522-021-00209-4>.

**Correspondence** and requests for materials should be addressed to S.C.S.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021