General Response to Referee #1

The presented work showcases the capabilities of a modified version of a PLAIR UV-LIF to distinguish certain bacteria from background noise as well as one bacterial species from the others. These kinds of investigations are certainly relevant for aerobiology as optical methods represent a new frontier of aerobiological sampling, and these kinds of studies are generally limited to pollen taxa.

We thank the reviewer for their positive assessment of our work and for highlighting its relevance within the field of aerobiology. We agree that optical methods remain largely focused on pollen taxa, and one of the motivations of this study was precisely to explore the potential of UV-LIF instruments for bacterial discrimination. We appreciate the reviewer's constructive comments and suggestions, which have helped us to improve the clarity and robustness of the manuscript. Below we address each of the points raised to the best of our knowledge.

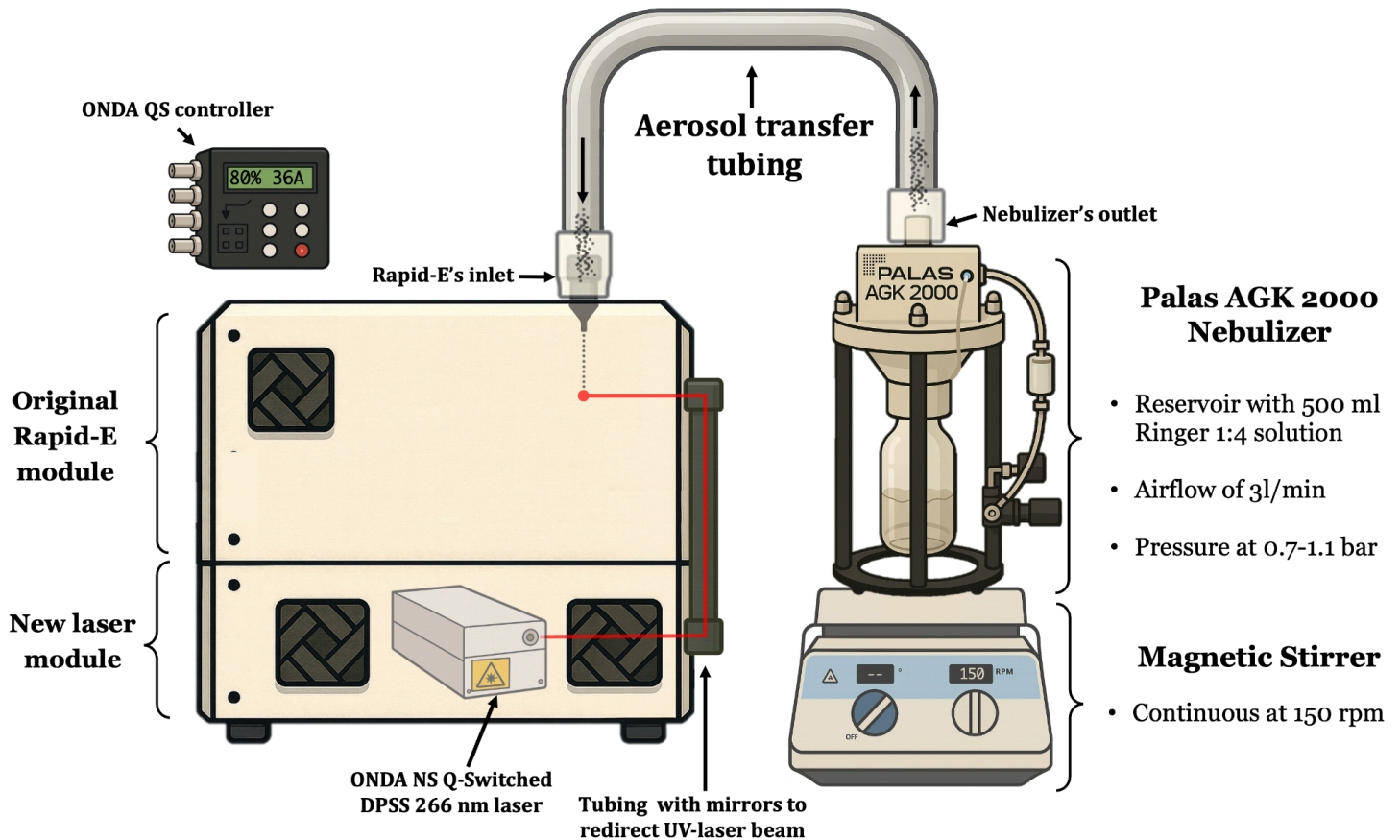While the work is interesting, I would have some comments for the authors:

- Can you please (even as supplementary material) give more information about the modifications made to the PLAIR? I think it would be interesting for other researchers as well to understand more in detail how such a modification can be made. Also, was this modification purely driven by optical considerations or from a previous experience where the instrument "as-is" failed to detect bacteria? A comparison of non-modified and modified PLAIR would also be interesting to understand the degree of improvement in detection.

We have added a section in the supplementary material where we describe in more detail the modifications performed on the Rapid-E device to integrate the new laser. In addition, we generated a schematic diagram that illustrates these changes and the overall aerosolization set-up in greater detail. We have now replaced Figure 2 in the manuscript, which was a simple picture of the updated device, with this to provide a more comprehensive idea of the modifications and the aerosolization process. The figure shows the process by which samples are aerosolized using the Palas AGK 2000 nebulizer, transferred through aerosol tubing, and introduced into the Rapid-E's inlet.

The diagram also highlights the integration of the ONDA NS 266 nm laser module, which was soldered to the instrument with an external controller allowing for intensity modulation.

A high-resolution version of the figure is also available in the project's GitHub repository, since we cannot attach images over 500x500px here:

https://github.com/AlFontal/lif-bacteria-aerosols-ms/blob/main/output/figures/combi_ms/fig_2_aerosolization_uv_integration_diagram.png

ONDA QS controller

80% 36A

Aerosol transfer tubing

Rapid-E's inlet

Nebulizer's outlet

PALAS AGK 2000

Original Rapid-E module

New laser module

Palas AGK 2000 Nebulizer

- Reservoir with 500 ml Ringer 1:4 solution
- Airflow of 3l/min
- Pressure at 0.7-1.1 bar

ONDA NS Q-Switched DPSS 266 nm laser

Tubing with mirrors to redirect UV-laser beam

Magnetic Stirrer

- Continuous at 150 rpm

150 RPM

Regarding the motivation for this modification, the version of the Rapid-E originally available to us was equipped with a 335 nm UV laser, which is mainly optimized for chlorophyll excitation (which made sense because indeed, this was a device thought to identify and classify pollen).

Based on discussions with the manufacturer and on previous literature, the consensus at the time was that this wavelength would not provide sufficient sensitivity for detecting bacterial fluorophores. As discussed in the text, and in particular, key biomolecules such as tryptophan, NADH, and FAD show stronger excitation in the deep-UV range around 260–280 nm. For this reason, and to maximize discrimination capability, we opted to integrate the 266 nm DPSS laser, which is known to better capture the intrinsic fluorescence signatures of bacteria (Pan et al., 1999; Sivaprakasam et al., 2004; Hill et al., 2015).

Unfortunately, we did not run a comparable test with the original device "as is" previous to the decision, and since the modification process was long and iterative, we did not perform a direct comparison between the original configuration and the modified version. We agree with the reviewer that such a side-by-side test would be very interesting to quantify the exact gain in classification power provided by the actual modifications. However, this evaluation fell outside the scope of the present work, where our aim was instead to demonstrate that the modified instrument could indeed classify bacterial aerosols with sufficient discriminatory power.

- Paragraph 2.2.1. Can you please provide a figure of your aerosolization set-up?

See the answer to the previous query and the new Figure 2 on the manuscript.

- Paragraph 2.2.3. Can you please provide details about bacterial growth (media, time, temperature of growth, …) as well as the technique used for identification of the species (how was MS-TOF used in this context?).

Filter portions were placed in contact with nutrient agar plates, then removed, and the plates were incubated at 37 °C for 24 h. Morphologically distinct colonies were subsequently subcultured under the same conditions to obtain pure isolates. These isolates were identified using MALDI-TOF MS (LT MicroFLex, Bruker Daltonics, Germany). For each isolate, a small fraction of biomass was spotted onto the target plate, after which 1 µL of a saturated HCCA matrix solution was added and allowed to dry. Each sample was spotted in duplicate, and each spot was measured twice, yielding four mass spectra per isolate. The resulting spectra were compared with the Bruker bacterial library v.9.0.

- Paragraph 2.3.1. At line 158 is stated that "No transformation was needed for incorporating fluorescence spectra and lifetime data into the models". I would like to better understand this point since, as far as I understand, per each particle (i.e.: sample) your features were all different in size. The fluorescence spectrum should be a [1,32] vector (intensity vs. 32 wavelengths); the lifetime a [4,64] matrix (4 bands vs. 64 nanoseconds) and the scattering a [24,60] matrix (24 angles vs. 60 microseconds following your cropping). How was this difference handled to generate a consistent input for the random forest classifier?

We appreciate the comment and understand how it might lead to an ambiguous understanding in the way we originally phrased it. What we meant is that both fluorescence spectra and fluorescence lifetime outputs are consistently acquired by the instrument as fixed-size arrays, so they could be directly incorporated into the models without further preprocessing, unlike the scattering data. Specifically, the fluorescence spectra are recorded as 32-channel vectors at 8 different timepoints (so, actually, a [8,32] matrix) and the lifetime outputs as [4,64] matrices (Figure 1 should clarify this). These matrices are then flattened into one-dimensional vectors [256] and [256] before being passed to the classifier.

In contrast, scattering signals varied in duration between particles, which is why we applied the cropping, zero-padding, and normalization procedure described in the text to ensure consistent input dimensionality.

Random Forests (and basically all tree-based methods) don't rely on distance metrics so they are not sensitive to the scale of the inputs, which in our case is advantageous since this eases the ease of incorporating heterogeneous inputs as ours. The structural nature of data, however, gets lost, so each feature is understood as independent from each other by the RF, and it needs to "learn" it if relevant. This doesn't seem to be a problem for the fluorescence spectra/lifetimes, but might explain why we see little predictive power gain from the scattering images.

To clarify this point, we have rephrased the relevant sentence in the manuscript as follows:

"No additional transformation is required for incorporating fluorescence spectra and lifetime data into the models, as these are acquired by the instrument in fixed dimensions (32-channel spectra x 8 acquisitions and 4 lifetime ranges x 64-channels, all later flattened into vectors). However, light scattering images present a challenge due to their irregular

shapes, as the total number of acquisitions depends on the duration of the detected scattered light signal"

- Figure 4. From the figure and from the text it is unclear to me how the samples were labelled. Were simply all the spectra from the aerosolization of a given bacteria considered as pure bacterial spectra with no filtering (except for the time cropping and the fluorescence threshold)? So, essentially, all spectra were "auto-labelled" depending on their source assuming no interference? Also, how was the multiclass classifier trained?

Indeed, this is one of the main challenges when generating training data from bioaerosols: we *know* that some of the particles produced will contain the species of interest, but we cannot guarantee that every droplet entering the device will do so (or even that they will always carry any biological material at all). This makes the labelling process closer to a pseudo-labelling approach rather than generating a real ground truth. With this in mind, however, we have done our best to attempt to minimize the potential labelling errors in our process:

First, we apply a strict fluorescence threshold, using a higher cutoff than in previous studies (e.g., Šaulienė et al. 2019), which also relied on thresholds to exclude unwanted particles such as pollen. Our 2000 a.u. cutoff lies within the 95th to 99th percentile range across all sample groups, so the majority of "empty" particles are discarded at this step.

Second, we observed that the control group also contained particles with fluorescence signals. This means that fluorescence alone cannot serve as the only variable to distinguish bacteria-containing particles from those in bacterial-free aerosols. For this reason, we first train the binary classifier to separate bacteria-containing particles from all others (using the subset already passing the fluorescence threshold), and then train the multiclass classifier exclusively on the bacterial-enriched aerosols.

That said, we acknowledge that many empty particles likely remain among those labelled as bacterial, and conversely, that many valid bacterial particles may have been discarded because they did not reach the fluorescence threshold. The fraction of non-bacterial particles incorrectly labelled as bacterial is likely limited, since the models show strong discrimination performance at the binary level on completely unseen data. Moreover, even simple two-dimensional PCA projections of the predictors already reveal clear separability between groups, which would be unlikely if both contained largely overlapping "empty" signals.

- Figure 6C, what's in the y axis of the plots? Each line is one particle (so it would be "Particle # [1-100]")?

Thanks for the remark, as indeed the current plot was rather ambiguous without labels or ticks in the y-axis. The y-axis here basically represents the index of each particle, and as the selected particles were the top 100 with the highest peak fluorescence at any acquisition time, it is an indicator of the top *n* particles with respect to that metric for each of the groups. For clarity, we have updated the panel C figure to include the label and the indexes 1, 50 and 100 as ticks. We have also opted to reverse the order of the y-axis, since the previous sorted 100 to 1 in a top-to-bottom order, which is rather confusing especially taking into account that now we explicitly label the indexes. For further clarification, I attach with the

comment a high-resolution PNG and a SVG of the figure, and the construction of the data behind the figure and commands can be openly accessed here:

https://alfontal.github.io/lif-bacteria-aerosols-ms/fluorophores_ms
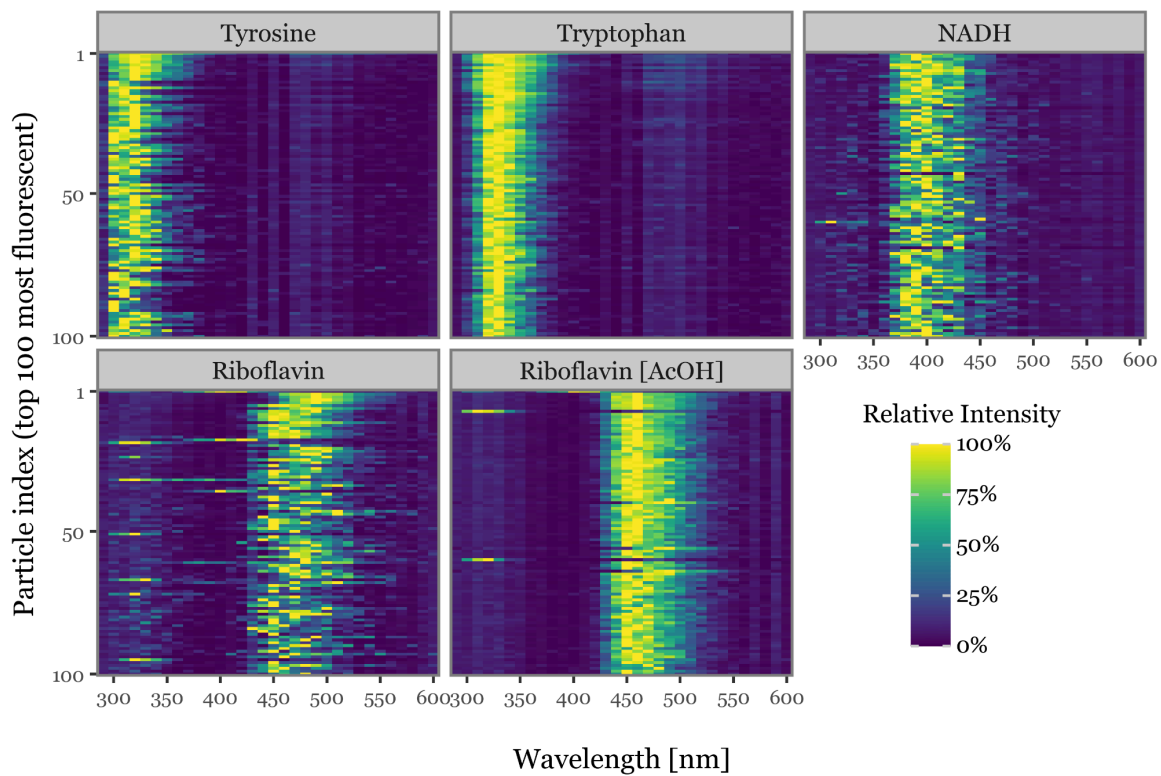
General Response to Referee #2

Response 0: General comments

The authors present the results of laboratory testing of a modified Rapid-E bioaerosol detector and a machine learning identification method, to assess their suitability for identification of airborne bacteria in bioaerosols. The presented work provides a significant step towards this goal and encouraging results are shown. However, significant open questions remain regarding the applicability of this approach for the stated goal of identification under environmental conditions. These challenges should be emphasised more clearly along with further discussion and more concrete suggestions as to how the gap may be bridged between the laboratory study presented here and the intended goal of environmental deployment and application.

The paper is mostly clear and well written, with the exception of some aspects that are listed below. The work is important because there does not appear to be an existing method to automatically detect and classify bacterial particles in bioaerosols in indoor and outdoor conditions. I therefore recommend that this work be published, subject to the specific comments/questions listed as follows being suitably addressed by the authors:

We would like to thank the reviewer for their careful reading of our manuscript and for the constructive feedback provided. We have carefully revised the text throughout to address all comments in detail, expanding the discussion on the complexities of real-world applications, particularly the implications of heterogeneous aerosol mixtures on classification and signal interpretation. The introduction now places the study in clearer context, emphasizing that our prototype represents an early proof-of-concept towards real-time microbial bioaerosol detection rather than a ready-to-deploy system.

In response to the reviewer's suggestions, we have also clarified several methodological aspects, including the nature and treatment of control samples, the potential influence of background fluorescence, and our approach to contamination and signal filtering. Additional text was included to discuss the limitations of the current setup and the steps required to translate this work into field-ready applications.

We believe that these changes have substantially improved the clarity, scope, and interpretive depth of the manuscript.

Please see the following author comments as individual responses, grouped thematically according to the reviewer's remarks. We hope these successfully address all points raised.

Response 1: Title change suggestion

Line 1:

The word 'benchmarking' suggests a comparison being made against a standard approach. What comparison is made here ie. given there is no standard approach? Suggest to revise title and eg. emphasise the scope of laboratory testing undertaken here.

Fine, thank you. We accept this as a valid remark since indeed literally, the term *benchmarking* implies that a comparison against an established standard (which simply does not exist for an emerging field such as real-time bacterial bioaerosol detection) was made. However, our use of the 'benchmark' mention was made in a more conceptual/figurational way to highlight that despite some similar methodologies have been developed, not any one has managed until now to objectively present proof-of-concept of near-real time differentiation of microbial taxa in bioaerosols. Acknowledging this, we have revised the title to reflect more accurately the scope of the study as a laboratory demonstration and prototyping of a potential new technology use. The title now reads:

"Laser-Induced Fluorescence coupled to Machine Learning as an effective approach for real-time identification of bacteria in bioaerosols."

Response 2: Prior work references

Line 26:

Provide reference(s) for this prior work.

We have now included references that cover representative studies and reviews on soil, aquatic, and surface microbiomes, as well as one addressing the current state of knowledge and remaining challenges in our understanding of the aerobiome. Hopefully, these provide enough context for how microbial communities in other environments are already well established in the literature (or at the very least, to a higher depth) compared to atmospheric microbiome research.

Lines 26-34 now read:

"While extensive research has been conducted on the microbiome present in soil (Banerjee and van der Heijden, 2023; Fierer, 2017), water (Cole, 1999), and different surfaces in both urban and rural environments (Gilbert and Stephens, 2018), the atmospheric microbial ecosystem (aerobiome) has remained relatively unexplored until recently (Amato et al., 2023; Tastassa et al., 2024). The atmosphere is a dynamic environment where microorganisms can be transported over long distances (Rodó et al., 2024), influencing both local and global ecological patterns. These airborne microorganisms play crucial roles in nutrient cycling, weather patterns, and even disease transmission (Griffin, 2007; Morris et al., 2011; Fröhlich-Nowoisky et al., 2016; Tellier et al., 2019). Despite their importance, the study of aerobiomes has been hampered by significant technical challenges (Behzad et al., 2015; Luhung et al., 2021a)"

References added:

Cole, J. J. (1999). Aquatic microbiology for ecosystem scientists: new and recycled paradigms in ecological microbiology. **Ecosystems**, 2(3), 215-225.

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. **Nature Reviews Microbiology**, 15(10), 579-590.

Banerjee, S., & Van Der Heijden, M. G. (2023). Soil microbiomes and one health. **Nature Reviews Microbiology**, 21(1), 6-20.

Gilbert, J. A., & Stephens, B. (2018). Microbiology of the built environment. **Nature Reviews Microbiology**, 16(11), 661-670.

Amato, P., Mathonat, F., Nuñez Lopez, L., Péguilhan, R., Bourhane, Z., Rossi, F., ... & Ervens, B. (2023). The aeromicrobiome: the selective and dynamic outer-layer of the Earth's microbiome. **Frontiers in microbiology**, 14, 1186847.

Tastassa, A. C., Sharaby, Y., & Lang-Yona, N. (2024). Aeromicrobiology: A global review of the cycling and relationships of bioaerosols with the atmosphere. **Science of the Total Environment, 912**, 168478.

Luhung, I., Uchida, A., Lim, S.B.Y. et al. Experimental parameters defining ultra-low biomass bioaerosol analysis. **npj Biofilms Microbiomes 7**, 37 (2021).

Luhung, I., Uchida, A., Lim, S.B.Y. et al. Experimental parameters defining ultra-low biomass bioaerosol analysis. **npj Biofilms Microbiomes 7**, 37 (2021).

Response 3: Challenges of real-word application.

Line 57-59:

More discussion should be provided here on the complexities of real world application eg. implications of heterogeneous mixtures on classification attempts, etc.

See below a joint reply and details of the edits made to this text.

See a joint reply below.

Line 303-304:

It is encouraging to see some separation in groups. But it is important to note the challenges in real air samples where mixtures of different bacteria and aerosols could be found in the same sample. A discussion of the implications and challenges of extending these results to mixtures is then important in assessing the usefulness of this work eg. what assumptions/conditions might be required to separate them in a statistically significant way?

See a joint reply below.

Line 387-390:

This point needs to be emphasised more prominently throughout and elaborated on/discussed in greater detail here.

Line 452-453:

As with the previous comment, this point should be elaborated on and concrete suggestions provided for how to realise such real world applications based on the status of this current work.

Agreed, this is a very good point and the main challenge to overcome if this prototype were to finally be operationalized. In natural conditions, microorganisms rarely appear isolated as single particles in the way our aerosolization process generates: they usually come attached to or embedded within other particles, in complex mixtures and diverse matrices.

Because the instrument interrogates one particle (or aggregate) at a time and returns a single feature vector, assigning more than one class to a single event is not realistic. Training for all possible mixtures found in ambient air is also infeasible, since the combination space is too large. There are two challenges here:

First, biological co-aggregates mix taxa (for example, two bacterial groups or fungi with bacteria). In such cases a single event may contain multiple true classes, which we cannot resolve with one excitation and one readout. Given the purpose of the device, we should prioritize precision for bacterial detections over exhaustive recall and allow events to remain unclassified or flagged as "bacterial-like" without finer typing, since in this case false negatives (unidentified taxa) are much more tolerable than false positives (wrongly identified taxa). This behaviour is acceptable as long as a stable fraction of incoming events is relatively pure and can be classified reliably.

Second, many particles carry inorganic or organic matrices that shift the signal through absorption, scattering, and quenching. This is a background-separation problem rather than a multi-label problem. A follow-up phase should aerosolise bacterial isolates together with realistic particulate matrices to test detection under lifelike conditions.

This later phase would also involve generating controlled mixtures and testing whether their composite signatures are consistent and learnable. We consider those exercises however to lie beyond the scope of this first prototype, which focuses on single-event labels under a single excitation band, and the feasibility of finding any discriminative signal in aerosolized samples in a pseudo-controlled scenario.

With all that being said, we have made several edits to the manuscript to incorporate and emphasise these challenges more clearly.

Text in lines 56-60 of the introduction now elaborate more on the issue and read as:

"UV-LIF can achieve fine-scale discrimination in the lab, even for small viral particles when paired with machine-learning classifiers (Gabbarini et al., 2019). Translating this capability to real-time, field-deployable systems is challenging because microbial particles in ambient air rarely occur as isolated single particles: they often appear as biological co-aggregates or are embedded within heterogeneous inorganic and organic matrices, both of which complicate both their detection and classification."

We've also added 10 lines to the discussion, with the last paragraph now reading (lines 441-461):

"While the two-step model demonstrated a substantial degree of generalization on our validation set, this capability remains inherently linked to the specific conditions of the training data: artificially generated bioaerosols. Although aerosolizing bacterial samples in Ringer solution 1:4 represents a step towards greater complexity than pure cultures in idealized media, it still falls short of replicating the full heterogeneity of real environmental aerosols. Natural aerosols comprise a dynamic and diverse array of biological (e.g., pollen, fungal spores, plant debris) and non-biological particles (e.g., dust, soot, pollutants), all capable of scattering light and emitting fluorescence to varying degrees (Hill et al., 2001; Després et al., 2012; Pöhlker et al., 2012). In such environments, the spectral, lifetime, and morphological signatures of target bacterial particles may be obscured, mimicked, or altered by other components, leading to potential misclassification or reduced detection sensitivity. Addressing this will require the development of extensive training datasets that more accurately reflect this real-world variability, a substantial although highly-promising task given the dynamic nature and sheer diversity of environmental aerosols across different locations and times. Two sources of heterogeneity are especially relevant for classification: biological co-aggregates that mix microbial groups within a single event, and inorganic or organic matrices that shift signals through absorption, scattering, and quenching (Savage et al., 2017; Calvo et al., 2018). To avoid this, in the short term, we suggest an approach with conservative thresholds, permitting unclassified or "bacterial-like" classes when mixtures or strong backgrounds are suspected or model uncertainty is high. Finally, taxonomy is hierarchical, a property that our current model omits, but that means that targets can be structured across ranks with a back-off rule: assign species or genus only when confidence is high, and default to higher ranks when it is not. This is standard practice in sequence-based classifiers such as the RDP naïve Bayesian classifier and MEGAN's lowest-common-ancestor approach, which label to the highest taxonomic rank supported by the data in metagenomics studies (Huson et al., 2007; Wang et al., 2007). The same logic could be adapted here if the class set and evaluation follow the rank

structure. Any real usage in the field will need validation with co-located reference sampling to establish site-specific baselines and thresholds."

And finally, we've also added a couple of lines in the conclusion in line with the comments of the reviewer (lines 468-471):

"Next research steps should focus on testing realistic particulate matrices that resemble ambient bioaerosols, and on mixtures of bacterial (and fungal) isolates to assess whether composite signatures are consistent and learnable. Meeting these milestones would bridge the technology shown in this study from prototype to robust field deployable tool. "

New references added:

Savage, N. J., Krentz, C. E., Collins, D. R., & Huffman, J. A. (2017). Systematic characterization of WIBS-4A using fluorescence standards and primary biological aerosol particles. **Atmospheric Measurement Techniques**, 10, 4279–4302. https://doi.org/10.5194/amt-10-4279-2017

Calvo, A. I., Baumgardner, D., Castro, A., Fernández-González, D., Vega-Maray, A. M., Valencia-Barrera, R. M., Oduber, F., Blanco-Alegre, C., & Fraile, R. (2018). Daily behavior of urban Fluorescing Aerosol Particles in northwest Spain. **Atmospheric Environment**, 184, 262–277. https://doi.org/10.1016/j.atmosenv.2018.04.027

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. **Genome Research**, 17(3), 377–386. https://doi.org/10.1101/gr.5969107

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. **Applied and Environmental Microbiology**, 73(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07

Response 4: Particle-aggregation clarification.

Line 74-75:

It seems important to this technique that fluorescence from only one particle is measured at a time. How often is this the case in practice eg. how often is fluorescence from multiple particles measured at the same time as a function of particle number density, size, and flow rate?

It is true that, for both training and inference, aggregates are a complication. As noted elsewhere in this response (see responses 3 and 12) and in the manuscript, natural bioaerosols often include biological co-aggregates and particles embedded in heterogeneous inorganic or organic matrices, which can shift or mix signals. This is a major challenge for field deployment, and we already reason that a practical system should favor precision and allow abstention in its classification outputs, understanding that a sizable amount of the particles entering the system will simply not be classifiable.

With that said, in our setup, the instrument scans one particle at a time in a laminar flow: a red laser first triggers the event via multi-angle scattering, then the UV excitation follows for fluorescence collection. During all experiments the connection to the aerosol generator was sealed and the flow into Rapid-E was held at about 3 L/min, which shortens residence time and should help limit collisions at the inlet. We did not directly quantify coincidence rates as a function of number density, or size, but we were operating near the device's counter limit of about 5,000 particles per minute during all training except for the cleaning steps.
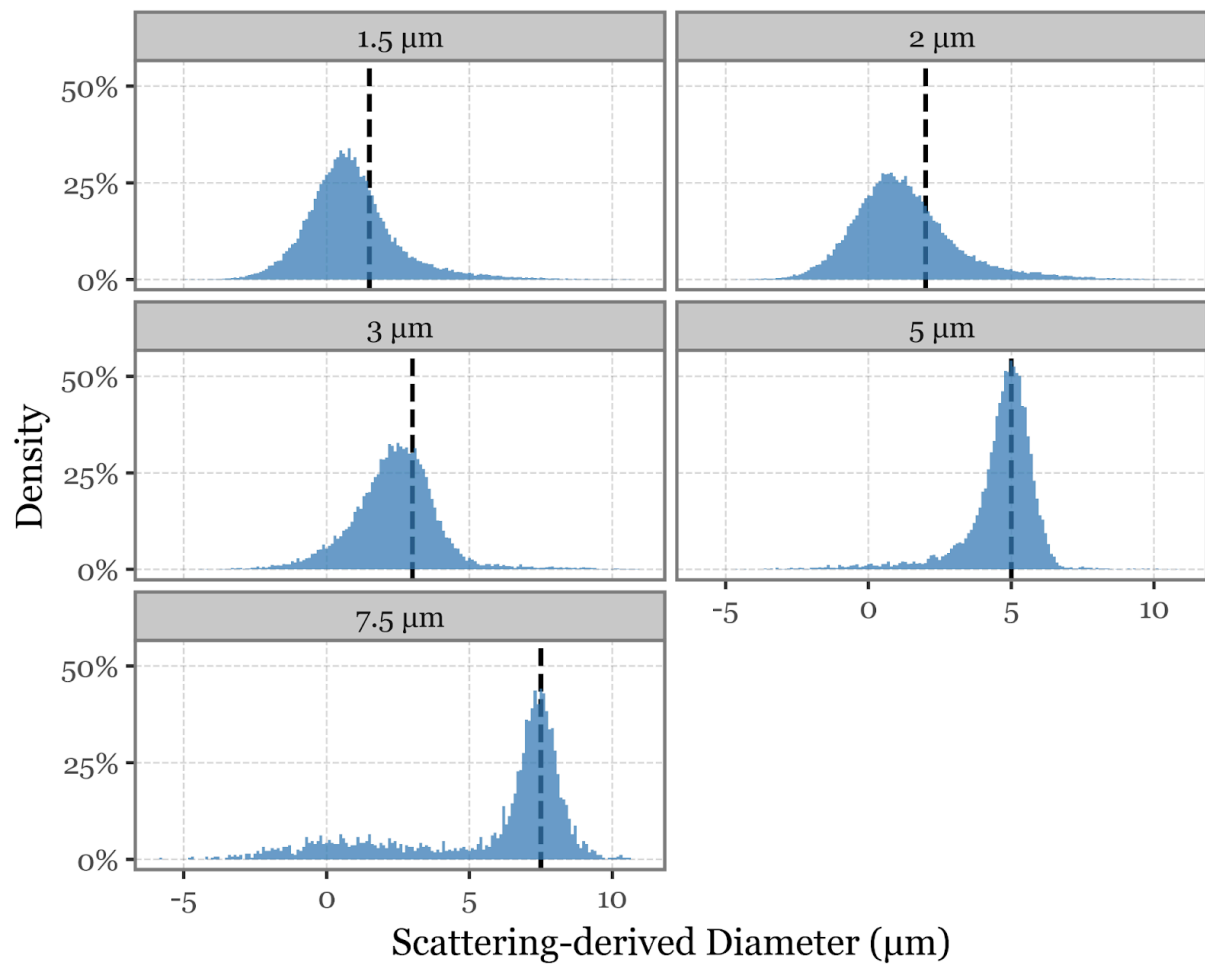
We did however run a small test of size resolution as a first check of the new laser setup with aerosolized monodisperse polystyrene microspheres between 1.5 µm and 7.5 µm (roughly in the same range as bacteria) under the same conditions as the bacteria and fluorophores, which might give us some data to quantify the degree of aggregation we might be achieving in the aerosolization process. The device reports a diameter derived from an internal calibration tied to Mie-type scattering.

In the attached figure we provide the estimates of diameters returned by the Rapid-E device, which tracked the expected modal sizes and remained stable up to roughly the 7–8 µm range in our tests. At very low scattering intensities the vendor's size proxy becomes less reliable, so we treat those tails with caution rather than as true negative or sub-physical diameters. Importantly, we did not observe distinct secondary modes that would suggest frequent large aggregates: instead, any right tails over the expected diameter appear limited. For the bacterial samples themselves, we already show in the manuscript how most particles fall within the 4–5 µm range (Table 2, Supp. Fig. 2B), with longer right tails. This pattern could reflect occasional aggregation, non-sphericity, or limits of the size proxy for irregular particles, but in no case they seem to suggest a significant number of aggregated particles.

We have added the attached figure as Supplementary Figure 8 with a short explanation of the microspheres tested that reads:

"**Suppl. Figure 8. Size distributions of monodisperse particles.** Scattering-derived diameters reported by the Rapid-E for aerosolized monodisperse polystyrene microspheres with nominal sizes of 1.5, 2, 3, 5 and 7.5 µm, generated under the same flow and aerosolization conditions as the bacterial experiments. The vertical dashed lines indicate the nominal diameters. Distributions are centered close to the expected values, with only limited right tails and no clear secondary modes, suggesting that the instrument's Mie-based size proxy is stable over

Response 5: Diagram modifications.

Line 95:

Suggest using a unique symbol in the diagram for each modification and mapping to description in the text eg. could number them on the diagram as in the text currently, to identify more clearly each modification from the diagram

We have now added a numbered red dagger (†) besides each of the 3 modifications in Figure 3. See attachment. We also adapted the figure caption to mention it:

"Figure 1. Schematic representation of Rapid-E's operational mechanism, illustrating the data points generated for each particle as it passes through the two integrated laser systems. The red daggers (†) denote modifications made with respect to the commercial version of Rapid-E: (1) the UV laser's excitation wavelength changed from 337 nm to 266 nm, (2) the fluorescence spectra acquisition range was changed from 350-800 nm to 290-600 nm, and (3) the first lifetime channel was shifted from 350-400 nm to 300-340 nm."
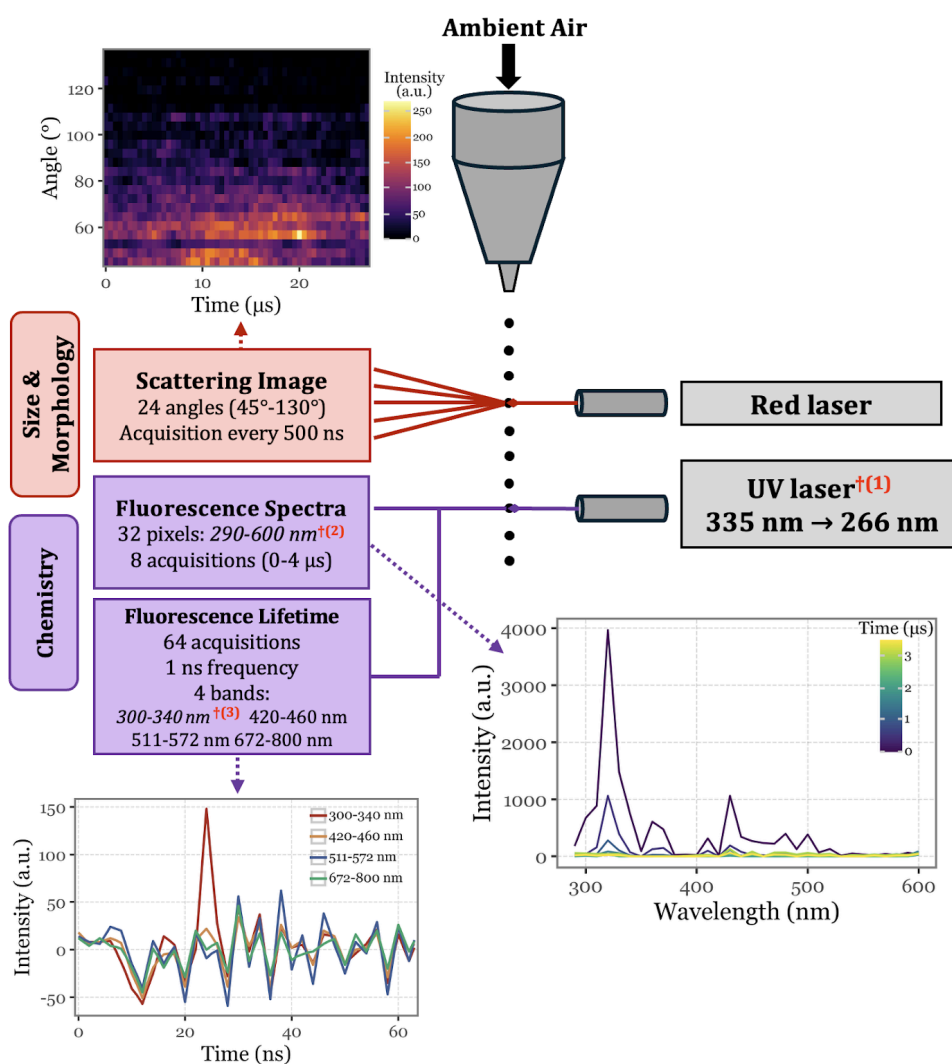
Response 6: 266 nm wavelength choice.

Line 89-91:

The presented approach critically depends on this claim that emission spectra for bacterial fluorophores are distinct in this wavelength range. More discussion/justification for this claim is therefore needed here. Is it truly distinct or do any other atmospheric particulates fluoresce similarly/in overlapping ways in this wavelength range? Include further references showing this property, or state if it is just an assumption.

Line 102:

Why was this specific wavelength chosen? Is it optimal somehow? Are there other/better wavelengths that could be tested in future, or perhaps a combination of wavelengths could be useful? Some more discussion on this should be provided.

Additional discussion should be included on the sensitivity of this technique to background noise sources eg. what are the dominant noise sources under laboratory and environmental conditions, and does this limit the technique significantly or not? Do other particulates typically found in the environment fluoresce under 266nm light that would complicate identification under real environmental conditions?

Thanks for the comments. We will try to address all the queries concerning the choice of wavelength

We chose this wavelength with our main goal in mind, which is bacterial identification/discrimination, and given the current hardware limitation of a single excitation band. The key native emitters are the aromatic amino acids in proteins (tryptophan dominant), plus metabolic cofactors such as NADH and flavins. Aromatic residues absorb most strongly in the deep-UV, so excitation in the 260–280 nm range produces stronger and more distinctive protein-like signals in bacteria than longer UV. Modeling and single-particle studies support this, with 266 nm sitting in that optimal window (Hill et al., 2013; Hill, 2015; Sivaprakasam et al., 2004; Pan et al., 2015).

With regard to the potential use of other wavelengths or combinations: Dual or multi-excitation could improve class separation performance by adding independent biochemical contrast and enabling ratiometric features which are not possible in our current setup. That being said, under our current single-excitation constraint, 266 nm remains the best standalone choice because it maximizes bacterial signal strength from aromatic residues.

If we wanted to add excitation channels, we would likely go for either the 340-370nm or the 440-470nm range:

340–370 nm would preferentially excite NADH, which emits around 440–470 nm (Lakowicz, 2006; Monici, 2005), and 410–450 nm would target flavins such as FAD/FMN, which emit near 520–535 nm (Lakowicz, 2006; Pöhlker et al., 2012). Existing instruments support these choices, with 266/355 nm two-pulse and 280/370 nm WIBS-style setups showing strong performance in bioaerosol characterization (Sivaprakasam et al., 2004; Savage et al., 2017).

Following up on the previous question, but now focusing on environmental interference: Yes, some common non-biological particles can fluoresce under 266 nm excitation. Potential interferences

include organic "brown carbon" and humic-like substances from natural decay, soot that carries polycyclic aromatic hydrocarbons (PAHs) from traffic, secondary organic coatings on mineral dust, and plant debris or fragmented pollen. These materials contain aromatic chromophores that absorb in the deep-UV, similar range as the aromatic amino acids we mentioned before, and can emit in the blue–green (Pöhlker et al., 2012; Savage et al., 2017; Calvo et al., 2018). However, their emission spectra typically differ from our bacterial targets, their fluorescence lifetimes are different, and their scattering patterns also tend to differ. Lifetime is our most informative feature, which is why we lean on it heavily. Background will therefore matter and can reduce signal cleanliness, and field use will require site-specific baselines and conservative thresholds, but these interferents should not preclude practical separation of biological aerosols in real air.

References:

Hill, S. C., Pan, Y.-L., & Chang, R. K. (2013). Fluorescence of bioaerosols: mathematical model including primary fluorescing and absorbing molecules in bacteria. **Optics Express**, 21(19), 22285–22313.

Hill, S. C. (2015). Fluorescence of bioaerosols: model extensions and implications for excitation in the deep-UV. **Applied Optics**, 54(31), 9352–9368.

Sivaprakasam, V., Huston, A. L., Scotto, C., & Eversole, J. D. (2004). Multiple UV-wavelength excitation and fluorescence of bioaerosols. **Optics Express**, 12(19), 4457–4466.

Pan, Y.-L., Santarpia, J. L., Hill, S. C., et al. (2015). Single-particle fluorescence spectroscopy of bioaerosols with deep-UV excitation. **Journal of Quantitative Spectroscopy and Radiative Transfer**, 153, 144–162.

Lakowicz, J. R. (2006). **Principles of Fluorescence Spectroscopy** (3rd ed.). Springer.

Monici, M. (2005). Cell and tissue autofluorescence: research and diagnostic applications. **Journal of Photochemistry and Photobiology B: Biology,** 82(3), 141–154.

Pöhlker, C., Huffman, J. A., & Pöschl, U. (2012). Autofluorescence of atmospheric bioaerosols: spectral patterns and physical properties. **Atmospheric Measurement Techniques**, 5, 37–71.

Savage, N. J., Krentz, C. E., Collins, D. R., & Huffman, J. A. (2017). Systematic characterization of WIBS-4A using fluorescence standards and primary biological aerosol particles. **Atmospheric Measurement Techniques,** 10, 4279–4302.

Response 7: Cross-sample contamination

Line 128-130:

How effective was this method for reducing the impacts of cross sample contamination? Can measurements be included that confirm the validity of this approach?

Line 148-149:

As with the previous comment, measurements should be included to confirm the validity of this contamination handling approach.

Thank you for the very relevant point raised. As described in the cited lines of the manuscript, we used a 30-minute Milli-Q aerosol flush between samples and analyzed only the last 10 minutes of each run to provide an extra buffer.

To verify this in practice, we tracked minute-by-minute totals and the share of particles crossing the fluorescence threshold during each flush and at the start of every new sample. Immediately after a switch the counter was often near the instrument limit (~5,000 particles/min). During the Milli-Q flush both the total counts and the fluorescent share fell sharply to a stable background, and we only began the next sample several minutes after that stabilization (at least, 30). Only the buffered 10-minute window was used for modeling.

This does not demonstrate zero carryover, but the measurements indicate that any residual was likely negligible for training and validation, given the long flush, the fixed conditions, and the low fraction of fluorescent events used in the models. The clear separability observed between groups is consistent with minimal cross-contamination and suggests it did not materially affect discrimination.

For clarity, we have added in the subsection 2.2.3 of the Methods section, the following lines extending the current explanation:

"Each bacterial isolate was aerosolized for at least 15 minutes, and to prevent cross-contamination, Milli-Q water was aerosolized for 30 minutes between each sample. During method development we verified this procedure by tracking minute-by-minute particle counts and the fraction of events above the fluorescence threshold, and confirmed that both dropped to a low, stable background before starting the subsequent sample"

Response 8: Filtering & Sampling Biases?

Line 151-152:

Does this filtering introduce sampling biases? Would real samples meet this intensity requirement or could the training data be inadequate for real conditions with identification based on weaker signals eg. due to variability in scattering from distributions in particle shape/size/etc in an environmental sample? Some added discussion on this would be helpful.

Yes, the 2000 a.u. fluorescence cutoff introduces a sampling bias, and it is intentional. We prefer to train and evaluate on particles with strong signal-to-noise so that discrimination is reliable, rather than include weak events that would add label noise. The instrument's throughput is high (~5000 particles per minute), so we do not need to classify every event. In practice only about 1–4% of detected particles exceeded the threshold, yielding 7141 particles for modeling, which we consider an acceptable trade-off between sample size and label quality.

As noted in the Methods, our cutoff is higher than the 1,500 a.u. used by Šaulienė et al. 2019, for pollen because our laser and spectral conditions differ and we observed substantial background below 2000 a.u., so a stricter filter was appropriate for these bacterial experiments.

This choice means the current models are not tasked with classifying very weak events. For deployment, thresholds should be tuned to site-specific background and the system should favor precision with the option to abstain on weak or mixed signals, as we already discuss in the manuscript and replies about real-world use. In short, the filter creates a deliberate bias toward clean signals. It does not aim to represent all ambient particles, but it improves the validity of the training labels under the reported conditions.

We have added some text for clarity in the Methods, lines 159-161:

"In practice, only about 1-4 % of detected particles exceeded this threshold. This intentionally biases the dataset toward particles with high signal to noise so that label quality is prioritised over exhaustive sampling of all detected events."

Response 9: Clarification on Ringer Control

Figure 3:

I am confused by the shown results. I had the impression that your control used was a Ringer solution with no particulates added. Yet the graph here and subsequent discussion indicate similar particle scattering properties for the control and for the bacterial samples.

Does this indicate that there are bacteria/similarly fluorescent particles in the Ringer solution? Therefore, does this complicate interpretations of thus non independent samples? Or is the fraction of bacteria particulates in the Ringer solution somehow known to be low enough to not give ambiguous results and measurements are primarily from the aerosolised salts in the Ringer solution?

Or, am I misinterpreting this entirely?

In either case, a clearer explanation of the Control used would help.

Line 280-281:

What is the origin of these features, even in the Control? Is it an artefact of the Ringer solution used, and, if so, could more pure spectra be obtained with an alternative medium in future?

Thank you for this query as it helps clarify our results in the new adapted text.

The control is, indeed, Ringer 1:4 solution aerosolized on its own without any added particulates (bacterial or otherwise). The purpose of Figure 3 was to show that the 30 μs threshold that we used to pre-process the scattering data keeps essentially all useful scattering signals. It was not intended to suggest that scattering alone would be able to separate classes.

Regarding the similarity between control and bacteria in scattering: that is correct and expected. The generator produced a narrow size distribution, and across all groups most particles were in the 4–5 μm range, including the control (Table 2), so similar sizes lead to similar scattering morphologies. In our data, light scattering contributes little discriminative power relative to fluorescence, and performance gains come mainly from lifetime features. This is stated in the Results and Discussion and supported by the PCA and model summaries.

For the time-aggregated fluorescence spectra in Figure 7, all groups share a strong band near 330 nm. This is consistent with protein-like emission under deep-UV excitation, where aromatic residues such as tryptophan produce emission in the 320–360 nm region, and it matches our tryptophan reference measured with 266 nm excitation, noting the instrument's wavelength shift discussed in the manuscript. At the same time, UV-LIF instruments often show background structure from optics and matrix effects, and a protein-like band can also be reinforced by non-biological backgrounds common in LIF systems (Sivaprakasam et al. 2011, Savage et al. 2017). We therefore interpret the control's 320–340 nm feature as a combination of instrument or matrix background plus weak protein-like signals, rather than as evidence that salts fluoresce on their own.

Spectra alone are not sufficiently distinctive in our dataset (Figures 7 and 8), which mirrors prior Rapid-E findings where spectra looked alike yet still supported classification once models combined multiple features (Šaulienė et al., 2019). In our case, lifetimes carry most of the usable contrast: PCA

projections (Figure 7) and model ablations (Figure 8) show that including LT provides the largest gains, while FS and especially LS on their own are weak. This is a setting where machine-learning models help by down-weighting the shared background signal and exploiting the more nuanced temporal signatures of lifetime decays.

To address the final comment: yes, it is possible that a different medium could yield cleaner/purer spectra together with a stronger background handling in the data pre-processing steps. In our setup, a true blank is hard to obtain because (1) the aerosolization step introduces a big variability in the morphology of particles entering the device and (2) we cannot use pure water as the carrier because cells require an isotonic medium to avoid osmotic shock and lysis, which Ringer provides.

For future testing, lower-fluorescence isotonic buffer or reduced salt concentrations should definitely be tested, but again, since our focus was not the interpretability of the signals but on proving the discriminative capability that they carry with the new setup, the fact that the ML models were able to distinguish between classes at such a decent rate in unseen data gives us a very promising head start. And even if that were achieved, comparison with real-world particles or mixtures would not necessary be improved, as as mentioned, real air samples may appear to contain high-salt concentrations and specific matrices bearing some base fluorescence. This should always be benchmarked in every real situation for a proper calibration and validations of results.

We have now added the following lines in the Discussion section (L438-441) to acknowledge the potential issues with the added background noise of the media and potential attenuations:

"As a practical next step, evaluation of lower-fluorescence isotonic carriers and reduced-salt formulations to further suppress background while preserving cell integrity should be tested given the degree of background noise observed in the time-aggregated spectra."

References:

Šaulienė, I., Šukienė, L., Daunys, G., Valiulis, G., Vaitkevičius, L., Matavulj, P., et al. 2019. Automatic pollen recognition with the Rapid-E particle counter: the first-level procedure, experience and next steps. **Atmospheric Measurement Techniques**, 12, 3435–3452.

Sivaprakasam, V., Lin, H. B., Huston, A. L., & Eversole, J. D. (2011). Spectral characterization of biological aerosol particles using two-wavelength excited laser-induced fluorescence and elastic scattering measurements. **Optics Express**, 19(7), 6191–6208.

Savage, N. J., Krentz, C. E., Collins, D. R., & Huffman, J. A. (2017). Systematic characterization of WIBS-4A using fluorescence standards and primary biological aerosol particles. **Atmospheric Measurement Techniques,** 10, 4279–4302.

Response 10: Clarifications on Spectral emission signal shift/missalignment

Line 231-233:

More discussion is needed as to the nature of this shift eg. the shift magnitude varies based on the species; is this to be expected with the proposed explanation of misalignment?

Thank you for this interesting remark. We indeed share the need for further clarification: we strongly believe that the nature of this shift seems to come due to a difference in the step size of the fluorescence emission detector module. Specifically, while the manufacturer claims that they installed a module reading emissions starting at a wavelength of 290 nm and 10nm pixels (so, the complete emission spectrum covered ranges from 290 to 600 nm), the data seems to suggest that in reality, the pixel range is exactly the same as the one of the previous module (14 nm), just with a 40nm shift to lower wavelengths, so in reality our covered range is closer to 290-738nm.

If we plot the aerosolized fluorophores comparing the 10nm pixels vs the 14nm pixels, we see how the fluorescence patterns match exactly those from Pan et al., 2015 that we show in Figure 6B. We attach the comparison figure in the response.
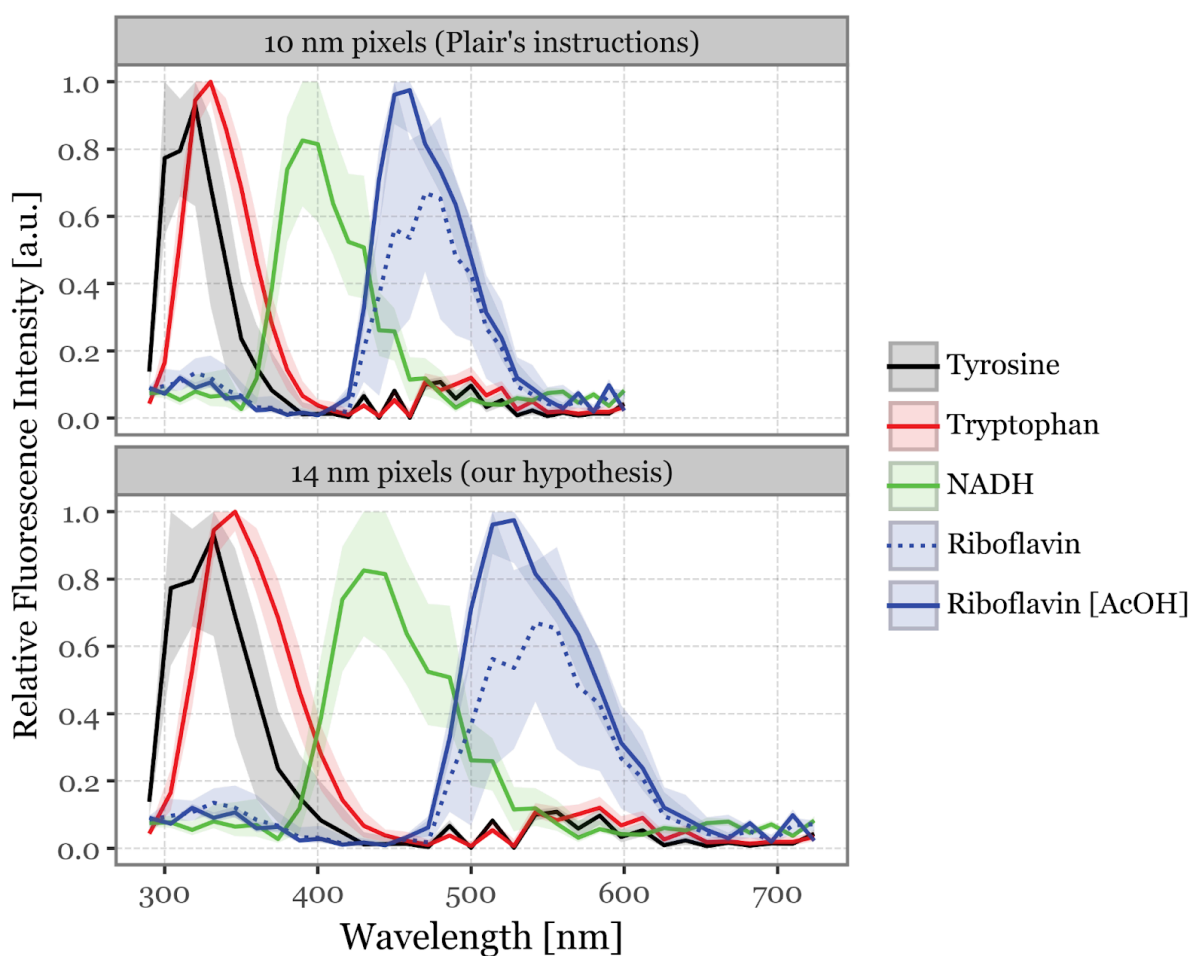
It thus makes sense that the shift increases with distance from the starting wavelength, because an additional 4 nm misalignment accumulates per pixel. This can make the shift appear species-dependent simply because different fluorophores peak at different wavelengths, not because

their spectra are biologically shifted. That said, the effect is negligible for all the results shown in the study, and it would simply imply that the device has capacity for a wider range of emission spectra at the expense of slightly lower resolution.

In any case, we did in fact already briefly discuss this issue in the Discussion section (lines 435-439) and we keep it there in the modified version of the manuscript.

Response 11: Removal of speculative text in the Results section.

Line 292-293:

Save such speculation for the discussion ie. after presenting the machine learning results.

We agree. Speculative remarks about fluorescence-lifetime features have now been removed from lines 292–293. The role of fluorescence lifetime in group discrimination is already covered in the Discussion, after the machine-learning results.

Response 12: Improvement of discrimination power and model choice.

Line 379-380:

Can you suggest other metrics for better discrimination in future? The current discriminatory power is promising but it seems like more is required for applications outside the controlled lab conditions.

We agree. Discrimination is already good in this controlled scenario, and our usage of random forest baseline was meant to show that the data generated by the modified device produces a separable signal by bacterial group that would be pickable by an "easy-to-train" model.

That said, for a post-prototype stage, we see two main lines of improvement:

First, make evaluation and models uncertainty-aware so the system predicts only when confidence is warranted. In practice, what matters here is not classifying every single particle, but being right when we do classify (for this use-case, precision >> recall). We should quantify probability quality and let the model abstain when uncertainty crosses a threshold. In practice, this could mean using approaches such as conformal prediction (Angelopoulos & Bates, 2023) or reject options (Geifman & El-Yaniv, 2019) that would enable the models to have an estimate of prediction quality and "abstain" or back off if a threshold isn't met.

Second, make targets and metrics taxonomy-aware. Our current prototype treats all targets as equidistant, but taxonomy is hierarchical. Mistaking *Bacillus cereus* for *Bacillus endophyticus* is far less severe than mistaking *Staphylococcus hominis* for a fungus or an abiotic particle. Encoding taxonomic structure could improve both training and evaluation: use hierarchical labels with back-off to genus/family when species-level confidence is low, and weight errors by taxonomic distance. This approach mirrors the common practice in metagenomic taxonomy annotation such as the RDP classifier and MEGAN's lowest-common-ancestor (LCA) (Wang et al., 2007; Huson et al., 2007; Kim et al., 2016), but also falls in line with hierarchical-classification methods applied to several domains (Silla & Freitas, 2011).

In short, making the system aware of both the uncertainty of its predictions and the hierarchical nature of its targets should improve discrimination where it matters, keep false positives low, and scale as we expand classes to fungi and common interferents.

We have added a mention to these two targets in the final paragraphs of the discussion, which now reads on lines 462-468:

"Finally, taxonomy is hierarchical, a property that our current model omits, but that means that targets can be structured across ranks with a back-off rule: assign species or genus only when confidence is high, and default to higher ranks when it is not. This is standard practice in sequence-based classifiers such as the RDP naïve Bayesian classifier and MEGAN's lowest-common-ancestor approach, which label to the highest taxonomic rank supported by the data in metagenomics studies (Huson et al., 2007; Wang et al., 2007). The same logic could be adapted here if the class set and evaluation follow the rank structure. Any real usage in the field will need validation with co-located reference sampling to establish site-specific baselines and thresholds."

References:

Angelopoulos, A. N., & Bates, S. (2023), "Conformal Prediction: A Gentle Introduction", **Foundations and Trends in Machine Learning**: Vol. 16: No. 4, pp 494-591

Geifman, Y., & El-Yaniv, R. (2019). SelectiveNet: A deep neural network with an integrated reject option. **ICML**.

Kim, D., Song, L., Breitwieser, F. P., & Salzberg, S. L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. **Genome Research**, 26(12), 1721–1729.

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. **Genome Research**, 17(3), 377–386.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. **Applied and Environmental Microbiology**, 73(16), 5261–5267.

Silla, C. N., & Freitas, A. A. (2011). A survey of hierarchical classification across different application domains. **ACM Computing Surveys**, 44(1), 1–35.