

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

Giorgia Palladino^{1,3}, Pietro Morozzi², Elena Biagi^{1,3}, Erika Brattich⁴, Silvia Turrone¹, Simone Rampelli¹, Laura Tositti^{2*}, Marco Candela^{1,3*}

¹HolobioME (Holobiont Microbiome and Microbiome Engineering) Unit, Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy

²Department of Chemistry “Giacomo Ciamician”, University of Bologna, Italy

³Fano Marine Center (FMC), Fano, Italy

⁴Department of Physics and Astronomy, University of Bologna, Italy

*Corresponding authors:

Prof. Laura Tositti, Department of Chemistry “Giacomo Ciamician”, University of Bologna, Via Selmi 2, 40126, Bologna, Italy. Phone: +39 051 20 9 9488; mail: laura.tositti@unibo.it

Prof. Marco Candela, Department of Pharmacy and Biotechnology, University of Bologna, Via Belmeloro 6, 40126, Bologna, Italy. Phone: +39 051 20 9 9727; mail: marco.candela@unibo.it

Running title: Airborne microbiome variation in urban settlement

Keywords: air microbiome, particulate matter, meteorology, terrestrial biome, human health

Competing interests statement: The authors declare no competing financial interests.

Abstract

25

26 Here we explore how the chemical composition of particulate matter (PM) and meteorological
 27 conditions combine in shaping the air microbiome in a heavily inhabited industrial urban settlement.
 28 During the observation time, the air microbiome was highly dynamic, fluctuating between different
 29 compositional states, likely resulting from the aerosolization of different microbiomes emission
 30 sources. This dynamic process depends on the combination of local meteorological parameters and
 31 particle emission sources, which may affect the prevalent aerosolized microbiomes. In particular, we
 32 showed that, in the investigated area, industrial emissions and winds blowing from the inlands
 33 combine with an airborne microbiome that includes faecal microbiomes components, suggesting
 34 multiple citizens' exposure to both chemicals and microorganisms of faecal origin, as related to
 35 landscape exploitation and population density. In conclusion, our findings support the need to include
 36 monitoring of the air microbiome compositional structure as a relevant factor for the final assessment
 37 of local air quality.

Introduction

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 PM₁₀ 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, plants and animal faeces [4, 9] - 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 [10, 11]. PBA size spans from a few nanometres up to about a tenth of a millimetre [5], with bacteria-containing particles ranging around 2-4 μm in diameter [12] and accounting for “5–50% of the total number of atmospheric particles $>0.2 \mu\text{m}$ in diameter” [13]. 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 [14]. In the troposphere, the AM concentration ranges from 10^2 to 10^5 cells/ m^3 [12], being the densest in the planetary boundary layer, whose thickness depends on micrometeorological factors and geographic location, with marked daily and seasonal fluctuations [15, 16]. 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 [17], 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 [18-20]. 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 [21]. These factors may likely interfere with AM, shaping its structure and dispersion throughout the troposphere, with direct consequences on the terrestrial biome [22]. However, as far as we know, the current state of knowledge on the connections between AM, atmospheric processes and atmospheric pollution is still fragmentary, especially due to the lack of a cross-cut approach. Therefore, in this work, we explore 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 PM₁₀ 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 behaviour in an area affected by both natural and anthropogenic emission sources, determining the local bacterial microbiome from PBA contained in PM₁₀ and its main features as a function of local meteorological and environmental characteristics.

Materials and methods

Site description.

The PM₁₀ samples treated in this work were collected in Savona, one of the main towns in the Ligurian region (**Figure 1**). The whole region overlooks the Tyrrhenian sea but is entirely occupied by the Appenninic 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) [23-24] 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 [25], while sea-land breeze regimes prevail in the warm season, usually starting from March [26, 27].



Figure 1 – 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 PM₁₀ sampling site with a 3D view of the surroundings (bottom panel) (source: Google Earth; map data: SIO, NOAA, U.S. Navy, NGA, GEBCO, TerraMetrics).

Sample collection and atmospheric parameters.

A total of 184 daily PM₁₀ 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 (Whatman 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 PM₁₀ 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 S1**), i.e. at the same time resolution as the PM₁₀ samples, using the “openair” package [28] of the R software (version 3.6.1; <https://www.r-project.org/>).

Chemical characterisation of the samples.

Chemical characterization of PM₁₀ filters was carried out using several analytical techniques. First, PM₁₀ mass load (μg/m³) was determined by gravimetric analysis. Elemental and organic carbon were determined on quartz membranes by thermal-optical transmittance analysis (TOT), as previously described [29]. 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 (NH₄⁺, K⁺, Mg²⁺, NO₃⁻, SO₄²⁻, Na⁺, Cl⁻, Ca²⁺, 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 [30]. Elemental analysis by ICP-MS was carried out according to the UNI EN 14902, 2005 for PM₁₀ 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.

157 Positive Matrix Factorization (PMF) is an advanced multivariate factor analysis technique widely
 158 used in receptor modelling for the chemometric evaluation and modelling of environmental datasets
 159 [3, 31-36]. PMF allows the identification and quantification of the emissive profile of a receptor site,
 160 i.e. the monitoring site where an air quality station is operated. We applied EPA PMF 5.0 software
 161 [37]. The dataset was checked and re-arranged prior to PMF modelling according to the model criteria
 162 previously described [37] and, after data pre-processing, a concentration matrix of 98 samples \times 25
 163 variables was obtained. After careful evaluation of the input data and uncertainty matrices, an
 164 optimum number of factors was found by analysing the values of Q, a parameter estimating the
 165 goodness of the fit performed [38], and the distribution of residuals. In order to assess the reliability
 166 of the model reconstruction, measured (input data) and reconstructed (modeled) values together with
 167 the distribution of residuals were compared. Our results indicated a good general performance of the
 168 model in reconstructing PM₁₀ (coefficient of determination equal to 0.79) for most variables. In order
 169 to confirm the results of receptor modelling, the origin of the air masses associated with the factors
 170 obtained was investigated through the creation of wind polar plots using the source contribution of
 171 the factors produced by PMF. In particular, polar plots were produced for each single PMF factor
 172 using the “openair” package of R [28], utilizing the conditional probability function (CPF) [39] with
 173 an arbitrary threshold set to the 75th percentile.

174
 175 *Microbial DNA extraction, 16S rRNA gene amplification and sequencing.*

176 Microbial DNA extraction was performed on quartz membrane filter using the DNeasy PowerSoil
 177 Kit (Qiagen, Hilden, Germany) with the following modifications: the homogenization was performed
 178 with a FastPrep instrument (MP Biomedicals, Irvine, CA) at 5.5 movements per sec for 1 min, and
 179 the elution step was preceded by a 5-min incubation at 4°C [40, 41]. Extracted DNA samples were
 180 quantified with NanoDrop ND-1000 (NanoDrop Technologies, Wilmington, DE) and stored at -20°C
 181 until further processing. The V3-V4 hypervariable region of the 16S rRNA gene was PCR amplified
 182 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 [42], 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). Sequence reads were deposited in the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA; BioProject ID XXXXX).

Bioinformatics and statistics.

A pipeline combining PANDAseq [43] and QIIME 2 [44] was used to process raw sequences. DADA2 [45] 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 [46] and the SILVA database (December 2017 release) [47], 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) [48], 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 were performed in R using "Made4" [49] 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, PANDAsseq assembled paired-end reads were also processed with the QIIME1 [50] pipeline for OTUs (Operational Taxonomic Units) clustering based on 97% similarity threshold. Taxonomy was then assigned using the SILVA database. OTUs were processed through the R package “MaAsLin2” [51] 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/>).

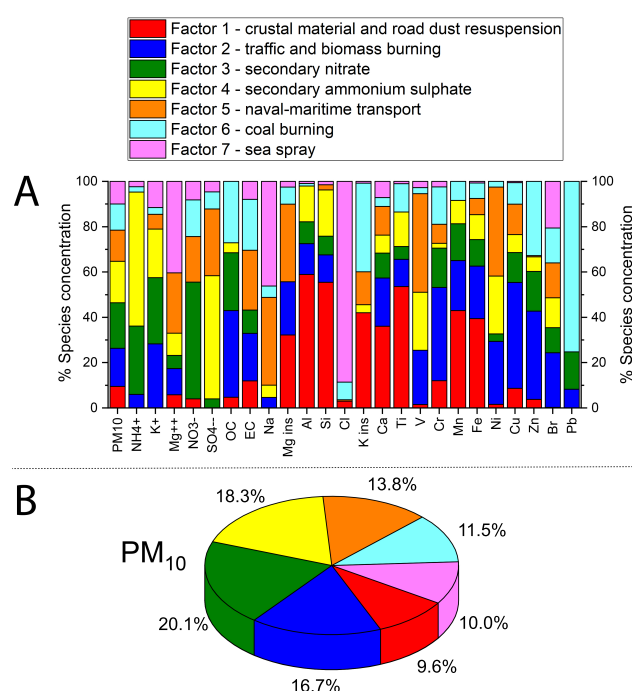
Results

Particulate Matter emission sources and atmospheric parameters.

The PMF model application on PM₁₀ samples resulted in a solution with an optimum number of seven source factors at the receptor site, *i.e.* the station where the PM₁₀ 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 S2**.

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 (**Figure 2A**). Moreover, the percentage contribution of the seven identified sources to the total variable was reported (PM₁₀, **Figure 2B**). 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 [34, 52]. 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 [53]. 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 [54-56]. Therefore, this factor was identified as a combination of “traffic and biomass burning” sources. Factor 3 is mainly associated with NO₃⁻ from gas-to-particle conversion of NO_x (g) in the atmosphere to which traffic and other high-temperature combustion processes may contribute [57, 58]; 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 SO₄²⁻ and NH₄⁺ from gas-to particle reactions, leading to secondary ammonium sulphate [59-61]. 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 SO₂ and the ubiquity of NH₃ (g) [62, 63]. 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 [64-66]. 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 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 [67]. 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 [68].



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

287

288 *AM overall composition.*

289 Next generation sequencing of the V3-V4 hypervariable region of the 16S rRNA gene from the total
290 microbial DNA extracted from PM₁₀ air filters resulted in 98 samples containing more than 1 000
291 reads per samples which were retained for the rest of the study, for a total of 797 781 high-quality
292 sequences with an average of $8\,058 \pm 3\,410$ (mean \pm SD) paired-end reads per sample, binned into 4
293 189 ASVs. According to our data, AM is dominated by the phyla Proteobacteria (mean relative
294 abundance \pm SD = $42.8 \pm 19.4\%$) and Firmicutes ($27.4 \pm 18.9\%$), with Actinobacteria ($14.8 \pm 10.9\%$)
295 and Bacteroidetes ($9.2 \pm 8.6\%$) being subdominant. At the family level, the most represented taxa are
296 *Comamonadaceae* ($6.1 \pm 13.4\%$) and *Sphingomonadaceae* ($4.3 \pm 5.0\%$), both belonging to
297 Proteobacteria. Other represented families are *Ruminococcaceae* ($3.9 \pm 7.6\%$), *Enterobacteriaceae*
298 ($3.7 \pm 5.9\%$), *Clostridiaceae* ($3.6 \pm 6.8\%$), *Bacillaceae* ($3.5 \pm 5.0\%$) and *Flavobacteriaceae* ($3.4 \pm$
299 5.7%). Please see **Supplementary Figure S2** for a graphical representation of the overall
300 compositional structure of AM throughout the entire sampling period.

301 In order to explore connections between the AM structure and seasonality, we compared the levels
302 of AM diversity over the different months (**Figure 3**). Diversity measurements indicated a general
303 trend of microbial richness to decrease from winter to summer, although the differences did not reach
304 statistical significance (Kruskal-Wallis test, FDR corrected p-value > 0.05) (**Figure 3A**). Conversely,
305 the PCoA of unweighted UniFrac distances between the AM compositional profiles showed sample
306 segregation according to the month of sampling (**Figure 3B**) (FDR-corrected permutation test with
307 pseudo-F ratio, p-value = 0.012), meaning that seasonality significantly affects the overall
308 compositional AM structure.

309

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

312 To further explore the overall AM variation across the sampling period, a clustering analysis of the
313 AM compositional profiles was carried out. Hierarchical Ward-linkage clustering based on the

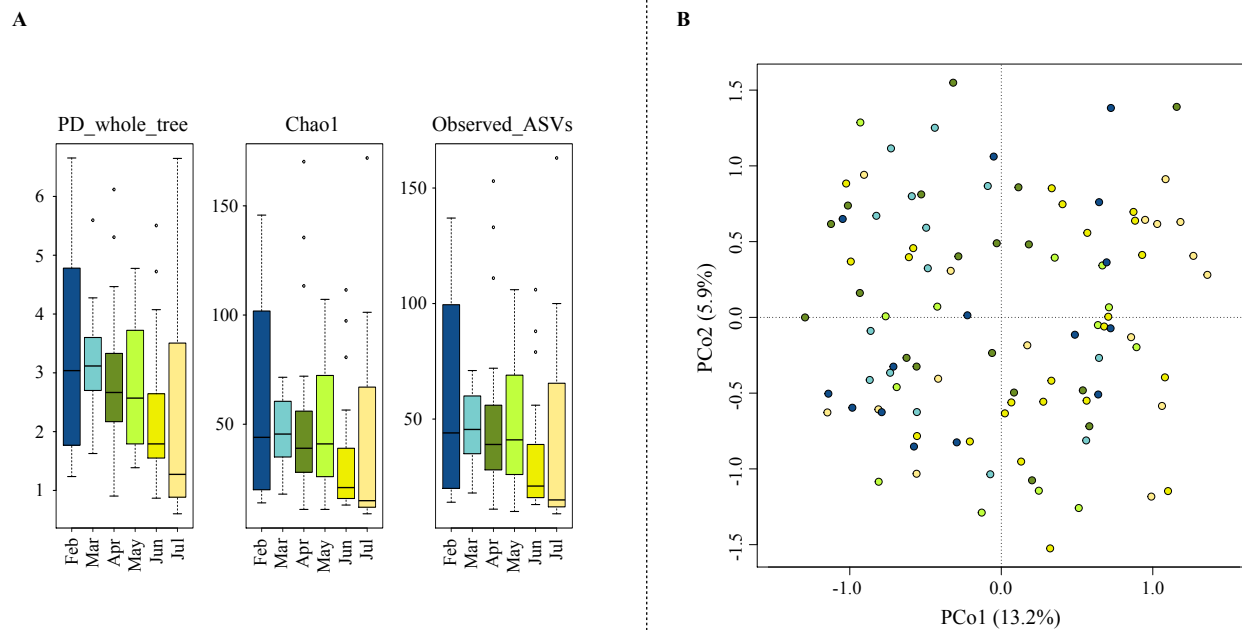
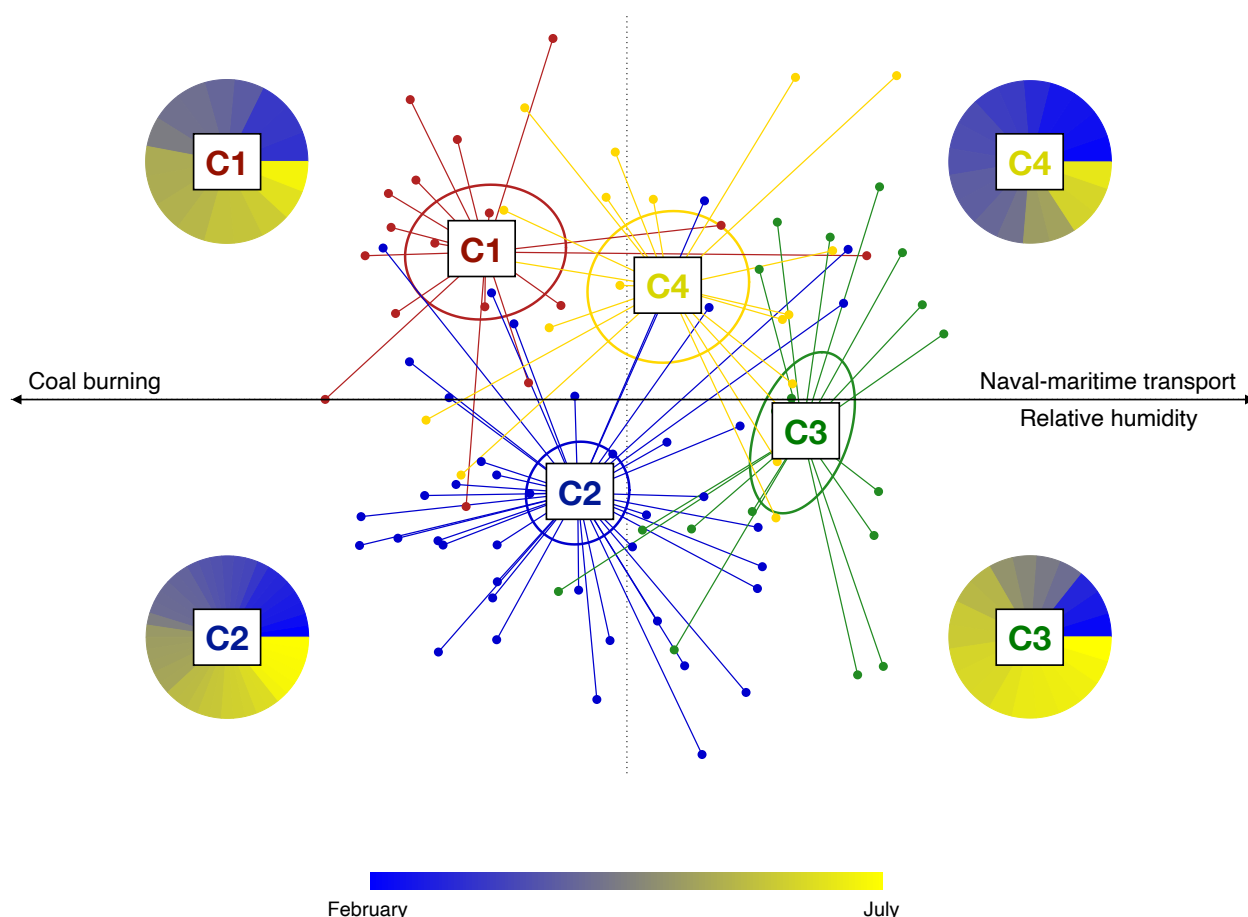


Figure 3 – AM alpha and beta diversity throughout the sampling period. (A) Box-and-whiskers distribution of Faith's Phylogenetic Diversity (PD_whole_tree), Chao1 index for microbial richness and number of observed ASVs, for each month of sampling. The data show a trend towards reduced microbial richness 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.

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) (**Figure 4**). 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 (**Figure 5**). Interestingly, when we searched for correlations between PCoA coordinates and measured meteorological parameters or PMF factors (**Supplementary Tables S1**

whole tree, chao1, and observed ASVs, mean \pm 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).



350

Figure 5 – 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 Figure 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.

362

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*, *Lachnospiraceae*, *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* [69-71], the cattle rumen components *Succinivibrio dextrinosolvens* [72] and

389 *Oscillibacter ruminantium* [73], and the porcine gut microbiome member *Treponema porcinum* [74].
 390 Differently, the cluster C2 is characterized by OTUs assigned to microorganisms isolated from plant
 391 roots and leaves, including *Curtobacterium flaccumfaciens* [75], *Glutamicibacter halophytocola* [76]
 392 and *Frigoribacterium endophyticum* [77], as well as by a specific pattern of environmental bacteria,
 393 from soil, air, and fresh and marine water ecosystems. Similarly, both clusters C3 and C4 are
 394 characterized by a peculiar combination of environmental microorganisms from different sources,
 395 including soil, fresh and marine waters, and airborne microbial ecosystems.

396

397 Discussion

398

399 In order to explore connections between the local air microbiome, atmospheric pollution and
 400 meteorological factors, here we provide a longitudinal survey of the near-ground AM, atmospheric
 401 particulate and atmospheric parameters in Savona, Italy. According to our findings, the local AM
 402 appears dominated by the phyla Proteobacteria, Firmicutes, Actinobacteria and Bacteroidetes, well
 403 matching the general layout of an AM community [4]. The application of the PMF receptor modelling
 404 on the chemical compositional pattern of the PM₁₀ samples collected during the field campaign
 405 allowed the identification of seven emission sources: “crustal material and road dust resuspension”,
 406 “traffic and biomass burning”, “secondary nitrate”, “secondary ammonium sulphate”, “naval-
 407 maritime transport”, “coal burning” and “sea spray”. Each source factor was subsequently subjected
 408 to anemological analysis based on polar plots, allowing each emission source to be associated with
 409 the corresponding wind direction to which the receptor site is downwind. Specifically, emission
 410 sources as “crustal material and road dust resuspension”, “secondary nitrate”, “secondary ammonium
 411 sulphate” and “coal burning” were associated with winds blowing from the inland toward the
 412 sampling site, intercepting traffic and industrial particulate sources. Conversely, emission sources
 413 such as “naval-maritime transport” and “sea spray” were associated with a sea breeze, supporting a

414 marine origin for both. Finally, the “traffic and biomass burning” emission source mostly showed a
415 local origin.

416 When we explored the AM structure variation during the observation period, we were able to identify
417 four distinct clusters of samples, named C1 to C4. Interestingly, the four clusters were associated with
418 a peculiar combination of seasonality, meteorological variables and emission sources. In particular,
419 the AM cluster C1 was associated with the “coal burning” emission source, suggesting not actually
420 the industrial facility as a microbiome source, but rather the influence of an air mass whose transport
421 over a given district harvests chemical and microbiological components along the same tropospheric
422 path. Instead, the cluster C3, most represented in the warm period, probably has a marine origin due
423 to its association with the “naval-marine transport” emission source and high relative humidity.
424 Finally, the clusters C2 and C4 did not show any specific association with the aerosol sources assessed
425 by PMF, even if they showed a different seasonal behaviour, with C4 being more represented in the
426 cold period.

427 The four AM clusters revealed a distinct, well-defined compositional structure, each being enriched
428 with a specific set of microbial families and OTUs. The specificity of each bacterial profile de facto
429 serves as a microbiological fingerprint, allowing to single out the probable microbiome sources
430 characterizing each cluster that, similar to what occurs to abiotic particles, allow to trace back the
431 origin of the air mass. In particular, the clusters C3 and C4 substantially reflect interconnected
432 environmental microbiomes, encompassing a specific combination of microorganisms from soil
433 resuspension, as well as from marine and fresh waters (possibly from rivers and streams flowing into
434 the Ligurian Sea) and from the air. C2 cluster reveals the plant microbiome as an additional source,
435 showing a further combination of plant-associated and environmental microorganisms, due to the
436 contact of air masses over a vegetation landscape. Interestingly, the feasibility of air mass tracing also
437 using bacterial species clearly emerges when we observe in detail the compositional structure of C1.
438 This is in fact the only AM cluster carrying a recognizable pool of bacterial moieties of faecal origin,
439 which are consistently part of the animal gut microbiome, suggesting not only a well-defined origin

440 but also the potential use of this information in the assessment of microbiological impacts. It should
 441 be noted that in the area upwind C1 no sewage treatment plant as a possible source of faecal
 442 microbiome was present at the time of sampling. However, the area is very densely populated and
 443 forested areas populated by local fauna are closely found within a few kilometres.

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

457 In conclusion, our results suggest that, in an urban settlement, air pollution may increase the
 458 proportion of aerosolized faecal microorganisms in the atmosphere, ultimately increasing citizens'
 459 exposure to faecal microbes. Similar results have recently been obtained by exploring AM in Beijing
 460 over 6 months [22]. Our findings strengthen the importance of including the monitoring of the AM
 461 compositional structure as a determinant factor in the currently used air quality indexes. Indeed, in
 462 urban areas, the possible increased exposure to faecal-associated microbiomes as a result of
 463 increasing pollution can pose a possible threat to human health, particularly in regions with high-
 464 intensity animal farming, due to the inherent propensity of opportunistic pathogens to aerosolize.

465

Acknowledgments

The authors wish to acknowledge Google Earth and the data providers SIO, NOAA, U.S. Navy, NGA, GEBCO and TerraMetrics for providing maps and 3D views of the sampling site.

This study represents partial fulfilment of the requirements for the PhD thesis of G. Palladino at the PhD course of Innovative Technologies and Sustainable Use of Mediterranean Sea Fishery and Biological Resources (FishMed – University of Bologna, Italy).

Competing Interests Statement

The authors declare no competing financial interests.

References

1. Tositti L. Physical and Chemical Properties of Airborne Particulate Matter. In: Capello F, Gaddi A (eds). *Clinical Handbook of Air Pollution-Related Diseases*. Springer, Cham, Switzerland, 2018, pp 7-32. .
2. Tositti L. The Relationship Between Health Effects and Airborne Particulate Constituents. In: Capello F, Gaddi A (eds). *Clinical Handbook of Air Pollution-Related Diseases*. Springer, Cham, Switzerland, 2018, pp 33-54.
3. Hopke PK. Review of receptor modeling methods for source apportionment. *J Air Waste Manage* 2016; 66(3): 237-259.
4. Fröhlich-Nowoisky J, Kampf CJ, Weber B, Huffman JA, Pöhlker C, Andreae MO *et al.* Bioaerosols in the Earth system: Climate, health, and ecosystem interactions. *Atmos Res* 2016; 182: 346-376.
5. Castillo J A, Staton SJ, Taylor TJ, Herckes P, Hayes MA. Exploring the feasibility of bioaerosol analysis as a novel fingerprinting technique. *Anal Bioanal Chem* 2012; 403(1): 15-26.

6. Womack AM, Bohannon BJ, Green JL. Biodiversity and biogeography of the atmosphere. *Philos T R Soc B* 2010; 365(1558): 3645-3653.
7. Hyde P, Mahalov A. Contribution of bioaerosols to airborne particulate matter. *J Air Waste Manage* 2020; 70(1): 71-77.
8. Mescioglou E, Rahav E, Belkin N, Xian P, Eizenga JM, Vichik A *et al.* Aerosol microbiome over the mediterranean sea diversity and abundance. *Atmosphere* 2019; 10(8): 440.
9. Després V, Huffman JA, Burrows SM, Hoose C, Safatov A, Buryak G *et al.* Primary biological aerosol particles in the atmosphere: a review. *Tellus B* 2012; 64(1): 15598.
10. Wilson TW, Ladino LA, Alpert PA, Breckels MN, Brooks IM, Browse J *et al.* A marine biogenic source of atmospheric ice-nucleating particles. *Nature* 2015; 525(7568): 234.
11. Veron F. Ocean spray. *Annu Rev Fluid Mech* 2015; 47: 507-538.
12. Dommergue A, Amato P, Tignat-Perrier R, Magand O, Thollot A, Joly M *et al.* Methods to investigate the global atmospheric microbiome. *Front Microbiol* 2019; 10: 243.
13. Delort AM, Amato P (eds). *Microbiology of aerosols*. John Wiley & Sons, Inc: Hoboken, NJ, USA, 2018.
14. DeLeon-Rodriguez N, Lathem TL, Rodriguez-R LM, Barazesh JM, Anderson BE, Beyersdorf AJ *et al.* Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proc Natl Acad Sci U S A* 2013; 110(7): 2575-2580.
15. 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* 2013; 13(1): 225–243.
16. Sesartic A, Lohmann U, Storelvmo T. Bacteria in the ECHAM5-HAM global climate model. *Atmos Chem Phys* 2012; 12(18): 8645-8661.

17. Renard P, Canet I, Sancelme M, Matulova M, Uhliarikova I, Eyheraguibel B *et al.* Cloud Microorganisms, an Interesting Source of Biosurfactants. In: Dutta AK (ed). *Surfactants and Detergents*. IntechOpen: London, UK, 2019.
18. Huang RJ, Zhang Y, Bozzetti C, Ho KF, Cao JJ, Han Y *et al.* High secondary aerosol contribution to particulate pollution during haze events in China. *Nature* 2014; 514(7521): 218-222.
19. Cao C, Jiang W, Wang B, Fang J, Lang J, Tian G *et al.* Inhalable microorganisms in Beijing's PM_{2.5} and PM₁₀ pollutants during a severe smog event. *Environ Sci Technol* 2014; 48(3): 1499-507.
20. Michaud JM, Thompson LR, Kaul D, Espinoza JL, Richter RA, Xu ZZ *et al.* Taxon-specific aerosolization of bacteria and viruses in an experimental ocean-atmosphere mesocosm. *Nat Commun* 2018; 9(1): 2017.
21. von Schneidmesser E, Monks PS, Allan JD, Bruhwiler L, Forster P, Fowler D *et al.* Chemistry and the Linkages between Air Quality and Climate Change. *Chem Rev* 2015; 115(10): 3856-97.
22. Qin N, Liang P, Wu C, Wang G, Xu Q, Xiong X *et al.* Longitudinal survey of microbiome associated with particulate matter in a megacity. *Genome Biol* 2020; 21(1): 55.
23. Köppen W. Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt. *Geographische Zeitschrift* 1900; 6(11): 593-611.
24. 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. *Springer*, Berlin, Germany, 1954, pp 603–607.
25. Burlando M, De Gaetano P, Pizzo M, Repetto MP, Solari G, Tizzi M. Wind climate analysis in complex terrains. *J Wind Eng Ind Aerod* 2013; 123: 349-362.
26. Bentamy A, Ayina HL, Queffeuilou P, Croize-Fillon D, Kerbaol V. Improved near real time surface wind resolution over the Mediterranean Sea. *Ocean Sci* 2007; 3(2): 259-271.

27. Tositti L, Brattich E, Parmeggiani S, Bolelli L, Ferri E, Girotti S. Airborne particulate matter biotoxicity estimated by chemometric analysis on bacterial luminescence data. *Sci Total Environ* 2018; 640: 1512-1520.
28. Carslaw D, Ropkins K. openair --- An R package for air quality data analysis. *Environ Modell Softw* 2012; 27—28: 52–61.
29. 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* 2011; 11(7).
30. Lucarelli F, Nava S, Calzolari G, Chiari M, Udisti R, Marino F. Is PIXE still a useful technique for the analysis of atmospheric aerosols? The LABEC experience. *X-Ray Spectrometry* 2011; 40(3): 162-167.
31. Paatero P, Tapper U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 1994; 5(2): 111-126.
32. Hopke PK. A guide to positive matrix factorization. In EPA Workshop Proceedings Materials from the *Workshop on UNMIX and PMF as Applied to PM2*, Vol. 5, 2000, p. 600.
33. 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*, ISBN 978-92-79-12954-4, 2009.
https://publications.jrc.ec.europa.eu/repository/bitstream/JRC52754/reqno_jrc52754_final_pdf_version%5B1%5D.pdf
34. Tositti L, Brattich E, Masiol M, Baldacci D, Ceccato D, Parmeggiani S *et al.* Source apportionment of particulate matter in a large city of southeastern Po Valley (Bologna, Italy). *Environ Sci Pollut R* 2014; 21(2): 872-890.
35. Belis C, Favez O, Mircea M, Diapouli E, Manousakas M-I, Vratolis S *et al.* *European guide on air pollution source apportionment with receptor models - Revised version 2019.*
<https://doi.org/10.2760/439106>

36. Masiol M, Squizzato S, Formenton G, Khan MB, Hopke PK, Nenes A *et al.* Hybrid multiple-site mass closure and source apportionment of PM_{2.5} and aerosol acidity at major cities in the Po Valley. *Sci Total Environ* 2020; 704: 135287.
37. 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, Washington, DC. https://www.epa.gov/sites/production/files/2015-02/documents/pmf_5.0_user_guide.pdf.
38. Brown SG, Eberly S, Paatero P, Norris GA. Methods for estimating uncertainty in PMF solutions: examples with ambient air and water quality data and guidance on reporting PMF results. *Sci Total Environ* 2015; 518-519: 626-35. 5.
39. Ashbaugh LL, Malm WC, Sadeh WZ. A residence time probability analysis of sulfur concentrations at Grand Canyon National Park. *Atmos Environ* 1985; 19(8): 1263-1270.
40. Jiang W, Liang P, Wang B, Fang J, Lang J, Tian G *et al.* Optimized DNA extraction and metagenomic sequencing of airborne microbial communities. *Nat Protoc* 2015; 10(5): 768-79.
41. Shin SK, Kim J, Ha SM, Oh HS, Chun J, Sohn J *et al.* Metagenomic insights into the bioaerosols in the indoor and outdoor environments of childcare facilities. *PLoS One* 2015; 10(5): e0126960.
42. Turrone S, Fiori J, Rampelli S, Schnorr SL, Consolandi C, Barone M *et al.* Fecal metabolome of the Hadza hunter-gatherers: a host-microbiome integrative view. *Sci Rep* 2016; 6: 32826.
43. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 2012; 13(1): 31.
44. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019; 37(8): 852-857.

45. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016; 13(7): 581-3.
46. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 2016; 4: e2584.
47. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013; 41(Database issue): D590-6.
48. Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Conserv* 1992; 61(1): 1-10.
49. Culhane AC, Thioulouse J, Perrière G, Higgins DG. MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics* 2005; 21(11): 2789-2790.
50. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010; 7(5): 335-6.
51. Mallick H, Rahnavard A, McIver L. 2020. Maaslin2: Maaslin2. R package version 1.2.0, <http://huttenhower.sph.harvard.edu/maaslin2>.
52. Chow JC, Lowenthal DH, Chen LWA, Wang X, Watson JG. Mass reconstruction methods for PM 2.5: a review. *Air Qual Atmos Health* 2015; 8(3): 243-263.
53. Pachon JE, Weber RJ, Zhang X, Mulholland JA, Russell AG. Revising the use of potassium (K) in the source apportionment of PM2.5. *Atmos Pollut Res* 2013; 4(1): 14-21.
54. Alastuey A, Moreno N, Querol X, Viana M, Artiñano B, Luaces JA *et al.* Contribution of harbour activities to levels of particulate matter in a harbour area: Hada Project-Tarragona Spain. *Atmos Environ* 2007; 41(30): 6366-6378.
55. Thorpe A, Harrison RM. Sources and properties of non-exhaust particulate matter from road traffic: a review. *Sci Total Environ* 2008; 400(1-3): 270-282.

56. Gietl JK, Lawrence R, Thorpe AJ, Harrison RM. Identification of brake wear particles and derivation of a quantitative tracer for brake dust at a major road. *Atmos Environ* 2010; 44(2): 141-146.
57. Pathak RK, Wu WS, Wang T. Summertime PM 2.5 ionic species in four major cities of China: nitrate formation in an ammonia-deficient atmosphere. *Atmos Chem Phys* 2009; 9: 1711–1722,
58. Schaap M, Van Loon M, Ten Brink HM, Dentener FJ, Builtjes PJH. Secondary inorganic aerosol simulations for Europe with special attention to nitrate. *Atmos Chem Phys* 2004; 4: 857–874,
59. Hueglin C, Gehrig R, Baltensperger U, Gysel M, Monn C, Vonmont H. Chemical characterisation of PM_{2.5}, PM₁₀ and coarse particles at urban, near-city and rural sites in Switzerland. *Atmos Environ* 2005; 39(4): 637-651.
60. Rodríguez S, Querol X, Alastuey A, Viana MM, Alarcón M, Mantilla E *et al.* Comparative PM₁₀-PM_{2.5} source contribution study at rural, urban and industrial sites during PM episodes in Eastern Spain. *Sci Total Environ* 2004; 328(1-3): 95-113.
61. Vecchi R, Marcazzan G, Valli G, Ceriani M, Antoniazzi C. The role of atmospheric dispersion in the seasonal variation of PM₁ and PM_{2.5} concentration and composition in the urban area of Milan (Italy). *Atmos Environ* 2004; 38(27): 4437-4446.
62. Behera SN, Sharma M, Aneja VP, Balasubramanian R. Ammonia in the atmosphere: a review on emission sources, atmospheric chemistry and deposition on terrestrial bodies. *Environ Sci Pollut Res Int* 2013; 20(11): 8092-131.
63. Fioletov V, McLinden CA, Kharol SK, Krotkov NA, Li C, Joiner J *et al.* Multi-source SO₂ emission retrievals and consistency of satellite and surface measurements with reported emissions, *Atmos Chem Phys* 2017; 17: 12597–12616.

64. Jang HN, Seo YC, Lee JH, Hwang KW, Yoo JI, Sok CH *et al.* Formation of fine particles enriched by V and Ni from heavy oil combustion: Anthropogenic sources and drop-tube furnace experiments. *Atmos Environ* 2007; 41(5): 1053-1063.
65. Becagli S, Sferlazzo DM, Pace G, Di Sarra A, Bommarito C, Calzolari G *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* 2012; 12(7): 3479.
66. Viana M, Hammingh P, Colette A, Querol X, Degraeuwe B, de Vlieger I *et al.* Impact of maritime transport emissions on coastal air quality in Europe. *Atmos Environ* 2014; 90: 96-105.
67. Yu J, Yan C, Liu Y, Li X, Zhou T, Zheng M. Potassium: A tracer for biomass burning in Beijing. *Aerosol Air Qual. Res* 2018; 18(9): 2447-2459.
68. Grythe H, Ström J, Krejčí R, Quinn PK, Stohl A. A review of sea-spray aerosol source functions using a large global set of sea salt aerosol concentration measurements. *Atmos Chem Phys* 2014; 14: 1277–1297.
69. Zhang J, Guo Z, Xue Z, Sun Z, Zhang M, Wang L *et al.* A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. *ISME J* 2015; 9(9): 1979-90.
70. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. *Genome Med* 2016; 8(1): 51.
71. Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A *et al.* What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms* 2019; 7(1): 14.
72. Wang Y, Cao P, Wang L, Zhao Z, Chen Y, Yang Y. Bacterial community diversity associated with different levels of dietary nutrition in the rumen of sheep. *Appl Microbiol Biotechnol* 2017; 101(9): 3717-3728.

73. Lee GH, Kumar S, Lee JH, Chang DH, Kim DS, Choi SH *et al.* Genome sequence of *Oscillibacter ruminantium* strain GH1, isolated from rumen of Korean native cattle. *J Bacteriol* 2012; 194(22): 6362.
74. Nordhoff M, Taras D, Macha M, Tedin K, Busse HJ, Wieler LH. *Treponema berlinense* sp. nov. and *Treponema porcinum* sp. nov., novel spirochaetes isolated from porcine faeces. *Int J Syst Evol Microbiol* 2005; 55(Pt 4): 1675-1680.
75. Chen G, Khojasteh M, Taheri-Dehkordi A, Taghavi SM, Rahimi T, Osdaghi E. Complete Genome Sequencing Provides Novel Insight into the Virulence Repertoires and Phylogenetic Position of Dry Beans Pathogen *Curtobacterium flaccumfaciens* pv. *flaccumfaciens*. *Phytopathology* 2020; 10.1094/PHYTO-06-20-0243-R.
76. Feng WW, Wang TT, Bai JL, Ding P, Xing K, Jiang JH *et al.* *Glutamicibacter halophytocola* sp. nov., an endophytic actinomycete isolated from the roots of a coastal halophyte, *Limonium sinense*. *Int J Syst Evol Microbiol* 2017; 67(5): 1120-1125.
77. Wang HF, Zhang YG, Chen JY, 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*. 2015;65(Pt 4):1207-1212. doi:10.1099/ijs.0.000081

Supplementary Information Description

Supplementary Table S1 (2 pages) – Meteorological parameters during the PM sampling period. The first column reports the sample ID, while the second indicates the sampling date. The meteorological parameters taken into account are temperature (T, °C), relative humidity (RH, %), pressure (P, mbar), rainfall (Rain, mm), wind speed (ws, m/s, and wind direction (wd, °). All values were taken every 30 min and averaged on a daily basis.

692 **Supplementary Table S2 (2 pages) – Normalized contributions per sample of the seven factors**
 693 **resolved by PMF analysis.** The first column reports the sample ID. All the other columns represent
 694 the contribution of each factor identified by PMF on the corresponding sample.

695 **Supplementary Table S3 (provided as Excel file) – Characteristics of the OTUs accounting for**
 696 **the compositional specificity of the four AM clusters.** For each OTU, the following information is
 697 given: unique OTUs ID, taxonomy as assigned with SILVA database, the cluster/s to which each
 698 OTU is significantly correlated (i.e. the cluster/s in which the given OTU is significantly more
 699 represented), the BLAST best hit resulting from blasting OTU fasta sequences against the NCBI 16S
 700 rRNA sequence database, the percentage of identity (ID (%)) and coverage (coverage (%)) between
 701 the OTU sequences and the corresponding best hit, and the isolation source of each best hit as reported
 702 in the GenBank database.

703 **Supplementary Figure S1 – Association between the factors obtained by PMF analysis and the**
 704 **wind direction and intensity.** Polar plots of the seven factors obtained by the PMF model. ws, wind
 705 speed; CPF, conditional probability function.

706 **Supplementary Figure S2 - AM overall composition.** Pie charts summarizing the microbiota
 707 composition of air filter samples at phylum (A) and family (B) level. Only phyla with relative
 708 abundance >1.5% in at least 10 samples and families with relative abundance >3% in at least 10
 709 samples are shown.

710 **Supplementary Figure S3 – AM bacterial families differentially represented among the four**
 711 **microbial clusters.** Box plots showing the bacterial families whose relative abundance is
 712 significantly differently distributed among the microbial clusters C1-C4 (Kruskal-Wallis test, FDR-
 713 corrected p-value $\leq 0.05^*$, p-value $\leq 0.01^{**}$ and p-value $\leq 0.001^{***}$). The central box represents the
 714 distance between the 25th and 75th percentiles. The median is marked with a black line. Whiskers
 715 identify the 10th and 90th percentiles.

716 **Supplementary Figure S4 (2 pages) – AM-associated OTUs showing different distribution**
 717 **across microbial clusters.** Box plots showing the OTUs whose relative abundance is significantly

718 differently distributed among the four microbial clusters C1-C4 (Kruskal-Wallis test, FDR-corrected
 719 $p\text{-value} \leq 0.05^*$, $p\text{-value} \leq 0.01^{**}$ and $p\text{-value} \leq 0.001^{***}$). The central box represents the distance
 720 between the 25th and 75th percentiles. The median is marked with a black line. Whiskers identify the
 721 10th and 90th percentiles. unc., unclassified; amb., ambiguous.