scientific reports

OPEN

Particulate matter emission sources and meteorological parameters combine to shape the airborne bacteria communities in the Ligurian coast, Italy

Giorgia Palladino[1,3], Pietro Morozzi[2], Elena Biagi[1,3], Erika Brattich[4], Silvia Turroni[1], Simone Rampelli[1], Laura Tositti[2✉] & Marco Candela[1,3✉]

Aim of the present study is to explore how the chemical composition of particulate matter (PM) and meteorological conditions combine in shaping the air microbiome in Savona (Italy), a medium-size, heavily inhabited urban settlement, hosting a wide range of industrial activities. In particular, the air microbiome and PM_{10} were monitored over six months in 2012. During that time, the air microbiome was highly dynamic, fluctuating between different compositional states, likely resulting from the aerosolization of different microbiomes emission sources. According to our findings, this dynamic process depends on the combination of local meteorological parameters and particle emission sources, which may affect the prevalent aerosolized microbiomes, thus representing further fundamental tools for source apportionment in a holistic approach encompassing chemical as well as microbiological pollution. In particular, we showed that, in the investigated area, industrial emissions and winds blowing from the inlands combine with an airborne microbiome which include faecal microbiomes components, suggesting multiple citizens' exposure to both chemicals and microorganisms of faecal origin, as related to landscape exploitation and population density. In conclusion, our findings support the need to include monitoring of the air microbiome compositional structure as a relevant factor for the final assessment of local air quality.

Airborne Particulate Matter (APM) is a complex system of particles in suspension in the atmosphere, usually defined as aerosol. Atmospheric aerosol is contributed by a multiplicity of sources of both natural and anthropogenic origin, including both biogenic and abiotic chemical components, and producing extremely complex and variable matrices that can be processed and solved for their origin using appropriate analytical processing and computational tools[1,2]. In particular, the aerosol composition consists of a series of macrocomponents, which make up the mass of APM, as well as an even larger series of different trace components, the latter being of primary relevance as including the most toxic species and providing the highest chemical fingerprinting potential[3]. These aerosol bulk components can be emitted directly into the atmosphere (Primary Aerosol) or, otherwise, they can be abundantly produced within the atmosphere, following chemical reactions on gaseous precursors previously emitted (Secondary Aerosol). Primary Biological Aerosol (PBA), in short bioaerosol, represents the APM fraction including atmospheric particles released from the biosphere to the atmosphere[4]. PBA comprise living and dead organisms, their dispersal units (e.g. pollen and spores) as well as tissue fragments from decay processes[5,6]. The overall mass contribution of PBA to conventional APM metrics is to date a very challenging task though some authors have recently estimated that it may account for about 16% of PM_{10} in different cities examined[7]. The PBA fraction including microorganisms is defined as "airborne microbiome" (AM) and represents a highly dynamic and diversified assemblage of active and inactive microorganisms[8]. Indeed, AM can originate from multiple terrestrial and marine sources—including resident microbiomes in soil, waterbodies,

[1]Unit of Microbiome Science and Biotechnology, Department of Pharmacy and Biotechnology, University of Bologna, Via Belmeloro 6, 40126 Bologna, Italy. [2]Department of Chemistry "Giacomo Ciamician", University of Bologna, Via Selmi 2, 40126 Bologna, Italy. [3]Fano Marine Center (FMC), Viale Adriatico 1, 61032 Fano, Italy. [4]Department of Physics and Astronomy, University of Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy. ✉email: laura.tositti@unibo.it; marco.candela@unibo.it

plants and animal faeces[4,9,10]—whose relative importance depends season, location, altitude and meteorological and atmospheric factors. Further, in agricultural and suburban locations, other sources relevant to AM emissions are represented by man-made systems, such as agricultural waste, composting, and wastewater treatment plants. AM emission mechanisms include erosion or abrasive dislodgement from terrestrial sources and, from open waters, bubble-bursting at the air–water interface[11,12]. PBA size spans from a few nanometres up to about a tenth of a millimetre[6], with bacteria-containing particles ranging around 2–4 μm in diameter[13] and accounting for "5–50% of the total number of atmospheric particles > 0.2 μm in diameter"[14]. Due to the small size, AM can be transported over large distances, across continents and oceans, and reach the upper troposphere, where it actively contributes to ice nucleation and cloud processing[15]. In the troposphere, the AM concentration ranges from 10^2 to 10^5 cells/m^3[13], being the densest in the planetary boundary layer, whose thickness depends on micro-meteorological factors and geographic location, with marked daily and seasonal fluctuations[16,17]. In particular, the near-ground AM is the one most influenced by local sources, including local meteorology and atmospheric composition. AM is then removed from the troposphere by wet and dry deposition processes. The former is the major sink for atmospheric aerosol particles, in the form of precipitation[18], while the latter, being less important on the global scale, is particularly relevant with respect to local air quality[4,8].

Recently, an increasing perception of the strategic importance of PBA—AM in particular—for the Earth system and, ultimately, for the planet and human health, has arisen[19–21]. For instance, besides its relevance to atmospheric processes, AM has been found to control the spread of microorganisms over the planet surface, affecting the geographical biome, with key implications on agriculture and, ultimately, human health. This awareness raised concern about the potential impact of anthropic activities on PBA and, in particular, on the AM fraction. For example, changes in aerosol composition due to extensive human influence on the planetary scale give rise to air pollution, the inherent modification of atmospheric reactivity and, ultimately, climate change[22]. These factors may likely interfere with AM, shaping its structure and dispersion throughout the troposphere, with direct consequences on the terrestrial biome[23]. However, as far as we know, the current state of knowledge on the connections between AM, atmospheric processes and atmospheric pollution is still fragmentary, with relatively few pioneer studies that take into account multidimensional datasets to analyse the connection between air chemical composition and its microbiome structure and variations, e.g. comparing different land-use types[24,25], different sites within cities[26–29], or worldwide locations[8,30,31].

Our work fits in this scenario by exploring the ability of an interdisciplinary approach combining chemical speciation and metagenomics in shedding light on the complex relationships among abiotic and microbiome components of local ambient aerosol. The study is based on a series of about one hundred PM$_{10}$ samples from a coastal district in north-western Italy, collected daily over six months, from February to July 2012, to cover the cold-to-warm seasonal transition. The chemical composition of each sample was obtained, and a receptor modelling approach was used to identify and quantitatively apportion the chemical species determined in the samples to their sources. Owing to the cutoff adopted in APM sampling, the samples were deemed suitable for total DNA extraction and microbiome characterization by Next-Generation Sequencing using the 16S rRNA gene as target. In our work, we were able to finely reconstruct the overall aerosol behavior in an area affected by both natural and anthropogenic emission sources, determining the local bacterial microbiome from PBA contained in PM$_{10}$ and its main features as a function of local meteorological and environmental characteristics.

Results

Particulate matter emission sources and atmospheric parameters.

The PMF model application on PM$_{10}$ samples resulted in a solution with an optimum number of seven source factors at the receptor site, *i.e.* the station where the PM$_{10}$ samples were collected. Like other multivariate methods, these factors correspond to linear combinations of the original compositional parameters, each potentially identifiable as an emission source profile. The fractional contribution per sample for each of the seven factors is reported in Supplementary Table S1.

In order to associate the factors with specific emission sources, prior knowledge about the receptor site (Savona, Italy) was used together with a critical analysis of the factor fingerprints (Fig. 1a). Moreover, the percentage contribution of the seven identified sources to the total variable was reported (PM$_{10}$, Fig. 1b). As a result, the seven factors extracted by PMF analysis can be described as follows. Factor 1 is characterized by the prevalence of elements attributable to the geochemical composition because of the high percentages of Si, Al, and Ti. Therefore, this factor was identified as "crustal material and road dust resuspension", deriving from the soil and/or road surface[32,33]. Factor 2 is linked to organic carbon (OC), Cu, Zn, Cr, and K$^+$. OC and K$^+$ are strictly related to combustion processes, including biomass burning, as previously described[34]. Cu, Zn, and Cr are associated with traffic: Cu and Cr are well-known tracers of the brakes of motor vehicles, while Zn is known as a tracer of tire wear[35–37]. Therefore, this factor was identified as a combination of "traffic and biomass burning" sources. Factor 3 is mainly associated with NO$_3^-$ from gas-to-particle conversion of NO$_x$ (g) in the atmosphere to which traffic and other high-temperature combustion processes may contribute[38,39]; as such it can hardly be attributed to a single well-defined source, especially in such a complex emissive scenario. Therefore, this factor was identified collectively as "secondary nitrate". Factor 4 relates to SO$_4^{2-}$ and NH$_4^+$ from gas-to particle reactions, leading to secondary ammonium sulphate[40–42]. Similarly to secondary nitrate, this component can be contributed by various sources (both natural and anthropogenic) due to the multiplicity of fossil fuel sources of the precursor gaseous SO$_2$ and the ubiquity of NH$_3$ (g)[43,44]. Therefore, this factor was collectively identified as "secondary ammonium sulphate". Factor 5 is associated with Na, Mg ins, V, and Ni. The distinctive association of V and Ni reveals emissions attributable to the combustion of heavy oil[45–47]. The association of these species with Na and Mg suggests a "naval-maritime transport" source. Factor 6 is mainly characterized by high scores of Pb, K ins, Zn, OC, and elemental carbon (EC). The fine particles produced by coal combustion are characterized by significant fractions
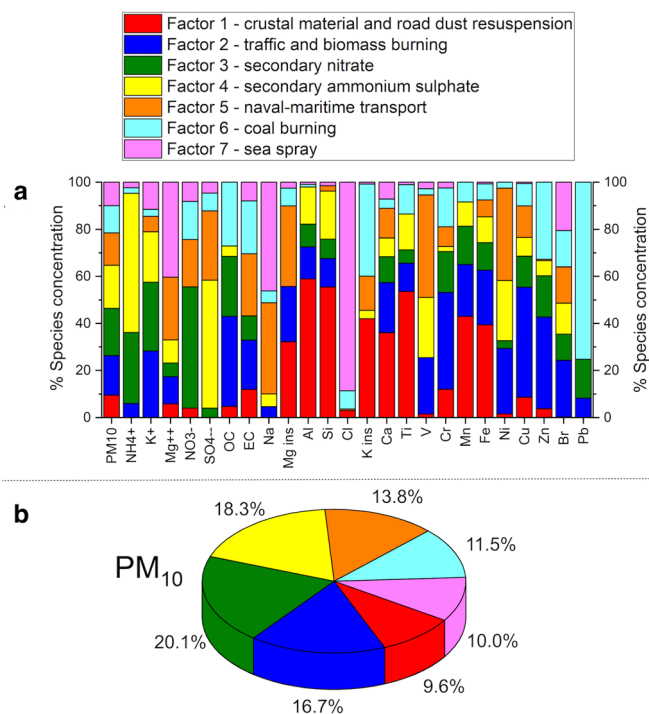
Figure 1. Emission sources identified by PMF analysis. (**a**) Stacked bar chart of the percentage concentration of each chemical species contributing to each of the seven factors that represent the chemical profile of each source identified in the PMF model. (**b**) Pie chart representing the contribution of the seven sources to PM_{10} mass. The seven factors were identified as reported at the top.

of OC and K together with typical elements such as Zn, while other semi-volatile elements condense on the surface of fine particles of K ins[48]. Therefore, this factor was identified as "coal burning". Factor 7 is connected to a large score of Cl⁻, Na, and Mg⁺⁺, and clearly identified as "sea spray" aerosol[49].

In order to confirm the PMF analysis results, the origin of the polluted air masses was investigated by analyzing the PMF factors as a function of wind direction, calculating the respective cumulative distribution functions and generating the corresponding wind polar plots. This method associates the emissive profile obtained by PMF with wind direction and intensity to which the receptor site is downwind. The plots obtained are shown in Supplementary Figure S1. In particular, factors 1, 3, 4, and 6 (respectively crustal material and road dust resuspension, secondary nitrate, secondary ammonium sulphate, and coal burning) are associated with winds blowing from the inland towards the coast covering traffic and industrial sources. Factor 5 (naval-maritime transport) is oriented downwind from the sea, confirming that it is associated with the fuel oil used for sea shipping. Finally, while factor 2 shows a local origin indicating sources in the proximity of the receptor site, factor 7 is meridionally oriented, indicating once more a marine origin. It should be noted, however, that, unlike factor 5 characterized by elements typical of the submicron fraction likely flushed back and forth by sea-land breezes from the harbor, factor 7 is associated with coarse particles requiring different meteorological conditions (possibly more intense winds from the open sea in order to sustain heavier particles).

AM overall composition.

Next generation sequencing of the V3-V4 hypervariable region of the 16S rRNA gene from the total microbial DNA extracted from PM_{10} air filters resulted in 98 samples containing more than 1 000 reads per samples which were retained for the rest of the study, for a total of 797,781 high-quality sequences with an average of $8,058 \pm 3,410$ (mean ± SD) paired-end reads per sample, binned into 4 189 ASVs. According to our data, AM is dominated by the phyla Proteobacteria (mean relative abundance ± SD = $42.8 \pm 19.4\%$) and Firmicutes ($27.4 \pm 18.9\%$), with Actinobacteria ($14.8 \pm 10.9\%$) and Bacteroidetes ($9.2 \pm 8.6\%$) being subdominant. At the family level, the most represented taxa are *Comamonadaceae* ($6.1 \pm 13.4\%$) and *Sphingomonadaceae* ($4.3 \pm 5.0\%$), both belonging to Proteobacteria. Other represented families are *Ruminococcaceae* ($3.9 \pm 7.6\%$), *Enterobacteriaceae* ($3.7 \pm 5.9\%$), *Clostridiaceae* ($3.6 \pm 6.8\%$), *Bacillaceae* ($3.5 \pm 5.0\%$) and *Flavobacteriaceae* ($3.4 \pm 5.7\%$). Please see Supplementary Figure S2 for a graphical representation of the overall compositional structure of AM throughout the entire sampling period.

In order to explore connections between the AM structure and seasonality, we compared the levels of AM diversity over the different months (Fig. 2). Diversity measurements indicated a general trend of microbial richness to decrease from winter to summer, although the differences did not reach statistical significance (Kruskal–Wallis test, FDR corrected p-value > 0.05) (Fig. 2a). Conversely, the PCoA of unweighted UniFrac distances between the AM compositional profiles showed sample segregation according to the month of sampling
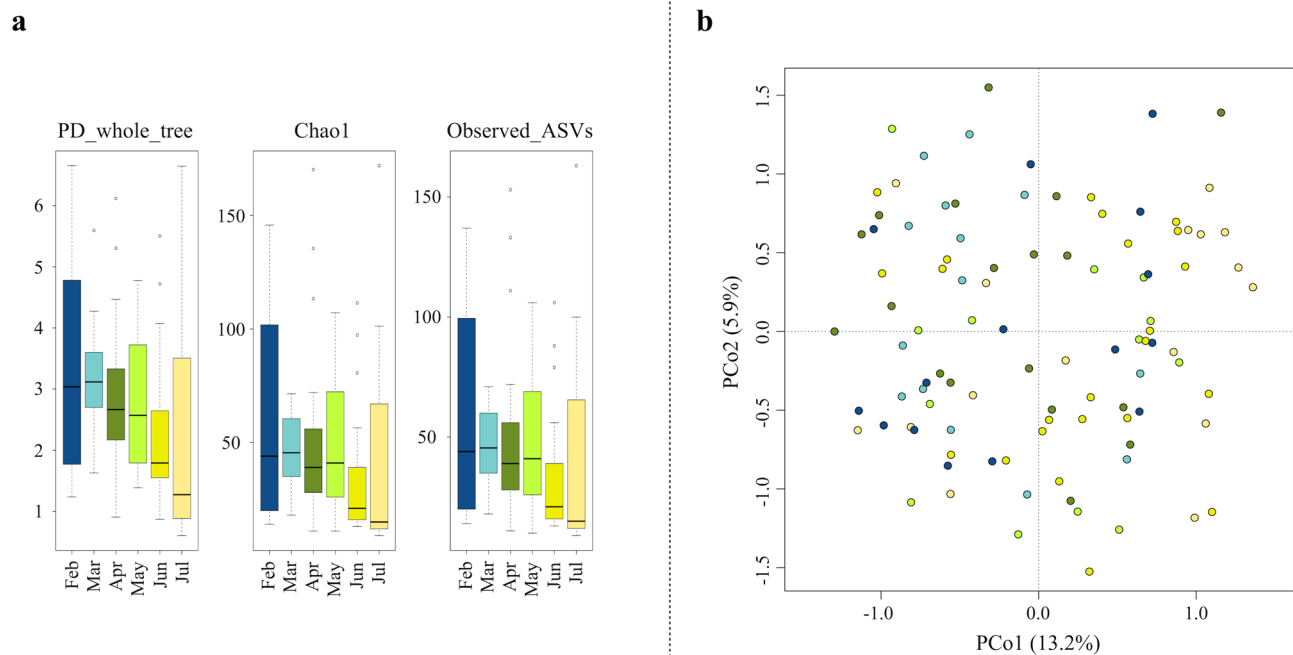
Figure 2. AM alpha and beta diversity throughout the sampling period. (**a**) Box-and-whiskers distribution of Faith's Phylogenetic Diversity (PD_whole_tree), Chao1 index for microbial richness and number of observed ASVs, for each month of sampling. The data show a trend towards reduced microbial richness from winter to summer, although the differences did not reach statistical significance (Kruskal–Wallis test, FDR-corrected p-value > 0.05). (**b**) Principal Coordinates Analysis (PCoA) based on unweighted UniFrac distances between AM profiles, showing separation by sampling month (permutation test with pseudo-F ratio, p-value = 0.012) (same colour code as in panel A). The first and second principal components (PCo1 and PCo2) are plotted and the percentage of variance in the dataset explained by each axis is reported.

(Fig. 2b) (FDR-corrected permutation test with pseudo-F ratio, p-value = 0.012), meaning that seasonality significantly affects the overall compositional AM structure.

Variation of the AM topological structure and association with PM emission sources and meteorological parameters.

To further explore the overall AM variation across the sampling period, a clustering analysis of the AM compositional profiles was carried out. Hierarchical Ward-linkage clustering based on the Spearman correlation coefficients of family-level AM profiles resulted in the significant separation of 4 clusters, named C1, C2, C3 and C4, respectively (FDR-corrected permutation test with pseudo-F ratio, p-value ≤ 0.001) (Fig. 3). Confirming the robustness of the identified clusters, the PCoA of the unweighted UniFrac distances between samples revealed a sharp segregation based on the assigned cluster (Fig. 4). Interestingly, when we searched for correlations between PCoA coordinates and measured meteorological parameters or PMF factors (Supplementary Tables S2 and S1, respectively), we found that factor 5 (naval-maritime transport) and relative humidity (RH) were both positively correlated with the PCo1 axis (Kendal's test, FDR-corrected p-value ≤ 0.001), while factor 6 (coal burning) was negatively correlated with the PCo1 coordinates (p-value ≤ 0.001) (Fig. 4). This indirect gradient analysis allowed to highlight positive associations between clusters C1 and C3 and factors 6 and 5, respectively. Further, cluster C3 was found to be positively related to RH. As for seasonality, the clusters C3 and C4 are the most prevalent in summer and winter, respectively, while for C1 and C2 we did not observe any prevalence for a particular sampling period. We also compared the microbial diversity values of samples included in the different clusters, using three different diversity metrics. Our data indicated higher biodiversity in clusters C1 and C2 (PD whole tree, chao1, and observed ASVs, mean ± SD: 3.6 ± 1.3, 67.2 ± 43.9, and 65.0 ± 40.5 for C1, 3.2 ± 1.4, 59.1 ± 38.2 and 57.6 ± 36.0 for C2, respectively) compared to C3 and C4 (1.4 ± 0.7, 18.7 ± 9.4, and 18.5 ± 9.3 for C3, 2.4 ± 1.2, 31.2 ± 18.4 and 31.2 ± 18.3 for C4), with C3 having the lowest biodiversity (Kruskal–Wallis test, FDR corrected p-value ≤ 0.001).

Compositional specificity and prevalent microbiological source of the four AM clusters.

We subsequently compared the relative abundance of AM families among the four clusters in order to find out the most distinctive families of each of them (Supplementary Figure S3). According to our findings, the discriminating families (i.e. families with significantly different relative abundance, based on Kruskal–Wallis test) for the microbial cluster C1 are *Prevotellaceae*, *Erysipelotrichaceae*, *Coriobacteriaceae*, *Christensenellaceae*, *Lachno-*
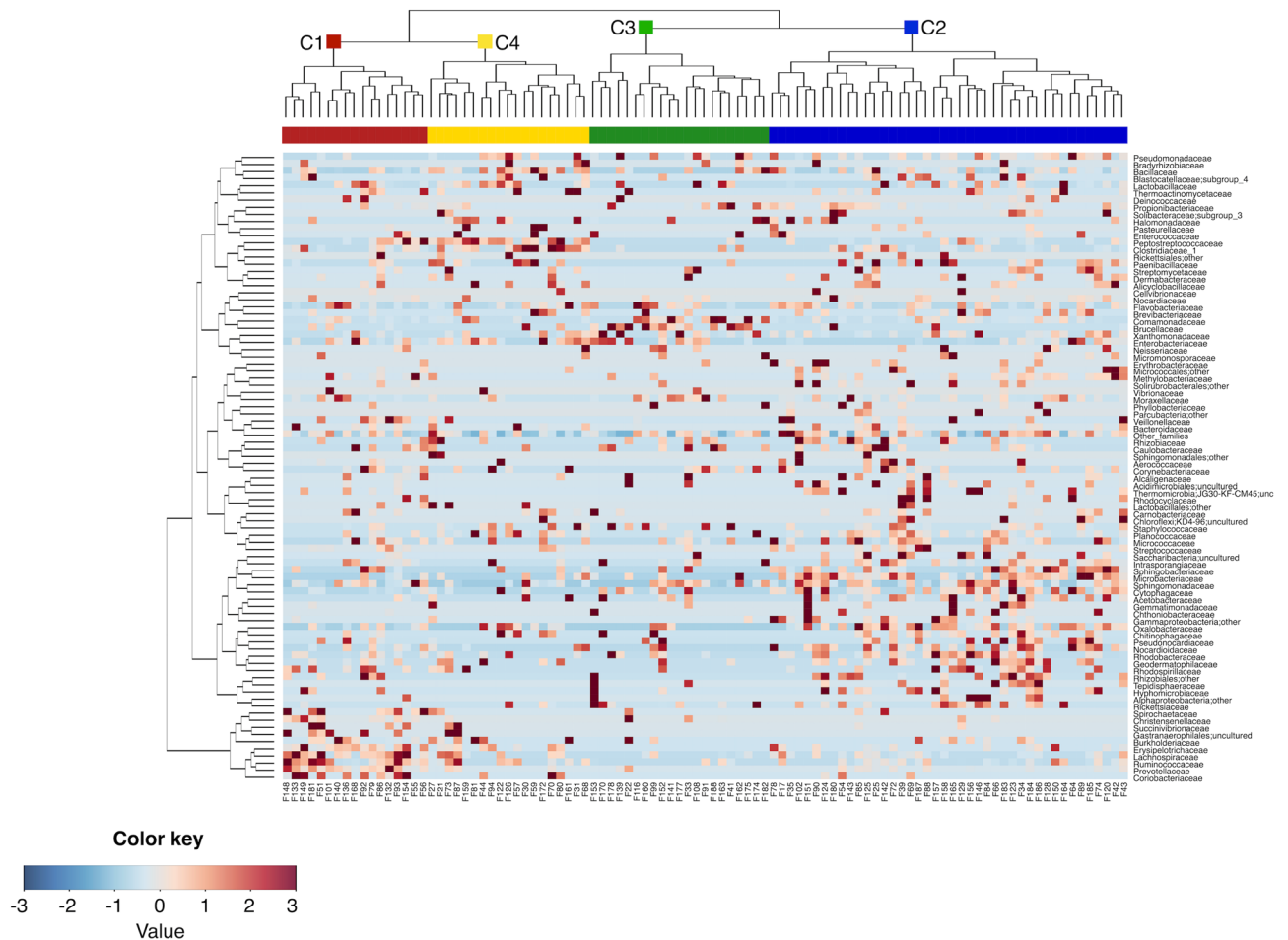
Figure 3. Family-level clusters of the airborne microbiome. Hierarchical Ward-linkage clustering based on the Spearman correlation coefficients of the proportion of families in the AM samples. Only families with relative abundance > 2% in at least 3 samples were retained. The four identified clusters (FDR-corrected permutation test with pseudo-F ratio, p-value ≤ 0.001) are labelled in the top tree and highlighted by different coloured squares (red, blue, green and yellow for the clusters C1, C2, C3 and C4, respectively).

spiraceae, *Ruminococcaceae,* and *Spirochaetaceae*. The microbial cluster C2 is instead characterized by higher abundance in the families *Microbacteriaceae*, *Cytophagaceae*, *Oxalobacteraceae*, *Sphingobacteriaceae*, *Nocardioidaceae*, *Methylobacteriaceae*, *Intrasporangiaceae*, *Rhodobacteraceae* and *Acetobacteraceae*. Only two proteobacterial families, namely *Brucellaceae* and *Comamonadaceae,* have a significantly higher abundance in cluster C3. Four families show higher abundance in cluster C4, i.e. *Peptostreptococcaceae*, *Clostridiaceae*, *Bacillaceae* and *Enterobacteriaceae*. It is also worth noting that the families *Planococcaceae* and *Paenibacillaceae* are highly represented in both C2 and C4 clusters, whereas *Sphingomonadaceae* members are equally represented in all clusters except for C4.

In an attempt to identify the most likely prevalent microbial origin of the four AM clusters, we first derived the respective compositional peculiarities at the OTU level. To this aim, 16S rRNA gene reads were clustered at 97% homology, resulting in 3 821 OTUs. By linear regression, we subsequently obtained 80 OTUs specifically discriminating the four clusters. In particular, for 52 of these OTUs a significantly different distribution in the four clusters was confirmed by a Kruskal–Wallis test, as shown in Supplementary Figure S4. For each of them, the isolation source of the closest BLAST match within the NCBI 16S rRNA sequence database was recovered (Supplementary Table S3). Interestingly, according to our findings, the cluster C1 is mainly characterized by OTUs of faecal origin. These OTUs include sequences assigned to typical components of the human gut microbiome, such as *Faecalibacterium prausnitzii*, *Ruminococcus faecis*, *Prevotella copri*, *Eubacterium eligens*, *Ruminococcus bromii*, *Roseburia inulinivorans* and *Blautia faecis*[50–52], the cattle rumen components *Succinivibrio dextrinosolvens*[53] and *Oscillibacter ruminantium*[54], and the porcine gut microbiome member *Treponema porcinum*[55]. Differently, the cluster C2 is characterized by OTUs assigned to microorganisms isolated from plant roots and leaves, including *Curtobacterium flaccumfaciens*[56], *Glutamicibacter halophytocola*[57] and *Frigoribacterium endophyticum*[58], as well as by a specific pattern of environmental bacteria, from soil, air, and fresh and marine water ecosystems. Similarly, both clusters C3 and C4 are characterized by a peculiar combination of environmental microorganisms from different sources, including soil, fresh and marine waters, and airborne microbial ecosystems.
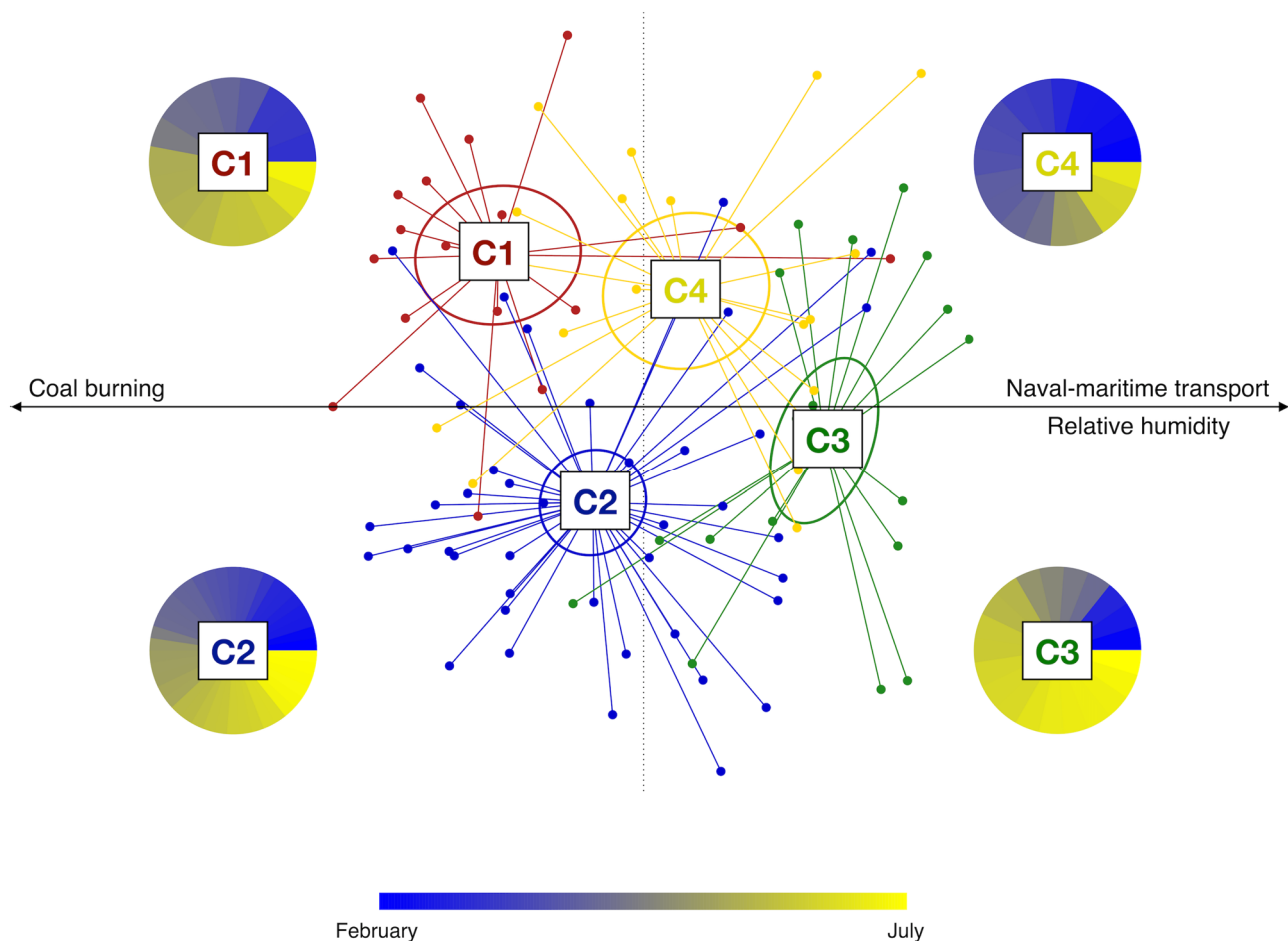
Figure 4. Variation of the AM topological structure and association with PM emission sources and meteorological parameters. Principal coordinates analysis (PCoA) based on the unweighted UniFrac distance shows separation between the microbial clusters (C1 to C4; permutation test with pseudo F-ratio, p-value ≤ 0.001; see also Fig. 3). The percentage of variance in the dataset explained by each axis, first and second principal component (PCo1 and PCo2), is 13.2% and 5.9%, respectively. Ellipses include 95% confidence area based on the standard error of the weighted average of sample coordinates. Significant Kendall correlations between PCoA axes and PMF factors and measured meteorological parameters are reported with a black arrow. Specifically, the emission source factor 5 (naval-maritime transport) and relative humidity are both positively correlated with the PCo1 axis (Kendall correlation test, FDR-corrected p-value ≤ 0.001), while the emission source factor 6 (coal burning) is negatively correlated with the PCo1 coordinates (p-value ≤ 0.001). For each AM cluster, the proportion of samples based on the sampling time (from February (dark blue) to July (yellow)) is shown as a pie chart.

Discussion

In order to explore connections between the local air microbiome, atmospheric pollution and meteorological factors, here we provide a longitudinal survey of the near-ground AM, atmospheric particulate and atmospheric parameters in Savona, Italy. According to our findings, the local AM appears dominated by the phyla Proteobacteria, Firmicutes, Actinobacteria and Bacteroidetes, well matching the general layout of an AM community[4,25]. The application of the PMF receptor modelling on the chemical compositional pattern of the PM_{10} samples collected during the field campaign allowed the identification of seven emission sources: "crustal material and road dust resuspension", "traffic and biomass burning", "secondary nitrate", "secondary ammonium sulphate", "naval-maritime transport", "coal burning" and "sea spray". Each source factor was subsequently subjected to anemological analysis based on polar plots, allowing each emission source to be associated with the corresponding wind direction to which the receptor site is downwind, similarly to the approach used by Innocente et al.[28]. Specifically, emission sources as "crustal material and road dust resuspension", "secondary nitrate", "secondary ammonium sulphate" and "coal burning" were associated with winds blowing from the inland toward the sampling site, intercepting substantially traffic and industrial particulate sources. Conversely, emission sources such as "naval-maritime transport" and "sea spray" were associated with a sea breeze, supporting a marine origin for both. Finally, the "traffic and biomass burning" emission source mostly showed a local origin.

It is known that bacterial communities structure and composition can be influenced by environmental conditions, such as seasons, air masses origin and landscape characteristics[28,31,59], as well as PM_{10} chemical

compositional pattern[27]. When we explored the AM structure variation during the observation period, we were able to identify four distinct clusters of samples, named C1 to C4. Interestingly, the four clusters were associated with a peculiar combination of seasonality, meteorological variables and emission sources, as integrated into factors by PMF analysis.

In particular, the AM cluster C1 was associated with the "coal burning" emission source, suggesting not actually the industrial facility as a microbiome source, but rather the influence of an air mass whose transport over a given district harvests chemical and microbiological components along the same tropospheric path. Instead, the cluster C3, most represented in the warm period, probably has a marine origin due to its association with the "naval-marine transport" emission source and high relative humidity. Finally, the clusters C2 and C4 did not show any specific association with the aerosol sources assessed by PMF, even if they showed a different seasonal behaviour, with C4 being more represented in the cold period. The results obtained by the application of such double-step multivariate analysis—associating multidimensional microbiome and chemical datasets—represent the true novelty of our study. Indeed, we were able to associate the AM composition to well discriminated emission sources, each with a characteristic chemical composition and origin.

The four AM clusters revealed a distinct, well-defined compositional structure, each being enriched with a specific set of microbial families and OTUs. The specificity of each bacterial profile de facto serves as a microbiological fingerprint, allowing to single out the probable microbiome sources characterizing each cluster that, similar to what occurs to abiotic particles, allow to trace back the origin of the air mass. In particular, the clusters C3 and C4 substantially reflect interconnected environmental microbiomes, encompassing a specific combination of microorganisms from soil resuspension, as well as from marine and fresh waters (possibly from rivers and streams flowing into the Ligurian Sea) and from the air. C2 cluster reveals the plant microbiome as an additional source, showing a further combination of plant-associated and environmental microorganisms, due to the contact of air masses over a vegetation landscape. Interestingly, the feasibility of air mass tracing also using bacterial species clearly emerges when we observe in detail the compositional structure of C1. This is in fact the only AM cluster carrying a recognizable pool of bacterial moieties of faecal origin, which are consistently part of the animal gut microbiome, suggesting not only a well-defined origin but also the potential use of this information in the assessment of microbiological impacts. Faecal material as a potential source of bacteria was already reported by Bowers et al.[24]. It should be noted that in the area upwind C1 no sewage treatment plant as a possible source of faecal microbiome was present at the time of sampling. However, the area is very densely populated and forested areas populated by local fauna are closely found within a few kilometres.

Taken together, our data on the temporal dynamics of the near-ground AM in Savona, highlight the relevant degree of plasticity of AM over time. As such, we demonstrated how meteorological factors (e.g. wind direction and humidity) and atmospheric pollution (particles emission sources) can combine in shaping the AM configuration. In particular, coal burning and winds blowing from the inlands mix to establish a characteristic AM with a prevalence of aerosolized faecal microorganisms, regardless of seasonality. Conversely, in the summer season, humidity, sea breeze and naval-marine transport pollutants result in an AM mainly originating from environmental microbiomes, including microorganisms that are typically found in seawater and soil. Even if we were not able to establish connections between the other identified emission sources and specific AM clusters, we would stress the importance of seasonality in shaping the AM structure. Indeed, the variation between the clusters C2 and C4, for which no connection with any emission source was observed, was shown to be dependent on the sampling period, with the cluster C2 most prevalent during the warm season and including plant microbiomes as possible characteristic sources, as previously observed by Franzetti et al.[60] and Bertolini et al.[61].

In conclusion, our results suggest that, in an urban settlement, the influence of air masses harvesting chemical components from industrial sources may increase the proportion of aerosolized faecal microorganisms in the atmosphere, ultimately increasing citizens' exposure to faecal microbes. Similar results have recently been obtained by exploring AM in Beijing over 6 months[23]. Our findings strengthen the importance of including the monitoring of the AM compositional structure as a determinant factor in the currently used air quality indexes. Indeed, in urban areas, the possible increased exposure to faecal-associated microbiomes as a result of increasing pollution can pose a possible threat to human health, particularly in regions with high-intensity animal farming, due to the inherent propensity of opportunistic pathogens to aerosolize.

Materials and methods

Site description. The PM_{10} samples treated in this work were collected in Savona, one of the main towns in the Ligurian region (Fig. 5). The whole region overlooks the Tyrrhenian sea but is entirely occupied by the Apennine range down to the coast, where only a narrow strip of plain is present. Therefore, the coastal area is densely inhabited and crossed by an extremely busy traffic road mainly connecting Italy to France. Besides being occupied by a medium-size heavily inhabited urban settlement, the Savona district also hosts a wide industrial area, including a coal-fired power plant active at the time of our experimental field activity and a large and very busy harbour. The climate of this site is classified as warm-temperate (Csa, according to Köppen and Geiger classification)[62,63] with an average annual temperature of 14.6 °C and average precipitation of 910 mm (https://en.climate-data.org, accessed 28/07/2020). Intense northern winds characterize the circulation in winter[64], while sea-land breeze regimes prevail in the warm season, usually starting from March[65,66]. The air sampling site was located at an altitude of 12 m on the sea level in a rural area as classified according to the EU Directive 2008/50/EC on Air quality, with a distance of 2 km from the sea.

Sample collection and atmospheric parameters. A total of 184 daily PM_{10} samples were collected from February 1, 2012, until July 20, 2012 with low-volume samplers (SWAM Dual Channel, 55.6 m³/day, FAI, Italy) to allow simultaneous collection of both quartz (Whatman ®QM-A quartz) and PTFE membranes (What-

Figure 5. Location of the sampling site. Map providing the location of Savona in Italy (indicated with a yellow balloon with a star; other Italian cities are indicated with a green balloon) (top panel), and snapshot of the PM$_{10}$ sampling site with a 3D view of the surroundings (bottom panel) (source: Google Earth; map data: SIO, NOAA, U.S. Navy, NGA, GEBCO, TerraMetrics). Adapted from the original file of P. Christener (https://commons.wikimedia.org/wiki/File:Vado-Ligure.jpg), CC BY-SA 3.0 (), via Wikimedia Commons.

man PM$_{2.5}$ PTFE). Samples were stored frozen in the dark at −10 °C until processing. In this work, PTFE membranes were used for gravimetry, ion chromatography and elemental analysis with particle induced X-ray emission and inductively coupled plasma mass spectrometry, while quartz membranes were used for the analysis of carbonaceous macrocomponents and microbiology. A subset of 98 samples, uniformly distributed across the sampling period, was used for the analyses reported in the present paper. During the sampling campaign, meteorological parameters were measured simultaneously on site using a Davis Vantage Pro2 Weather Station (Davis Instruments, Hayward, CA), placed in proximity of the PM$_{10}$ sampler, for the measurement of temperature, pressure, relative humidity, rainfall, and wind direction and speed with a time resolution of 30 min. Subsequently, the data obtained were averaged on a daily scale (Supplementary Table S2), i.e. at the same time resolution as the PM$_{10}$ samples, using the "openair" package[67] of the R software (version 3.6.1)[68].

Chemical characterisation of the samples. Chemical characterization of PM$_{10}$ filters was carried out using several analytical techniques. First, PM$_{10}$ mass load (µg/m^3) was determined by gravimetric analysis. Elemental and organic carbon were determined on quartz membranes by thermal-optical transmittance analysis (TOT), as previously described[69]. For inorganic speciation, several analytical techniques were performed on PTFE filter portions: Ion Chromatography (IC) for the determination of the main water-soluble ion composition (NH$_4^+$, K$^+$, Mg^{2+}, NO$_3^-$, SO$_4^{2-}$, Na$^+$, Cl$^-$, Ca^{2+}, and a few low-level organic compounds, i.e. oxalates and methanesulfonate), and Particle Induced X-ray Emission (PIXE) and Inductively Coupled Plasma Mass Spectrometry (ICP-MS) for the simultaneous analysis of a series of metals and metalloids (Na, Al, Si, Cl, Ca, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, Br, Pb, Li, Co, Rb, Sr, Cd, La, Ce, Sb, Cs, Ba, Ti, Bi, As, Se, Sn). Elemental analysis by PIXE was carried out at the Tandetron 3 meV of LABEC-INFN, Florence (Italy), according to the method previously reported[70]. Elemental analysis by ICP-MS was carried out according to the UNI EN 14,902, 2005 for PM$_{10}$ as an extension of the DL 155, 2010 in agreement with the EU Directive 2008/50/EC on ambient air quality and cleaner air for Europe. In order to prevent data redundancy, insoluble magnesium (Mg ins) and insoluble potassium (K ins) were calculated as the difference between PIXE and IC concentrations and replaced the corresponding elementary concentration data.

Positive Matrix Factorization analysis.   Positive Matrix Factorization (PMF) is an advanced multivariate factor analysis technique widely used in receptor modelling for the chemometric evaluation and modelling of environmental datasets[3,32,71–75]. PMF allows the identification and quantification of the emissive profile of a receptor site, i.e. the monitoring site where an air quality station is operated. We applied EPA PMF 5.0 software[76]. The dataset was checked and re-arranged prior to PMF modelling according to the model criteria previously described[76] and, after data pre-processing, a concentration matrix of 98 samples × 25 variables was obtained. After careful evaluation of the input data and uncertainty matrices, an optimum number of factors was found by analysing the values of Q, a parameter estimating the goodness of the fit performed[77], and the distribution of residuals. In order to assess the reliability of the model reconstruction, measured (input data) and reconstructed (modeled) values together with the distribution of residuals were compared. Our results indicated a good general performance of the model in reconstructing PM_{10} (coefficient of determination equal to 0.79) for most variables. In order to confirm the results of receptor modelling, the origin of the air masses associated with the factors obtained was investigated through the creation of wind polar plots using the source contribution of the factors produced by PMF. In particular, polar plots were produced for each single PMF factor using the "openair" package of R[67], utilizing the conditional probability function (CPF)[78] with an arbitrary threshold set to the 75th percentile.

Microbial DNA extraction, 16S rRNA gene amplification and sequencing.   Microbial DNA extraction was performed on quartz membrane filter using the DNeasy PowerSoil Kit (Qiagen, Hilden, Germany) with the following modifications: the homogenization was performed with a FastPrep instrument (MP Biomedicals, Irvine, CA) at 5.5 movements per sec for 1 min, and the elution step was preceded by a 5-min incubation at 4 °C[79,80]. Extracted DNA samples were quantified with NanoDrop ND-1000 (NanoDrop Technologies, Wilmington, DE) and stored at − 20 °C until further processing. The V3-V4 hypervariable region of the 16S rRNA gene was PCR amplified in a 50-μL final volume containing 25 ng of microbial DNA, 2X KAPA HiFi HotStart ReadyMix (Roche, Basel, Switzerland), and 200 nmol/L of 341F and 785R primers carrying Illumina overhang adapter sequences. The thermal cycle was performed as already described[81] using 30 amplification cycles. PCR products were purified using Agencourt AMPure XP magnetic beads (Beckman Coulter, Brea, CA). Indexed libraries were prepared by limited-cycle PCR with Nextera technology and cleaned-up as described above. Libraries were normalized to 1 nM and pooled. The sample pool was denatured with 0.2 N NaOH and diluted to 6 pM with a 20% PhiX control. Sequencing was performed on an Illumina MiSeq platform using a 2 × 250 bp paired-end protocol, according to the manufacturer's instructions (Illumina, San Diego, CA).

Bioinformatics and statistics.   A pipeline combining PANDAseq[82] and QIIME 2[83] was used to process raw sequences. DADA2[84] was used to bin high-quality reads (min/max length = 350/550 bp) into amplicon sequence variants (ASVs). After taxonomy assignment using the VSEARCH algorithm[85] and the SILVA database (December 2017 release)[86], the sequences assigned to eukaryotes (*i.e.* chloroplasts and mitochondria) or unassigned were discarded. Three different metrics were used to evaluate alpha diversity—Faith's Phylogenetic Diversity (PD whole tree)[87], Chao1 index for microbial richness, and number of observed ASVs—and unweighted UniFrac distance was used for Principal Coordinates Analysis (PCoA). Permutation test with pseudo-F ratio (function "adonis" in the "vegan" package of R), Kruskal–Wallis test or Wilcoxon rank-sum test were used to assess data separation. Kendall correlation test was used to determine associations between the PCoA coordinates of each sample and the factors identified by the PMF analysis. P-values were corrected for multiple testing with the Benjamini–Hochberg method, with a false discovery rate (FDR) ≤ 0.05 considered statistically significant. All statistical analyses and respective figures were produced with the R software[68] using "Made4"[88] and "vegan" (https://cran.r-project.org/web/packages/vegan/index.html) packages. Clustering analysis of family-level AM profiles, filtered for family subject prevalence of at least 20%, based on the SILVA taxonomy assignment, was carried out using hierarchical Ward-linkage clustering based on the Spearman correlation coefficients. We verified that each cluster showed significant correlations between samples within the group (multiple testing using the Benjamini–Hochberg method) and that the clusters were statistically significantly different from each other (permutational MANOVA using the Spearman distance matrix as input, function adonis of the vegan package in R).

Additionally, PANDAseq assembled paired-end reads were also processed with the QIIME1[89] pipeline for Operational Taxonomic Units (OTUs) clustering based on 97% similarity threshold. Taxonomy was then assigned using the SILVA database. OTUs were processed through the R package "MaAsLin2"[90] to determine their association with microbial clusters. Kruskal–Wallis test was used to find OTUs whose relative abundance was significantly different among microbial clusters. The resulting OTUs were taxonomically assigned against the NCBI 16S rRNA database using the BLAST algorithm (https://blast.ncbi.nlm.nih.gov/).

Data availability
Sequence reads were deposited in the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA; BioProject ID PRJNA664273; reviewer link https://dataview.ncbi.nlm.nih.gov/object/PRJNA664273?reviewer=k646phv540av66bt8v1jduj699).

References
1. Tositti, L. Physical and chemical properties of airborne particulate matter. In *Clinical Handbook of Air Pollution-Related Diseases* (eds Capello, F. & Gaddi, A.) 7–32 (Springer, Cham, 2018).

2. Tositti, L. The relationship between health effects and airborne particulate constituents. In *Clinical Handbook of Air Pollution-Related Diseases* (eds Capello, F. & Gaddi, A.) 33–54 (Springer, Cham, 2018).
3. Hopke, P. K. Review of receptor modeling methods for source apportionment. *J. Air Waste Manage* **66**, 237–259 (2016).
4. Fröhlich-Nowoisky, J. *et al.* Bioaerosols in the Earth system: climate, health, and ecosystem interactions. *Atmos. Res.* **182**, 346–376 (2016).
5. Womack, A. M., Bohannan, B. J. & Green, J. L. Biodiversity and biogeography of the atmosphere. *Philos. Trans. R. Soc. B.* **365**, 3645–3653 (2010).
6. Castillo, J. A., Staton, S. J., Taylor, T. J., Herckes, P. & Hayes, M. A. Exploring the feasibility of bioaerosol analysis as a novel fingerprinting technique. *Anal. Bioanal. Chem.* **403**, 15–26 (2012).
7. Hyde, P. & Mahalov, A. Contribution of bioaerosols to airborne particulate matter. *J. Air Waste Manage* **70**, 71–77 (2020).
8. Mescioglu, E. *et al.* Aerosol microbiome over the Mediterranean Sea diversity and abundance. *Atmosphere* **10**, 440 (2019).
9. Després, V. *et al.* Primary biological aerosol particles in the atmosphere: a review. *Tellus B* **64**, 15598 (2012).
10. Mhuireach, G. Á., Betancourt-Román, C. M., Green, J. L. & Johnson, B. R. Spatiotemporal controls on the urban aerobiome. *Front. Ecol. Evol.* **7**, 43 (2019).
11. Veron, F. Ocean spray. *Annu. Rev. Fluid Mech.* **47**, 507–538 (2015).
12. Wilson, T. W. *et al.* A marine biogenic source of atmospheric ice-nucleating particles. *Nature* **525**, 234 (2015).
13. Dommergue, A. *et al.* Methods to investigate the global atmospheric microbiome. *Front. Microbiol.* **10**, 243 (2019).
14. Delort, A. M. & Amato, P. *Microbiology of aerosols.* (John Wiley & Sons, Inc, 2018).
15. DeLeon-Rodriguez, N. *et al.* Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proc. Natl. Acad. Sci. USA* **110**, 2575–2580 (2013).
16. Sesartic, A., Lohmann, U. & Storelvmo, T. Bacteria in the ECHAM5-HAM global climate model. *Atmos. Chem. Phys.* **12**, 8645–8661 (2012).
17. Toprak, E. & Schnaiter, M. Fluorescent biological aerosol particles measured with the Waveband Integrated Bioaerosol Sensor WIBS-4: laboratory tests combined with a one year field study. *Atmos. Chem. Phys.* **13**, 225–243 (2013).
18. Renard, P. *et al.* Cloud microorganisms, an interesting source of biosurfactants. In: *Surfactants and Detergents* (ed. Dutta, A. K.) https://www.intechopen.com/books/surfactants-and-detergents/cloud-microorganisms-an-interesting-source-of-biosurfactants (IntechOpen, UK, 2019).
19. Cao, C. *et al.* Inhalable microorganisms in Beijing's PM2.5 and PM10 pollutants during a severe smog event. *Environ. Sci. Technol.* **48**, 1499–1507 (2014).
20. Huang, R. J. *et al.* High secondary aerosol contribution to particulate pollution during haze events in China. *Nature* **514**, 218–222 (2014).
21. Michaud, J. M. *et al.* Taxon-specific aerosolization of bacteria and viruses in an experimental ocean-atmosphere mesocosm. *Nat. Commun.* **9**, 2017. https://doi.org/10.1038/s41467-018-04409-z (2018).
22. von Schneidemesser, E. *et al.* Chemistry and the linkages between air quality and climate change. *Chem. Rev.* **115**, 3856–3897 (2015).
23. Qin, N. *et al.* Longitudinal survey of microbiome associated with particulate matter in a megacity. *Genome Biol.* **21**, 55 (2020).
24. Bowers, R. M. *et al.* Seasonal variability in bacterial and fungal diversity of the near-surface atmosphere. *Environ. Sci. Technol.* **47**, 12097–12106 (2013).
25. Liu, H. *et al.* The distribution variance of airborne microorganisms in urban and rural environments. *Environ. Pollut.* **247**, 898–906 (2019).
26. Woo, A. C. *et al.* Temporal variation in airborne microbial populations and microbially-derived allergens in a tropical urban landscape. *Atmos. Environ.* **74**, 291–300 (2013).
27. Gandolfi, I. *et al.* Spatio-temporal variability of airborne bacterial communities and their correlation with particulate matter chemical composition across two urban areas. *Appl. Microbiol. Biotechnol.* **99**, 4867–4877 (2015).
28. Innocente, E. *et al.* Influence of seasonality, air mass origin and particulate matter chemical composition on airborne bacterial community structure in the Po Valley, Italy. *Sci. Total Environ.* **593–594**, 677–687 (2017).
29. Li, H. *et al.* Spatial and seasonal variation of the airborne microbiome in a rapidly developing city of China. *Sci. Total Environ.* **665**, 61–68 (2019).
30. Graham, K. E., Prussin, A. J. 2nd., Marr, L. C., Sassoubre, L. M. & Boehm, A. B. Microbial community structure of sea spray aerosols at three California beaches. *FEMS Microbiol. Ecol.* **94**, fiy005. https://doi.org/10.1093/femsec/fiy005 (2018).
31. Tignat-Perrier, R. *et al.* Global airborne microbial communities controlled by surrounding landscapes and wind conditions. *Sci. Rep.* **9**, 14441. https://doi.org/10.1038/s41598-019-51073-4 (2019).
32. Tositti, L. *et al.* Source apportionment of particulate matter in a large city of southeastern Po Valley (Bologna, Italy). *Environ. Sci. Pollut. R.* **21**, 872–890 (2014).
33. Chow, J. C., Lowenthal, D. H., Chen, L. W. A., Wang, X. & Watson, J. G. Mass reconstruction methods for PM 2.5: a review. *Air Qual. Atmos. Health* **8**, 243–263 (2015).
34. Pachon, J. E., Weber, R. J., Zhang, X., Mulholland, J. A. & Russell, A. G. Revising the use of potassium (K) in the source apportionment of PM2.5. *Atmos. Pollut. Res.* **4**, 14–21 (2013).
35. Alastuey, A. *et al.* Contribution of harbour activities to levels of particulate matter in a harbour area: Hada Project-Tarragona Spain. *Atmos. Environ.* **41**, 6366–6378 (2007).
36. Thorpe, A. & Harrison, R. M. Sources and properties of non-exhaust particulate matter from road traffic: a review. *Sci. Total Environ.* **400**, 270–282 (2008).
37. Gietl, J. K., Lawrence, R., Thorpe, A. J. & Harrison, R. M. Identification of brake wear particles and derivation of a quantitative tracer for brake dust at a major road. *Atmos. Environ.* **44**, 141–146 (2010).
38. Schaap, M., Van Loon, M., Ten Brink, H. M., Dentener, F. J. & Builtjes, P. J. H. Secondary inorganic aerosol simulations for Europe with special attention to nitrate. *Atmos. Chem. Phys.* **4**, 857–874 (2004).
39. Pathak, R. K., Wu, W. S. & Wang, T. Summertime PM 2.5 ionic species in four major cities of China: nitrate formation in an ammonia-deficient atmosphere. *Atmos. Chem. Phys.* **9**, 1711–1722 (2009).
40. Rodríguez, S. *et al.* Comparative PM10-PM2.5 source contribution study at rural, urban and industrial sites during PM episodes in Eastern Spain. *Sci. Total Environ.* **328**, 95–113 (2004).
41. Vecchi, R., Marcazzan, G., Valli, G., Ceriani, M. & Antoniazzi, C. The role of atmospheric dispersion in the seasonal variation of PM1 and PM2.5 concentration and composition in the urban area of Milan (Italy). *Atmos. Environ.* **38**, 4437–4446 (2004).
42. Hueglin, C. *et al.* Chemical characterisation of PM2.5, PM10 and coarse particles at urban, near-city and rural sites in Switzerland. *Atmos. Environ.* **39**, 637–651 (2005).
43. Behera, S. N., Sharma, M., Aneja, V. P. & Balasubramanian, R. Ammonia in the atmosphere: a review on emission sources, atmospheric chemistry and deposition on terrestrial bodies. *Environ. Sci. Pollut. Res. Int.* **20**, 8092–8131 (2013).
44. Fioletov, V. *et al.* Multi-source SO2 emission retrievals and consistency of satellite and surface measurements with reported emissions. *Atmos. Chem. Phys.* **17**, 12597–12616 (2017).
45. Jang, H. N. *et al.* Formation of fine particles enriched by V and Ni from heavy oil combustion: anthropogenic sources and drop-tube furnace experiments. *Atmos. Environ.* **41**, 1053–1063 (2007).

46. Becagli, S. *et al.* Evidence for heavy fuel oil combustion aerosols from chemical analyses at the island of Lampedusa: a possible large role of ships emissions in the Mediterranean. *Atmos. Chem. Phys.* **12**, 3479–3492 (2012).
47. Viana, M. *et al.* Impact of maritime transport emissions on coastal air quality in Europe. *Atmos. Environ.* **90**, 96–105 (2014).
48. Yu, J. *et al.* Potassium: a tracer for biomass burning in Beijing. *Aerosol Air Qual. Res.* **18**, 2447–2459 (2018).
49. Grythe, H., Ström, J., Krejčí, R., Quinn, P. K. & Stohl, A. A review of sea-spray aerosol source functions using a large global set of sea salt aerosol concentration measurements. *Atmos. Chem. Phys.* **14**, 1277–1297 (2014).
50. Zhang, J. *et al.* A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. *ISME J.* **9**, 1979–1990 (2015).
51. Lloyd-Price, J., Abu-Ali, G. & Huttenhower, C. The healthy human microbiome. *Genome Med.* **8**, 51. https://doi.org/10.1186/s13073-016-0307-y (2016).
52. Rinninella, E. *et al.* What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* **7**, 14. https://doi.org/10.3390/microorganisms7010014 (2019).
53. Wang, Y. *et al.* Bacterial community diversity associated with different levels of dietary nutrition in the rumen of sheep. *Appl. Microbiol. Biotechnol.* **101**, 3717–3728 (2017).
54. Lee, G. H. *et al.* Genome sequence of *Oscillibacter ruminantium* strain GH1, isolated from rumen of Korean native cattle. *J. Bacteriol.* **194**, 6362. https://doi.org/10.1128/JB.01677-12 (2012).
55. Nordhoff, M. *et al.* *Treponema berlinense* sp. nov. and *Treponema porcinum* sp. nov., novel spirochaetes isolated from porcine faeces. *Int. J. Syst. Evol. Microbiol.* **55**, 1675–1680 (2005).
56. Chen, G. *et al.* Complete genome sequencing provides novel insight into the virulence repertories and phylogenetic position of dry beans pathogen *Curtobacterium flaccumfaciens* pv. *flaccumfaciens*. *Phytopathology* https://doi.org/10.1094/phyto-06-20-0243-r (2020).
57. Feng, W. W. *et al.* *Glutamicibacter halophytocola* sp. nov., an endophytic actinomycete isolated from the roots of a coastal halophyte, *Limonium sinense*. *Int. J. Syst. Evol. Microbiol.* **67**, 1120–1125 (2017).
58. Wang, H. F. *et al.* *Frigoribacterium endophyticum* sp. nov., an endophytic actinobacterium isolated from the root of *Anabasis elatior* (C. A. Mey.) Schischk. *Int. J. Syst. Evol. Microbiol.* **65**, 1207–1212 (2015).
59. Bowers, R. M., McLetchie, S., Knight, R. & Fierer, N. Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. *ISME J.* **5**, 601–612 (2011).
60. Franzetti, A., Gandolfi, I., Gaspari, E., Ambrosini, R. & Bestetti, G. Seasonal variability of bacteria in fine and coarse urban air particulate matter. *Appl. Microbiol. Biotechnol.* **90**, 745–753 (2011).
61. Bertolini, V. *et al.* Temporal variability and effect of environmental variables on airborne bacterial communities in an urban area of Northern Italy. *Appl. Microbiol. Biotechnol.* **97**, 6561–6570 (2013).
62. Köppen, W. Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt. *Geographische Zeitschrift* **6**, 593–611 (1900).
63. Geiger, R. *Landolt-Börnstein – Zahlenwerte und Funktionen aus Physik, Chemie, Astronomie, Geophysik und Technik, alte Serie Vol. 3, Ch. Klassifikation der Klimate nach W. Köppen* 603–607 (Springer, Berlin, 1954).
64. Burlando, M. *et al.* Wind climate analysis in complex terrains. *J. Wind Eng. Ind. Aerod.* **123**, 349–362 (2013).
65. Bentamy, A., Ayina, H. L., Queffeulou, P., Croize-Fillon, D. & Kerbaol, V. Improved near real time surface wind resolution over the Mediterranean Sea. *Ocean Sci.* **3**, 259–271 (2007).
66. Tositti, L. *et al.* Airborne particulate matter biotoxicity estimated by chemometric analysis on bacterial luminescence data. *Sci. Total Environ.* **640**, 1512–1520 (2018).
67. Carslaw, D. & Ropkins, K. Openair—an R package for air quality data analysis. *Environ. Modell. Softw.* **27–28**, 52–61 (2012).
68. R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria; https://www.R-project.org (2019).
69. Piazzalunga, A., Bernardoni, V., Fermo, P., Valli, G. & Vecchi, R. On the effect of water-soluble compounds removal on EC quantification by TOT analysis in aerosol samples. *Atmos. Chem. Phys. Discuss.* **11**, 19851–19880 (2011).
70. Lucarelli, F. *et al.* Is PIXE still a useful technique for the analysis of atmospheric aerosols? The LABEC experience. *X-Ray Spectrom.* **40**, 162–167 (2011).
71. Paatero, P. & Tapper, U. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **5**, 111–126 (1994).
72. Hopke, P. K. A guide to positive matrix factorization. In *EPA Workshop Proceedings Materials from the Workshop on UNMIX and PMF as Applied to PM2, Volume II, Appendix 1B.* https://www3.epa.gov/ttn/amtic/unmixmtg.html (2000).
73. Comero, S., Capitani, L. & Gawlik, B. M. *Positive Matrix Factorisation (PMF)—an introduction to the chemometric evaluation of environmental monitoring data using PMF*. JRC Scientific and Technical Reports. https://doi.org/10.2788/2497 https://op.europa.eu/it/publication-detail/-/publication/11e6a70c-6048-4442-9421-33a332893b8e/language-en/format-PDF/source-120358304 (2009).
74. Belis, C. *et al.* *European guide on air pollution source apportionment with receptor models—Revised version 2019*. https://doi.org/10.2760/439106 https://ec.europa.eu/jrc/en/publication/european-guide-air-pollution-source-apportionment-receptor-models (2019).
75. Masiol, M. *et al.* Hybrid multiple-site mass closure and source apportionment of PM2.5 and aerosol acidity at major cities in the Po Valley. *Sci. Total Environ.* **704**, 135287. https://doi.org/10.1016/j.scitotenv.2019.135287 (2020).
76. Norris, G., Duvall, R., Brown, S. & Bai, S. *Epa positive matrix factorization (PMF) 5.0 fundamentals and user guide prepared for the us environmental protection agency office of research and development.* United States Environmental Protection Agency report. https://www.epa.gov/sites/production/files/2015-02/documents/pmf_5.0_user_guide.pdf (2014).
77. Brown, S. G., Eberly, S., Paatero, P. & Norris, G. A. Methods for estimating uncertainty in PMF solutions: examples with ambient air and water quality data and guidance on reporting PMF results. *Sci. Total Environ.* **518–519**, 626–635 (2015).
78. Ashbaugh, L. L., Malm, W. C. & Sadeh, W. Z. A residence time probability analysis of sulfur concentrations at Grand Canyon National Park. *Atmos. Environ.* **19**, 1263–1270 (1985).
79. Jiang, W. *et al.* Optimized DNA extraction and metagenomic sequencing of airborne microbial communities. *Nat. Protoc.* **10**, 768–779 (2015).
80. Shin, S. K. *et al.* Metagenomic insights into the bioaerosols in the indoor and outdoor environments of childcare facilities. *PLoS ONE* **10**, e0126960. https://doi.org/10.1371/journal.pone.0126960 (2015).
81. Turroni, S. *et al.* Fecal metabolome of the Hadza hunter-gatherers: a host-microbiome integrative view. *Sci. Rep.* **6**, 32826. https://doi.org/10.1038/srep32826 (2016).
82. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinform.* **13**, 31. https://doi.org/10.1186/1471-2105-13-31 (2012).
83. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
84. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
85. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584. https://doi.org/10.7717/peerj.2584 (2016).

86. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Res.* **41**, D590–D596 (2013).
87. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
88. Culhane, A. C., Thioulouse, J., Perrière, G. & Higgins, D. G. MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics* **21**, 2789–2790 (2005).
89. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
90. Mallick, H., Rahnavard, A. & McIver, L. *Maaslin2: Maaslin2*. R package version 1.2.0. http://huttenhower.sph.harvard.edu/maaslin2 (2020).

Acknowledgements

Author contributions

Conceptualization: L.T., M.C.; Data curation: G.P., P.M.; Formal analysis: G.P., P.M., E.Br., S.R.; Investigation: G.P., P.M.; Methodology: E.Br., S.R., L.T., M.C.; Resources: L.T., M.C.; Writing—original draft: G.P., P.M., L.T., M.C.; Writing—review & editing: E.Bi, S.T.

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/s41598-020-80642-1.

Correspondence and requests for materials should be addressed to L.T. or M.C.

Reprints and permissions information is available at www.nature.com/reprints.

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