Phyllosphere bacterial communities in urban green areas throughout Europe relate to urban intensity

Muyshondt Babette[1,*], Wuyts Karen[1], Van Mensel Anskje[1], Smets Wenke[1], Lebeer Sarah[1], Aleixo Cristiana[2], Alós Ortí Marta[3], Casanelles-Abella Joan[4,5], Chiron François[6], Giacomo Puglielli[3], Laanisto Lauri[3], Moretti Marco[4], Niinemets Ülo[7], Pinho Pedro[2], Tryjanowski Piotr[8], Woszczyło Patrycja[8], Samson Roeland[1]

Key words: urban trees, phyllosphere, land use, microbial communities, urbanisation, diversity

[1] Environmental Ecology and Applied Microbiology (ENdEMIC), Department of Bioscience Engineering, University of Antwerp, Antwerp, Belgium

[2] Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal

[3] Chair of Biodiversity and Nature Tourism, Estonian University of Life Sciences, Tartu, Estonia

[4] Biodiversity and Conservation Biology, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

[5] Landscape Ecology, Institute of Terrestrial Ecosystems, ETH Zurich, Zurich, Switzerland

[6] Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, 91405, Orsay, France

[7] Chair of Crop Science and Plant Biology, Estonian University of Life Sciences, Tartu, Estonia

[8] Department of Zoology, Poznan University of Life Sciences, Poland

[*]Corresponding Author: ENdEMIC, Department of Bioscience engineering, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium. Telephone: +3232653574 Email: babette.muyshondt@uantwerpen.be

Abstract

The phyllosphere harbours a diverse and specific bacterial community, which influences plant health and ecosystem functioning. In this study, we investigated the impact of urban green areas connectivity and size on the composition and diversity of phyllosphere bacterial communities. Hereto, we evaluated the diversity and composition of phyllosphere bacterial communities of 233 *Platanus* x *acerifolia* and *Acer pseudoplatanus* trees in 77 urban green areas throughout six European cities. The community composition and diversity significantly differed between cities but only to a limited extent between tree species. We could show that urban intensity correlated significantly with the community composition of phyllosphere bacteria. In particular, a significant correlation was found between the relative abundances for 29 out of the 50 most abundant families and the urban intensity: the abundances of classic phyllosphere families, such as *Acetobacteraceae*, *Planctomycetes* and *Beijerinkiaceae*, decreased with urban intensity (i.e. more abundant in areas with more green, lower air pollution and lower temperature), while those related to human activities, such as *Enterobacteriaceae* and *Bacillaceae,* increased with urban intensity. The results of this study suggest that phyllosphere bacterial communities in European cities are associated with urban intensity and that effect is mediated by several combined stress factors.

Introduction

Our planet is rapidly urbanizing. This process is associated with vast land use changes, which in turn drive typical urban stressors, such as microclimate change and pollution, and create unique living conditions (Grimm *et al*. 2008). Urban green areas (UGAs), such as parks and forests, are often the most visible representations of nature in cities. UGAs are vital for supporting urban biodiversity (Aronson *et al*. 2014; Beninde *et al*. 2015; Ives *et al.* 2016) and provide important contributions to people in the form of ecosystem services (e.g. air pollution mitigation, microclimate regulation and recreation) (Alberti 2005; Diaz *et al*. 2018). However, the effectiveness of UGAs to deliver ecosystem services depends on several variables: patch size, patch shape, connectivity, land-use history, human population density in the surrounding matrix and management activities (Aronson *et al.* 2017). The outcomes of urbanization are heterogeneous and the urban biodiversity patterns and ecosystem processes vary across scales (e.g. when comparing different cities) (Aronson *et al.* 2014; Uchida *et al.* 2021). Urban stressors influence the provisioning of ecosystem services in general mediated by biodiversity (Emmet *et al*. 2017). The relationship between biodiversity and ecosystem service delivery has been extensively studied before, however, most studies focus on macro taxa and less on microorganisms (Hortal *et al.* 2015; Pinho *et al*. 2021). This is an important limitation since microorganisms are very abundant and key for ecosystem health and functioning, e.g. through plant health and performance (Knapp *et al*. 2021; Knief *et al.* 2012). Phyllosphere bacteria in specific play an important role in ecosystems because of their role in the nitrogen (Knief *et al*. 2012; Moyes *et al*. 2016) and carbon cycle (Jo *et al*. 2015) directly, and indirectly through their effects on host plant health and productivity (Laforest-Lapointe *et al*. 2017; Legein *et al.* 2020; Wei *et al.* 2017; Weyens *et al.* 2015).

The composition and diversity of phyllosphere bacterial communities have been shown to be influenced by both biotic and abiotic factors, including the host plant species (Laforest-Lapointe *et al*. 2016; Redford *et al.* 2010) and environmental factors such as temperature (Aydogan *et al.* 2018), solar radiation (e.g. Truchado *et al*. 2019), atmospheric pollutants (e.g. Pan *et al*. 2019; Wuyts *et al.* 2020) and fertilizers (e.g. McGarvey *et al.* 2019). Due to their potential impact on plant performance and health, ecosystem functioning, air purification and the development of a healthy human immune system (Laforest-Lapointe *et al.* 2017; Vacher *et al*. 2016), the diversity and composition of phyllosphere bacterial communities are of vital importance. However, insights in how phyllosphere bacterial communities of urban trees are influenced by urban stressors are scarce. A limited number of studies found that the composition of phyllosphere bacterial communities of trees in urban landscapes differs from those in rural settings (Brighigna *et al.* 2000; Laforest-Lapointe *et al.* 2017; Smets *et al.* 2016). According to Smets *et al.* (2016) and Wuyts *et al.* (2020), the composition of phyllosphere bacterial communities is significantly influenced by the level of exposure to (traffic-derived) air pollution, and particulate matter in particular. Moreover, the latter study also showed that the alpha diversity of phyllosphere bacterial communities of urban trees decreases and core taxa increase in relative abundance as the amount of green infrastructure in the tree's surroundings increases (Wuyts *et al.* 2020). However, these studies are limited to one geographic location. Moreover, most of these studies considered a restricted number of urban intensity classes, i.e. mainly urban versus rural areas, instead of a gradient in urbanisation and did not evaluate the role of separate urban stressors herein (with the exception of Wuyts *et al*. 2020). To our knowledge, none of the studies on urban phyllosphere bacteria investigated the variation in phyllosphere bacterial communities of urban trees over a large geographical extent and evaluated the influence of the characteristics of the UGAs where a tree is located in, such as their connectivity and size, on the composition and diversity of urban phyllosphere bacterial communities.

In this study, we aimed to assess the impact of urbanization on phyllosphere bacteria communities on a broad scale, and for that we determined the composition and diversity along a fragmentation

and urbanization gradient in cities across Europe. More specifically, we investigated the relationship of urban phyllosphere bacterial community composition and diversity with different urban variables including particulate matter pollution, microclimate and landscape characteristics such as green infrastructure cover in the surroundings and green area size. To do so, we sampled the phyllosphere bacterial communities of *Platanus x acerifolia* and *Acer pseudoplatanus* in 77 UGAs distributed over six European cities and determined their composition and diversity using 16S rRNA gene sequencing.

Materials and methods

Sampling collection

This study was conducted on leaves sampled from trees in six cities throughout Europe in the summer of 2018. Two sampling campaigns were undertaken: the first one, campaign 1, included sampling in Antwerp (Belgium), Tartu (Estonia), Poznan (Poland), Paris (France) and Zurich (Switzerland) from 15 June until 27 July 2018 and the second, campaign 2, included sampling in Antwerp (Belgium), Tartu (Estonia), Poznan (Poland) and Almada (Portugal) from 18 until 28 September 2018 (see Table 1). In each city, 12 UGAs were selected, aiming to maximise the variation according to two independent gradients: i) the size of the UGAs and ii) their structural connectivity with other UGAs, using the Proximity index. Pinho *et al.* (2021) provides a more detailed explanation of the selection process.

In each of these selected UGAs, three *P. x acerifolia* and/or *Acer pseudoplatanus* trees were chosen. Table 2 presents the number of samples in each campaign, per city and per tree species sampled. The locations of the sampled trees are mapped in figure A. These two tree species were chosen because they are common throughout Europe, easily recognisable and usually out of reach for people to touch, which minimizes contamination from the human-associated bacteria. When possible, preference was given to trees located at the centroid of the UGAs. From each tree separately, two types of samples were taken at the south side of the tree. One sample was taken for microbiological analysis and consisted of two randomly selected leaves that were put into a sterile 50 mL conical tube (Falcon, VWR). Gloves and scissors sterilized on site with 70% ethanol were used for sampling. The other sample was used for biomagnetic analysis and consisted of five leaves put into a paper envelope. Sampled leaves were fully-developed, healthy and undamaged (no visible moulds, disease or herbivorous damage). After sampling, the tubes were kept cool on ice until processing in the lab. Samples from Paris, Poznan, Tartu, Almada and Zurich were shipped to the EUREC-Air lab in Antwerp in insulating boxes with ice packs. The samples were processed immediately after arrival.

Microbiological analysis

In order to extract the bacteria from the leaves within 48 hours after sampling, the leaves in the tube were washed with 5 mL leaf wash buffer (10 mM Tris, 1 mM Ethylenediaminetetraacetic acid, pH 8; Redford 2010) and DNA extraction was executed with the Powerfecal DNA extraction kit (MoBio) following the manufacturer's instructions. The remaining of the microbiological analysis was done following the protocol described in Wuyts *et al.* (2020), with the two important modifications (i) plastid peptide nucleic acid (pPNA) (pPNA, 5'-GGCTCAACCCTGGACAG-3') and mitochondrial PNA (mPNA) clamps (mPNA, 5'-GGCA AGTGTTCTTCGGA-3') were added during the polymerase chain reaction (PCR), to specifically bind and block the amplification of plastid and mitochondrial DNA (Fitzpatrick *et al.* 2018) and (ii) DNA extraction kit blanks and two PCR blanks for each PCR plate were included to check for product contamination. The sequencing data from this study are available in the European Nucleotide Archive under the study accession number PRJEB51843.

Biomagnetic analysis

Subsequent to their arrival in Antwerp, the samples were analyzed for their leaf-area normalised saturation isothermal remanent magnetisation (SIRM). For this, projected leaf area of the samples was determined by scanning the fresh leaves with a leaf area meter (Li-300 leaf area meter, Li-COR, Lincoln Nebraska; accuracy, 0.01 cm^2). Afterwards, the leaves were oven-dried at 40 °C for four days in labelled paper bags and stored awaiting SIRM determination. SIRM determination was done following the method of Declercq *et al* (2020) with the JR6 magnetometer (AGICO, Czech Republic; sensitivity: 2.4×10^{-6} A/m) as a proxy for particulate matter originating from traffic and industry (Hofman *et al.* 2017, Kardel *et al.* 2012).

Landcover and remote sensing data

The European Urban Atlas (EEA 2012) was used to extract the composition of land use/cover around each sampled tree (Table 3). Circular buffers with a radius of 100 m, as put forward by Wuyts *et al*. (2020), were drawn around each sampling site, and herein the cover of each land use/cover class (in m^2) was calculated for each sampled tree separately. In addition, six land cover characteristics were extracted from remote sensing data for each sampled tree separately, as listed and described in Table 3. More details on the calculation of these variables can be found in the supplementary materials.

Data processing and statistical analysis

The raw sequence data were further processed with the DADA2 pipeline, as described in Callahan *et al.* (2016). Samples containing less than 500 reads, amplicon sequence variants (ASVs) occurring less than 100 times in the whole dataset and non-bacterial reads were removed. The ASVs that occurred in the blanks (both kit and PCR blanks), and that are not typically found in the phyllosphere, were removed from the data (Table A shows a list of the removed ASVs). All statistical analyses were performed in R v.4.0.3 (R core team 2019) and R Studio v. 1.3.1093 (Allaire 2011). The statistical analyses were performed separately for campaign 1 and 2 to unravel campaign-specific trends. The significance threshold is set at p value <0.05.

We calculated alpha diversity measures starting from the rarefied sequence counts of the bacterial community at ASV level, the rarefied data were calculated using the R package vegan (Oksanen 2016). For this rarefaction, a cut-off value of 1808 reads was used for both campaign 1 and 2, hereby removing 23 samples in campaign 1 and 18 samples in campaign 2. From this rarefied abundance table, the alpha diversity was calculated as ASV richness.

For analyses of composition, we made use of the count zero multiplicative approach, with the R package zCompositions (Palarea-Albaladejo & Martín-Fernández 2015), the centered log-ratio of the unrarefied reads (clr-abundances) at ASV level were calculated following the method described by Gloor *et al.* (2017). On these clr-abundances, a principal component analysis (PCA) was performed to reduce the dimensions of the dataset, using the R package vegan (Oksanen 2016), further on referred to as PCA_{clr}. Besides, we calculated two types of pairwise beta diversity: (i) the Bray-Curtis distance based on clr-abundance data using the R package vegan and (ii) the Sørensen dissimilarity based on incidence data. To visualise the Bray-Curtis distance we did a principal coordinates analysis. Next, we did an adonis and betadisper test, using the R package vegan, to test the differences in means and the spreading of the Bray-Curtis data for city and tree species. The Sørensen dissimilarity was calculated for its partitioning into the Simpson dissimilarity for species turnover and the nestedness-resultant dissimilarity using the R package betapart (Baselga *et al.* 2017). According to Baselga *et al.* (2013), for beta diversity partitioning, beta diversity of non-visible populations, such as

microorganisms, is better calculated on the incidence data rather than abundance data. Turnover refers to the replacements of species and nestedness indicates the degree to which species-poor communities are a subset of relatively richer communities (Baselga 2010).

We computed a PCA on the environmental variables in order to define new meaningful principal variables that describe the tree environment while reducing the dimensionality of the dataset. The PCA was performed on 15 environmental variables: UGA size, leaf SIRM, land surface temperature, normalized difference vegetation index, fraction of vegetation coverage, green normalized difference vegetation index, chlorophyll content in the leaves, canopy water content and urban area cover of Roads, Large structures, Green, Sport areas, Water, Urban high and Urban low, and is further on referred to as PCA_{env}. This was done on both campaigns with the R package InDaPCA (Podani *et al.* 2021) which allows for PCA on incomplete data. In order to avoid multicollinearity issues, significantly correlated variables were not tested together in a mixed model.

To explain the variation in richness and community composition, linear mixed models were built with the ASV richness and the scores of the first two components of the PCA_{clr} as dependent variables and the first two components of the PCA_{env}. as independent variables. The Akaike Information criterion was used to compare the performance of the models. The model that showed the significantly lowest Akaike Information criterion was used to do the final analysis (Bozdogan 1987). With Mantel tests (R package vegan), the relationships between the two types of pairwise beta diversity on the one hand and (i) the spatial distance between trees calculated from their geographical coordinates with the open source application Geographical Distance Matric Generator (Ersts 2019) and (ii) differences in environmental factors: SIRM, land surface temperature, UGA size and Green cover in a 100 m circular area around the sampled tree on the other hand were tested. Model performance was checked with diagnostic plots, including plots of residuals versus fitted data, normal QQ-plots and spread-location plots.Per campaign, the 50 most occurring bacterial families were determined of which the 10 most occurring families for each city were identified, based on the number of samples in which they were present. The clr-abundances of these 50 most dominant taxa were then tested for their correlation with PC1 and PC2 of PCA_{env}, using the Pearson correlation test. To account for multiple testing and false discovery rate, p values were corrected according to the Benjamini-Hochberg correction (Thissen *et al.* 2002). Finally, the 10 most occurring families for both *P. x acerifolia* and *A. pseudoplatanus* were determined, again based on the number of samples in which they were present.

Results

After filtering and processing the raw data by the DADA2 pipeline, a total of 32 930 ASVs was found for the two campaigns, of which 8638 (26.2%) were removed since they were not part of the kingdom Bacteria. In total 14 210 different bacterial ASVs were identified in campaign 1 and 12 343 in campaign 2. After the processing and quality check, campaign 1 comprised 138 samples, of which 36 originated from Antwerp, 31 from Poznan, 26 from Tartu, 25 from Zurich and 20 from Paris. Campaign 2 consisted of 93 samples, of which 5 originated from Antwerp, 27 from Almada, 29 from Poznan and 32 from Tartu. In total, 79 samples in campaign 1 and 50 samples in campaign 2 originated from *P. x acerifolia* trees and 59 samples in campaign 1 and 43 samples in campaign 2 were from *A. pseudoplatanus* trees. Of the total amount of different ASVs, 28 and 30% (respectively in campaign 1 and 2) were classified as Proteobacteria, of which 13 and 16% (campaign 1 and 2, respectively) belonged to the Alphaproteobacteria, 6 and 5% (campaign 1 and 2, respectively) were identified as part of the Gammaproteobacteria and 3% (in both campaign 1 and 2) belonged to the Betaproteobacteria. Thirteen and 12.5% (respectively in campaign 1 and 2) of the ASVs were

classified as Bacteroidetes and 9.5 and 11% (respectively in campaign 1 and 2) were identified as part of the Actinobacteria.

At city level, in campaign 1 (Table B), the *Sphingomonadaceae* were found to be the most abundant in all cities; also the *Acetobacteraceae* were one of the most abundant families, particularly in Tartu, Poznan and Zurich. The *Hymenobacteraceae* were most abundant in Paris, Tartu and Zurich, and to a lesser extent in Antwerp and Poznan. In Antwerp, the *Plactomyceataceae* were the second most abundant family, while this family was only in the top 10 in phyllosphere bacterial communities from Tartu and Poznan. In Antwerp and Poznan, the *Chitinophagaceae* were in the top three of most abundant families, while in other cities, this family was less abundant but still in the top 10. The *Methylobacteriaceae* were only seen in the top 10 of Paris and Zurich. The top 10 families in Zurich were distinct from other cities because of the high occurrence of the *AB374370_f,* the *Oxalobacteriaceae* were only in the top 10 of Paris and the *Rhodobacteraceae* were very abundant in Paris and Poznan, in contrast to the top 10 of the other cities.

In campaign 2 (Table C), the *Sphingomonadaceae* and the *Acetobacteraceae* were again the most abundant families in all cities; the *Hymenobacteraceae* were also in the top three except in Almada. In Almada, the *Planctomycetacea* were the most abundant family; they were still in the top 10 of the other cites but less prominent. The *Beijerinkiaceae* was only in the top 10 of Antwerp and Tartu and the *Methylobacteriaceae* were only in the top 10 of Antwerp and Poznan. The top 10 of Almada differed the most from other cities with five families that were ranked lower in the top 10 in other cities (*Geodermatophilaceae, Nostocaceae, Rhodospirillaceae, Bacillaeceae* and *Chroococcidiopsis*_f) and two missing families that were in the top 10 of the three other cities (*Cytophagaceae*, *Chitinophagaceae*).

At the level of tree species (Table D), the families *Sphingomonadaceae* (present in 92% of the samples of *A. pseudoplatanus* and in 94% of the samples of *P. x acerifolia)* and *Acetobacteraceae* (present in 91% of the *A. pseudoplatanus* samples and 93% of the *P. x acerifolia* samples) were found in most samples. In the *P. x acerifolia* trees, the *Planctomycetaceae* were most abundantly present *(*in *82% of the samples)*, while the *Hymenobacteraceae* were the most prominent in the *A. pseudoplatanus* trees *(in 95% of the samples).* Six families were found in the top 10 of both tree species. Four out of the 10 most abundant families were not found in the top 10 of the other tree species, although they were all present in the samples from both tree species, only less prominent. For *P. x acerifolia*, these top-10-unique families were: *CP04136_f, Rhodobacteraceae, Tepidisphaeraceae and Methylobacteriaceae* and for *A. pseudoplatanus* these families were *Beijerinckiaceae, Oxalobacteraceae, Chthoniobacteraceae* and *Enterobacteriacea.* More details on the top 10 of each city and the top 10 of each tree species can be found in the supplementary materials.

Environmental conditions in the surroundings of the tree gathered in an urbanisation gradient

The environmental conditions in the surroundings of the sampled trees differed between cities. The land surface temperature differed significantly between the cities in campaign 1 ($p<0.001$) and campaign 2 ($p<0.001$). The highest average land surface temperature around the sampled trees in summer was observed in Antwerp (29.3 °C in campaign 1 and 32.1 °C in campaign 2) and in Almada (32.8 °C for campaign 2). Tartu showed the lowest average land surface temperature, respectively 23.1 °C for campaign 1 and 23.1 °C for campaign 2. Paris (26.1 °C), Poznan (28.1°C for campaign 1 and 28.1 °C for campaign 2) and Zurich (28.2 °C) showed medium land surface temperature (figure B in the supplementary materials). Moreover, the leaf SIRM differed significantly between cities in

campaign 1 (p<0.001) and campaign 2 (p<0.001; see figure C in the supplementary materials). The highest average leaf SIRM values were found in Almada (18.55 ± 9.28 µA) and Poznan (18.54 ± 10.68 µA for campaign 2). Tartu (2.94 ± 1.72 µA for campaign 1 and 3.06 µA for campaign 2) and Zurich (4.03 ± 1.81 µA) showed the lowest average SIRM values.

Pearson correlations showed significant correlations between some of the variables describing the environmental conditions around the trees. The land surface temperature and the UGA size showed a negative correlation in campaign 1 (p=0.002 for campaign 1 and p=0.061 for campaign 2), while a positive correlation between land surface temperature and SIRM was found in both campaign 1 (p<0.001) and campaign 2 (p<0.001). No significant correlation between the leaf SIRM and the size of the UGAs could be found.

The PCA on the environmental variables ($PC1_{env}$) returned two axes explaining 47.6 % and 9.3% of the variation, respectively (Figure 1).The environmental variables land surface temperatureSum100, leaf SIRM and Roads (respectively D, B and M in Figure 1) were positively correlated to $PC1_{env}$, contrary to the variables Green cover and UGA size (respectively G and C in Figure 1), which were negatively correlated with $PC1_{env}$. The first axis can thus be seen as an urban intensity gradient, with positive values associated to more urbanization.

Alpha diversity of bacterial communities

The rarefied ASV richness, as a measure of alpha diversity, ranged between 15 (in Poznan during campaign 2) and 701 ASVs (in Almada during campaign 2), and had a mean value of 253 ASVs. The mean ASV richness per city ranged between 121 (in Zurich in campaign 1) and 357 (in Almada in campaign 2) (Figure 2). We found significant differences in the alpha diversity between cities in both campaign 1 and campaign 2. In campaign 1, the mean ASV richness was highest in Antwerp (335 ASVs). Paris (175 ASVs) and Zurich (121 ASVs) had the lowest ASV richness, while Poznan (277 ASVs) and Tartu (224 ASVs) had intermediate ASV richness . When comparing the ASV richness city by city in campaign 1, the ASV richness of Antwerp differed significantly from the ASV richness of Paris (p<0.001), Zurich (p<0.001) and Tartu (p=0.003). Besides, significant differences were observed between the ASV richness of Paris and Poznan (p=0.02) and between the ASV richness of Zurich and Poznan (p<0.001) and Tartu (p=0.03). In campaign 2, Almada (357 ASV) displayed the highest ASV richness, followed by Tartu (250 ASV) and Antwerp (246 ASV). Statistically, only the ASV richness of Poznan (151 ASV) was significantly different from the ASV richness of Almada (p<0.001) and Tartu (p=0.010) in campaign 2. No other significant intercity differences could be found. In the optimised mixed model (with UGA ID as random factor), ASV richness was significantly influenced by the city in both campaign 1 (p<0.001) and 2 (p<0.001), but not by tree species in campaign 1 (p=0.15) and 2 (p=0.67). According to the mixed model outcome, $PC1_{env}$ significantly contributed to the explanation of the variation in the ASV richness of campaign 1, this effect could not be found for the second campaign. $PC2_{env}$ did not significantly contribute to the variation in the ASV richness in campaign 1 and/or 2.

Compositional analysis via beta diversity

According to the significant and positive correlation, identified using a Mantel test (Table 4), between the Bray-Curtis beta diversity and the spatial distance between the sampled trees, the phyllosphere bacterial communities on trees that were located further from each other were more distinct in composition in campaign 2. This trend was not observed for campaign 1 or when tested for each city separately. A significant positive correlation between the Bray-Curtis beta diversity and the differences in SIRM values between the samples was also found for both campaign 1 and 2. A significant positive correlation between the Bray-Curtis beta diversity and the differences in land

surface temperature between samples was observed in campaign 2, but not in the first campaign. No significant correlations were observed between the Bray-Curtis beta diversity and the differences between the amount of green cover surrounding the sampled tree or the size of the UGAs in campaign 1 and/or 2. An overview of the Kendall's τ and p values for all the abovementioned tested correlations can be found in table 4.

For both campaign 1 and 2, city was statistically explaining the variation in Bray-Curtis beta diversity (p=0.001 for campaign 1 and campaign 2). Tree species was only significantly explaining the variation in campaign 2 (p=0.03) and to a much lower extent than city (R^2= 0.09 for city versus R^2=0.0086 for tree species). Besides, the permdisp test indicated a significant dispersion effect of both city (p=0.04 for campaign 1 and p= 0.01 for campaign 2) and tree species (p=0.04 for campaign 1 and p=0.01 for campaign 2) in both campaigns. Figure D shows the principal coordinates analysis plots for both city and tree species for campaign 1 and 2.

Following the partitioning of beta diversity and again making use of Mantel tests, the multi-site incidence based Sørensen beta diversity related significantly to the spatial distance between the trees for both campaign 1 and 2. The significant relation between the Sørensen beta diversity and the spatial distance was also found at city level for Zurich and Poznan in campaign 1 and for Almada in campaign 2. As part of the Sørensen beta diversity, the turnover component related significantly with the spatial distance in both campaign 1 and 2. In the second campaign, the Sørensen beta diversity and its turnover component related significantly with the differences in SIRM values, in Green in the 100 m buffer zone, in land surface temperature in the 100 m buffer zone and in the size of the UGAs. The nestedness component, as a second part of the Sørensen beta diversity, showed no significant relationship with the abovementioned parameters. Table 4 shows an overview of Kendall's τ and p values for all the tested correlations.

Principal component analyses of the bacterial clr-abundances

To look deeper into how the composition of the communities related with the explanatory variables (i.e. city, tree species and the urbanization gradient), the clr-abundances of a selection of ASVs across all samples were investigated. This selection was done in two approaches. In a first approach, analyses were based on the sample scores along the first two principal components of the clr-abundances in order to take into account as much variation as possible. The first ($PC1_{clr}$) and second ($PC2_{clr}$) principle components (PCs) of the PCA on the clr-abundances (PCA_{clr}) of campaign 1 explained 11.4 and 6.9% of the variation in abundances, respectively. For campaign 2, this was 24.3% for the first PC and 8.6% for the second PC. The biplot of the first two PC axes (Figure 3) shows an indication that PC1 divided the samples more on city level and PC2 also divided the samples on city level but also (very limited) on tree species level. In the final model, with UGA ID as random factor, a significant effect of city was found on both $PC1_{clr}$ and $PC2_{clr}$, but not for tree species on $PC1_{clr}$ or $PC2_{clr}$ for campaign 1 and 2.

In the optimised mixed model for campaign 1, with city and UGA ID as random factors, $PC1_{clr}$ and $PC2_{clr}$ sample scores were not significantly influenced by $PC1_{env}$. In campaign 2, $PC1_{env}$ significantly influenced the $PC1_{clr}$ and $PC2_{clr}$ sample scores. Table 5 shows the p values for each tested variable and a detailed overview for the city by city comparisons can be found in the supplementary materials.

Correlation of environmental variables with clr-abundances of the 50 most abundant families

In a second approach to gain more insight into how the community composition related with the explanatory variables, we considered the 50 most abundant families in each campaign covering 7327 ASVs or 51.6% of the total amount of ASVs for campaign 1 and 7109 ASVs or 57.2% for campaign 2. The lists of the 50 most abundant families in campaign 1 and 2 were very similar, see Table E for the complete list. Figure E shows the clr-abundances of the 50 most abundant families for campaign 1 and 2, of which 45 families occurred in the top 50 of both campaigns. In both campaigns, the top three of the most abundant families consisted of the *Sphingomonadaceae*, the *Hymenobacteraceae* and the *Acetobacteraceae* (935 ASVs for campaign 1 and 1129 ASVs for campaign 2). Only five families were different between the top 50 of the two campaigns and all of them are found in the second half of the top 50. However, despite their absence in the top 50 of respectively campaign 1 or 2, all of the previously mentioned families were present in at least one sample of both campaigns. Figure F to K shows how the abundances for the 50 most abundant families differed between cities in each campaign.

In campaign 1, only two families showed a significant correlation of their clr-abundances with $PC1_{env}$ *Deinococcaceae* and *Beijerinckiaceae*, a positive and negative correlation respectively. In campaign 2, 29 families showed a significant correlation with $PC1_{env}$ (Table 6). Of these 29 families, 12 were negatively correlated with $PC1_{env}$ (e.g. *Chitinophagaceae, Enterobacteriaceae, Clostridiaceae*) and 16 were positively correlated with $PC1_{env}$ (e.g. *Planctomycetaceae, Rhodospirillaceae, Kineosporiaceae*). Of the three most abundant families in both campaigns, i.e. *Sphingomonadaceae, Hymenobacteraceae* and *Acetobacteraceae*, only the latter showed a significant positive correlation of the clr-abundance to $PC1_{env}$ in campaign 2. Of the typical phyllosphere bacteria, the *Methylobacteriaceae* showed a negative correlation to $PC1_{env}$ in campaign 2 and the *Beijerinkiaceae* showed a negative correlation to $PC1_{env}$ in campaign 1. The *Deinococcaceae* were the only family that showed a significant correlation with $PC1_{env}$ in both campaigns. See Table 6 for an overview of all families that showed a significant correlation of their clr-abundances with $PC1_{env}$ in campaign 1 and/or 2.

Discussion

Diversity and composition across the cities and species

In this study, we aimed to determine the composition and diversity and its variation of phyllosphere bacterial communities of trees in UGAs on a large geographic scale. We assessed the impact of urbanization by testing the relationship of phyllosphere bacterial communities with different urban variables, including particulate matter pollution, microclimate and landscape characteristics such as green infrastructure cover in the surroundings and green area size. We also investigated the correlation between relative abundances of the dominant families in the phyllosphere bacterial communities and the aforementioned environmental variables.

The city appeared to be the most important driver of the phyllosphere bacterial communities in our study, since the city significantly affected both the community composition and the ASV richness . The differences in phyllosphere bacterial communities between different geographical locations was a turnover effect: some species disappeared while others appeared. In addition, when looking at the relative abundances, we found that the variation in sample scores along the first two axes of the PCA on clr-abundances were mainly explained by a city effect (see Table 5). The observation that city is a main driver of phyllosphere bacterial communities is consistent with the findings of other authors that state that geographic location is an important, if not the most important, driver of phyllosphere bacterial community composition and diversity (Finkel *et al.* 2011; Knief *et al.* 2010; Laforest-

Lapointe *et al.* 2016; Qvit-Raz *et al.* 2012). However, in contrast to the abovementioned studies, we focused on trees in urban areas, consequently excluding any variation caused by differences in ecosystems, and still we found big differences between different geographic locations/cities.

Our results strongly indicate that this geographic location effect can be the result of several factors that differed between the cities. Spatial distance effects can come from differences in the local species pool (Aronson *et al.* 2014), but our study also indicates differences in climate conditions and ambient air pollution. Significant differences between cities in land surface temperature and SIRM were indeed found (see figure B and C in supplementary materials), with the highest land surface temperature in Almada and Antwerp and the highest leaf SIRM values in Antwerp for campaign 1 and Almada and Poznan for campaign 2. City-mean fine particulate matter concentrations in 2019 (Table 1) differed greatly between cities, with highest concentrations in Poznan and Antwerp and lowest in Tartu and Zurich. The potential role of different urban stress factors that may have caused the differences in phyllosphere bacterial communities between cities is discussed below.

In contrast to the findings of, among others, Redford *et al*. (2010) and Laforest-Lapointe (2016), we only found a limited influence of host species on the composition and diversity of phyllosphere bacterial communities (Table 5). This might be due to the fact that we used other methods of analysis, or to the prevailing effect of city, which might have masked the species effect. Notable is that these tree species belong to different, not phylogenetically related orders (*A. platanoides*: Sapindales and *P. x acerifolia*: Proteales).

Urban intensity as a proxy cluster of urban stress factors

The first PC of the PCA on the environmental factors ($PC1_{env}$) can be seen as a measure of urban intensity, in which almost half of the variation in 15 environmental variables determined for the area surrounding the sampled trees are taken into account. A higher score of samples along this $PC1_{env}$ axis can be interpreted as a higher level of urbanisation. This is expressed in higher leaf SIRM values, more area covered by roads and higher land surface temperature surrounding the sampling site. Vice versa, a lower score along the $PC1_{env}$ means a higher cover by green land uses. This is expressed as larger UGAs and more area covered by green infrastructure surrounding the sampling site. This trade-off between air pollution, land surface temperature and more/larger green infrastructure is visualised in figure 1. UGAs are typically characterised by lower ambient air temperature and lower air pollution concentrations due to their location away from air pollution sources such as motorised traffic and the effect depending on the size of the UGA (Abhijith *et al.* 2017; Saaroni *et al.* 2018). This urban intensity level can be considered as a proxy for a cluster of interrelated variables, i.e. UGA size, air pollution and land surface temperature.

Effect of urban intensity on community composition and diversity

Our results show that the size of the UGA where the sample was taken has an effect on the community composition and/or the diversity. In addition, following the beta diversity results, phyllosphere bacterial communities become more and more distinct when the differences in SIRM and land surface temperature become larger (Table 4). Since the leaf SIRM and land surface temperature were significantly and positively correlated, and land surface temperature was negatively correlated with UGA size, we suggest that the bacterial communities in the phyllosphere of urban trees differ according to urban intensity, as described above. A positive relation between diversity and the amount of green infrastructure in urban areas has already been observed for several taxa, including birds and insects (Beninde *et al.* 2015). Specifically, for phyllosphere bacterial communities, Laforest- Lapointe *et al*. (2017) found that the diversity and richness of urban phyllosphere bacterial communities are higher than those of rural phyllosphere bacterial

communities. In addition, Wuyts e*t al*. (2020) observed decreasing alpha diversity of phyllosphere bacterial communities of urban trees with increasing cover of urban green and decreasing distance to urban green, while the core community becomes more dominant. Following these findings, it could be assumed that phyllosphere bacterial communities of trees located in larger UGAs have a lower diversity than those of trees located in smaller UGAs. However, we could not confirm such a negative relation between ASV richness and the size of the UGA or the amount of green area in a 100 m buffer zone. This lack of relationship at the large spatial scale of this study could be the result of the large heterogeneity in composition and structure of green area and UGAs across the cities investigated, i.e. larger than within the premises of one city such as in the studies by Laforest-Lapointe *et al*. (2017) and Wuyts *et al.* (2020). Urban green greatly varies between cities in vegetation cover, composition, structure and intensity of human activity and management, making it plausible that the influence of green in the surroundings of a tree is non-uniform.

Urban phyllosphere bacterial communities differed more in composition when located further from each other and possible explanations could be found in differences in size of the green area, land surface temperature and ambient air pollution as indicated by leaf SIRM (Table 4). This correlation of beta diversity and its turnover component with spatial distance can be a reflection of an increased chance of differences in this cluster of interrelated variables, but could as well point to restricted dispersal of phyllosphere bacteria (Finkel *et al.* 2011).

Digging deeper: urban intensity effect on relative abundances

To gain more insight into how the community composition related with the explanatory variables, we evaluated the variation in clr-abundances using two approaches, i.e. based on the sample scores along the first two axes of the PCA on clr-abundances at ASV level and based on the most abundant families. Regarding the first approach, we found that only in campaign 2 the urban intensity (i.e. the first $PC1_{env}$) significantly explained variation in the $PC1_{clr}$ and $PC2_{clr}$ sample scores (Table 5). Regarding the second approach: given the changes in composition in relation to urban intensity, we investigated how the clr-abundances of the most abundant families related to this urban intensity. For 29 out of the 50 most dominant families in campaign 1 and 2, a significant relation of their clr-abundance with either more urbanised or more green surroundings was observed (Table 6). *Acetobacteraceae, Planctomycetes* and *Beijerinkiaceae*, three of the most dominant families, showed a negative relation of clr-abundance with $PC1_{env}$.This means that the abundances of these families increased with an increasing proportion of green areas in a radius of 100 m around the tree. For the *Beijerinkiaceae*, whose members are typically diazotrophic (Kembel *et al.* 2014), the positive relationship with the amount of surrounding green is in line with the results of Wuyts et al. (2020) and Smets *et al.* (2016). The positive relations between these three families and the amount of green area in the surroundings suggest that these families can be considered to be typical classical phyllosphere bacterial families.

In contrast to the abovementioned families, the clr-abundances of *Enterobacteriaceae*, *Comamonadaceae* and *Bacillaceae* showed a positive correlation with more urbanised surroundings. *Enterobacteriaceae* are found in diverse environments, ranging from the gut of humans and other animals to (waste)water and soils (De Oliveira *et al.* 2017). Members of the family *Comamonadaceae* are frequently associated with the use of aromatic compounds, such as benzene, 2,4-dinitrotoluene, glycerol, acetate and sulfolane as an energy source (Blümel *et al.* 2001; Fahy *et al.* 2006; Satola *et al*. 2013). These aromatic compounds are typically associated with traffic and consequently more abundant in urban areas (Bravo *et al.* 2002; Dor *et al.* 1995). The family *Bacillaceae* is especially known for their robustness. *Bacillaceae* are able to resist heat, radiation, chemicals and drought by forming endospores, making them true specialists (Bottone 2010; Hoyles *et al.* 2012; Schmidt *et al*.

2011). The members of these families could therefore be considered to be able to better adapt and survive in more urban settings than the typical phyllosphere bacterial families discussed above, and could be seen as typical urban dwellers in the phyllosphere.

Phyllosphere bacterial communities in the city

In cities, phyllosphere bacterial communities face several stressors, such as high temperatures as a consequence of the urban heat island effect (Perry *et al.* 2013) and higher (air) pollution (Grimm *et al.* 2008). Water availability is a crucial parameter for phyllosphere bacteria to survive (Vorholt 2012). However, lower atmospheric humidity and higher temperature in urban areas can lead to water scarcity in both the soil and on the leaves. This is, among other things, an effect of the large areas of impervious surfaces, which in turn leads to lower evapotranspiration rates and lower (air) humidity. In addition, water scarcity in the soil could have an indirect effect on phyllosphere bacterial composition through the effect on the host plant. Air pollution in its turn can influence leaf traits such as the wettability of a leaf (Shepherd & Griffiths 2006) and as such indirectly influence phyllosphere bacterial communities composition (Wuyts *et al*. 2020). Besides, air pollution has toxic effects on living organisms, and thus most likely also on (some) phyllosphere bacterial communities, through oxidative stress (Gangwar *et al.* 2020). In addition (air) pollution could be used as a carbon source, consequently having a fertilizing effect. Lower (air) humidity, higher temperatures and (air) pollution can thus directly or indirectly influence phyllosphere bacterial communities. The correlation of phyllosphere bacterial families with the urban intensity variable complex could be a consequence of the abovementioned stressors, although an effect of a smaller local species pool was not ruled out. With climate change, it might be expected that these urban stressors will increase in magnitude, making it even harder for the typical classic phyllosphere bacteria to adapt and survive in favour of typical urban dwellers. This could consequently lead to a change in phyllosphere bacterial communities and, in turn, lead to changes in the functioning and health of urban trees and in the delivery of ecosystem services of plants (Haase *et al.* 2014). If we take the role of these phyllosphere bacterial communities in the functioning of the ecosystem into account, we risk that these trees can not function in the same way and/or deliver the exact same ecosystem services. To ensure preservation of the classical phyllosphere bacterial communities and the ecosystem services they provide, it is important to include enough large UGAs when constructing urban environments.

Future prospects and limitations

Our results show that sampling of phyllosphere bacterial communities on a large scale is possible. However, this study also shows that it comes with some points of attention. First, it is best to keep the sampling period as short as possible, to facilitate comparisons, e.g. in ASV richness . Sampling time is also of great importance, since sampling at the beginning of the growing season (campaign 1) or at the end of the growing season (campaign 2) showed big differences in the results. The effect of city overruled the effect of host species on phyllosphere bacterial communities, making it justifiable to sample different tree species in following monitoring campaigns if needed. It should be kept in mind that 16S rRNA gene amplicon analysis is limited in resolution, and thus, it does not provide quantitative data and it does not cover functional traits.

Conclusions

Our urban intensity complex, consisting of the interrelated variables air pollution, surface temperature and UGA size significantly influenced the phyllosphere bacterial community. This means that urbanization plays an important role in the diversity and composition of phyllosphere bacterial communities on urban trees. On the other hand, geographic location proved to be the prevailing factor for phyllosphere bacterial communities, in contrast to host species that was found to only

have a limited influence on the phyllosphere bacterial community. We see that typical classic phyllosphere bacterial families, such as *Planococcaceae*, *Beijerinckiaceae* and *Acetobacteraceae*, are more abundant in more green surroundings. On the contrary, more human associated bacterial families such as *Enterobacteriaceae* are more associated with grey or urbanized surroundings and they could be considered urban dwellers of the phyllosphere. To make sure that we preserve the typical classic phyllosphere bacterial communities in our city and the ecosystem services they provide, it is important to include enough, large UGAs when constructing urban environments. Lastly, because of the heterogeneous outcomes of the urbanization effect in different cities, including multiple cities in future studies concerning the phyllosphere in an urban context is very important.

Funding

This work was supported by the European ERA Net BiodivERsA project 'BIOVEINS: Connectivity of green and blue infrastructures: living veins for biodiverse and healthy cities.', by a PhD grant Strategic Basic Research of the Research Foundation - Flanders (FWO) [1S84819N] to B.M., by the Fundação para a Ciência e a Tecnologia [SFRH/BD/141822/2018] to C.A. and [2020.03415.CEECIND] to P.P., by a postdoctoral scholarship of the FWO [12ZJ821N] to W.S, by the Estonian Research council [PSG708] to G.P., by the European Research Council [852600] to S.L., by Poland [NCN/2016/22/Z/NZ8/00004], by the Swiss National Science Foundation [31BD30_172467] to J.C.A. and by European Social Fund's Dora Plus Programme to M.A.O..

Conflicts of interest

No conflicts of interest are declared.

Acknowledgements

The authors would like to thank all BIOVEINS partners and the Centre for Medical Genetics Antwerp for the use of the Illumina MiSeq sequencing system.

References

Abhijith KV, Kumar P, Gallagher J *et al.* Air pollution abatement performances of green infrastructure in open road and built-up street canyon environments – A review. *Atmospheric Environment* 2017;162: 71–86.

Alberti M. The effects of urban patterns on ecosystem function. *International Regional Science Review* 2005;28:168–192.

Allaire J. The R User Conference 2011. *R Studio: Intergrated Development Environment For R* 2011;770:14.

Aronson MFJ, La Sorte FA, Nilon CH *et al*. A global analysis of the impacts of urbanization on bird and plant diversity reveals key anthropogenic drivers. *Proceedings of the Royal Society B: Biological Sciences* 2014;281:1780.

Aronson MFJ, Lepczyk CA, Evans KL *et al*. Biodiversity in the city: key challenges for urban green space management. *Frontiers in Ecology and the Environment* 2017;15:189–196.

Aydogan EL, Moser G, Müller C *et al*. Long-term warming shifts the composition of bacterial communities in the phyllosphere of Galium album in a permanent grassland field-experiment. *Frontiers in Microbiology* 2018;9: 1–17.

Baselga A. Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* 2010;19:134–143.

Baselga A, Orme D, Villéger S *et al*. Betapart: Partitioning beta diversity into turnover and nestedness components. R Package version, 1-4-1. 2017. https://CRAN.R-project.org/package=betapart.

Beninde J, Veith M, Hochkirch A. Biodiversity in cities needs space: A meta-analysis of factors determining intra-urban biodiversity variation. *Ecology Letters* 2015;18:581–592.

Blümel S, Busse HJ, Stolz A *et al.* Xenophilus azovorans gen. nov., sp. nov., a soil bacterium that is able to degrade azo dyes of the Orange II type. *International Journal of Systematic and Evolutionary Microbiology* 2001;51:1831–1837.

Bozdogan H. "Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 1987;52: 345-370.

Bottone EJ. Bacillus cereus, a volatile human pathogen. *Clinical Microbiology Reviews* 2010;23: 382–398.

Bravo H, Sosa R, Sánchez P *et al*. Concentrations of benzene and toluene in the atmosphere of the southwestern area at the Mexico City Metropolitan Zone. *Atmospheric Environment* 2002; 36:3843–3849.

Brighigna L, Gori A, Gonnelli S *et al*. The influence of air pollution on the phyllosphere microflora composition of Tillandsia leaves (Bromeliaceae). *Revista de biología tropical* 2000;48:511-517.

Bulgarelli D, Schlaeppi K, Spaepen S *et al*. Structure and functions of the bacterial microbiota of plants. *Annual Review of Plant Biology* 2013;64:807–838.

Calfapietra C, Peñuelas J, Niinemets Ü. Urban plant physiology: Adaptation-mitigation strategies under permanent stress. *Trends in Plant Science* 2015;20:72–75.

Callahan B, McMurdie P, Rosen M *et al*. DADA2: High resolution sample inference from illumina amplicon data. *Encyclopedia of Medical Immunology 2016*;13:1–7.

Cilliers S, Cilliers J, Lubbe R et al. Ecosystem services of urban green spaces in African countries-perspectives and challenges. *Urban Ecosystems* 2013;16:681–702.

Declercq Y, Samson R, Van De Vijver E *et al*. A multi-proxy magnetic approach for monitoring large-scale airborne pollution impact. *Science of The Total Environment 2020;*743:140718.

Díaz S et al. Assessing nature's contributions to people. *Science* 2018;359.6373:270-272.

Dohrmann AB, Tebbe CC. Effect of elevated tropospheric ozone on the structure of bacterial communities inhabiting the rhizosphere of herbaceous plants native to Germany. *Applied and Environmental Microbiology* 2005;71:7750–7758.

Dor F, Le Moullec Y, Festy B. Exposure of city residents to carbon monoxide and monocyclic aromatic hydrocarbons during commuting trips in the Paris metropolitan area. *Journal of the Air and Waste Management Association* 1995;45:103–110.

Dubberke ER, Haslam DB, Lanzas C *et al.* The ecology and pathobiology of Clostridium difficile infections: An interdisciplinary challenge. *Zoonoses and Public Health* 2011;58: 4–20.

European Environment Agency (EEA) (2021) European city air quality viewer. https://www.eea.europa.eu/themes/air/urban-air-quality/european-city-air-quality-viewer. (11 Dec 2021, date last accessed)

Emmett DJ, Godwin CM, Cardinale BJ. Biodiversity effects in the wild are common and as strong as key drivers of productivity. *Nature* 2017;549:261–264.

Enya J, Shinohara H, Yoshida S *et al*. Culturable leaf-associated bacteria on tomato plants and their potential as biological control agents. *Microbial Ecology* 2007;53:524–536.

Ersts PJ. Geographic Distance Matrix Generator. (version 1.2.3). *American Museum of Natural History*. Center for Biodiversity and Conservation. Available from http://biodiversityinform atics.amnh.org/open source/gdmg 2019.

Eurostat (2021) Population density by NUTS 3 region. https://appsso.eurostat.ec.europa.eu/nui/show.do?query=BOOKMARK_DS-115323_QID_-5438C657_UID_-3F171EB0&layout=TIME,C,X,0;GEO,B,Y,0;UNIT,B,Z,0;INDICATORS,C,Z,1;&zSelection=DS-115323UNIT,PER_KM2;DS-115323INDICATORS,OBS_FLAG;&rankName1=UNIT_1_2_-1_2&rankName2 (11 February 2022, date last accessed).

Fahlgren C, Bratbak G, Sandaa RA *et al*. Diversity of airborne bacteria in samples collected using different devices for aerosol collection. *Aerobiologia* 2011;27:107–120.

Fahy A, McGenity TJ, Timmis KN *et al.* Heterogeneous aerobic benzene-degrading communities in oxygen-depleted groundwaters. *FEMS Microbiology Ecology* 2006;58:260–270.

Ferrari BC, Binnerup SJ, Gillings M. Microcolony cultivation on a soil substrate membrane system selects for previously uncultured soil bacteria. *Applied and Environmental Microbiology* 2005;71:8714–8720.

Finkel OM, Burch AY, Lindow SE *et al.* Geographical location determines the population structure in phyllosphere microbial communities of a salt-excreting desert tree. *Applied and Environmental Microbiology* 2011;77:7647–7655.

Fitzpatrick CR, Lu-Irving P, Copeland J *et al*. Chloroplast sequence variation and the efficacy of peptide nucleic acids for blocking host amplification in plant microbiome studies. *Microbiome* 2018;6:144.

Gandolfi I, Canedoli C, Imperato V *et al*. Diversity and hydrocarbon-degrading potential of epiphytic microbial communities on Platanus x acerifolia leaves in an urban area. *Environmental Pollution* 2017;220:650–658.

Gangwar RS, Bevan GH, Palanivel R *et al*. Oxidative stress pathways of air pollution mediated toxicity: Recent insights. *Redox Biology* 2020;34:101545.

Gloor GB, Macklaim JM, Pawlowsky-Glahn V *et al.* Microbiome datasets are compositional: and this is not optional. *Frontiers in microbiology*, 2017, 8: 2224.

Grimm NB, Faeth SH, Golubiewski NE *et al*. Global change and the ecology of cities. *Science* 2008;319:756–760.

Haase D, Larondelle N, Andersson E *et al.* A quantitative review of urban ecosystem service assessments: Concepts, models, and implementation. *Ambio* 2014;43:413–433.

Hallik L, Kazantsev T, Kuusk A *et al.* Generality of relationships between leaf pigment contents and spectral vegetation indices in Mallorca (Spain). *Regional Environmental Change* 2017;17:2097–2109.

Hofman J, Maher BA, Muxworthy AR *et al.* (2017). Biomagnetic monitoring of atmospheric pollution: a review of magnetic signatures from biological sensors. *Environmental Science & Technology* 2017;51:6648-6664.

Hortal J, De Bello F, Diniz-Filho JAF *et al.* Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics* 2015;46:523–549.

Hoyles L, Honda H, Logan NA *et al.* Recognition of greater diversity of Bacillus species and related bacteria in human faeces. *Research in Microbiology* 2012;163:3–13.

Ives CD, Lentini PE, Threlfall CG *et al.* Cities are hotspots for threatened species. *Global Ecology and Biogeography* 2016;25:117–126.

Jo Y, Cho JK, Choi H *et al.* Bacterial communities in the phylloplane of Prunus species. *Journal of Basic Microbiology* 2015;55:504–508.

Kardel F, Wuyts K, Maher BA *et al*. Intra-urban spatial variation of magnetic particles: monitoring via leaf saturation isothermal remanent magnetisation (SIRM). *Atmospheric Environment* 2012;55:111-120.

Kembel SW, O'Connor TK, Arnold HK *et al.* Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. *Proceedings of the National Academy of Sciences of the United States of America* 2014;111:13715–13720.

Kim M, Singh D, Lai-Hoe A *et al*. Distinctive Phyllosphere Bacterial Communities in Tropical Trees. *Microbial Ecology* 2012;63:674–681.

Kottek M, Grieser J, Beck C *et al*. World map of the Köppen-Geiger climate classification updated, *Meteorologische Zeitschrift* 2006;15:259–263.

Knapp S, Aronson M, Carpenter E *et al.*. research agenda for urban biodiversity in the global crisis. *BioScience* 2021;71:268–279.

Knief C, Delmotte N, Chaffron S *et al.* Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *ISME Journal* 2012 ;6 :1378–1390.

Knief C, Ramette A, Frances L *et al.* Site and plant species are important determinants of the Methylobacterium community composition in the plant phyllosphere. *ISME Journal* 2010;4(6):719–728.

Laforest-Lapointe I, Messier C, Kembel SW. Host species identity, site and time drive temperate tree phyllosphere bacterial community structure. *Microbiome* 2016;4:1–10.

Laforest-Lapointe I, Paquette A, Messier C *et al.* Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature* 2017;546:145–147.

Legein M, Smets W, Vandenheuvel D *et al.* (2020). Modes of Action of Microbial Biocontrol in the Phyllosphere. *Frontiers in Microbiology* 2020:1619.

Lindow SE, Brandl MT. Microbiology of the phyllosphere. *Applied and Environmental Microbiology* 2003;69:1875–1883.

Mardia KV, Kent JT, Bibby JM. Multivariate Analysis. *Academic Press* 1979.

Marquez MC, Sanchez-Porro C, Ventosa A. Hallophilic and haloakalophilic, aerobic endospore-forming bacteria in soil. *Endospore-Forming Soil Bacteria* 2011:309-339.

Marsella AJ. Urbanization, mental health, and social deviancy: A review of issues and research. *American Psychologist* 1998;53:624–634.

McGarigal K. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. *US Department of Agriculture, Forest Service, Pacific Northwest Research Station,* 1995;351.

McGarvey JA, Tran T, Han R. Bacterial population dynamics after foliar fertilization of almond leaves. *Journal of applied microbiology* 2019;126:945-953.

Morris CE, Kinkel LL *et al*. Fifty years of phyllosphere microbiology: significant contributions to research in related fields. *Phyllosphere microbiology* 2002: 365-375.

Mitchell MGE, Bennett EM, Gonzalez A *et al.* The montérégie connection: Linking landscapes, biodiversity, and ecosystem services to improve decision making. *Ecology and Society* 2015;20.

Moyes AB, Kueppers LM, Pett-Ridge J *et al.* Evidence for foliar endophytic nitrogen fixation in a widely distributed subalpine conifer. *New Phytologist* 2016;210:657–668.

Oksanen J, Blanchet FG, Friendly M et al. vegan: Community Ecology Package. *R package version* 2016;2:4–10. https://CRAN.R-project.org/package=vegan

De Oliveira DV, Nunes LS, Barth AL *et al.* Genetic background of β-lactamases in Enterobacteriaceae isolates from environmental samples. *Microbial ecology* 2017;74: 599-607.

Pakarinen J, Hyvärinen A, Salkinoja-Salonen M *et al*. Predominance of Gram-positive bacteria in house dust in the low-allergy risk Russian Karelia. *Environmental Microbiology* 2008;10:3317–3325.

Palarea-Albaladejo J, Martín-Fernández JA. ZCompositions - R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics and Intelligent Laboratory Systems* 2015;143:85–96.

Pan J. et al. The phyllosphere indigenous microbiota of Brassica campestris L. change its diversity in responding to di-n-butyl phthalate pollution. *World Journal of Microbiology and Biotechnology* 2019;35.2: 1-9.

Perry A, Thompson R. *Applied climatology: principles and practice:* Routledge, 2013.

Pinho P, Casanelles-Abella J, Luz AC *et al.* Research agenda on biodiversity and ecosystem functions and services in European cities. *Basic and Applied Ecology* 2021;53:124–133.

Podani J, Kalapos T, Barta B *et al.* Principal component analysis of incomplete data – A simple solution to an old problem. *Ecological Informatics* 2021;61.

Uchida K, Blakey RV, Burger JR *et al*. Urban Biodiversity and the Importance of Scale. *Trends in Ecology and Evolution* 2021;36:123–131.

QGIS Development Team (2020) 'QGIS Geographic Information System. Open Source Geospatial Foundation Project.' Available at: http://qgis.osgeo.org.

Qvit-Raz N, Finkel OM, Al-Deeb TM *et al.* Biogeographical diversity of leaf-associated microbial communities from salt-secreting Tamarix trees of the Dead Sea region. *Research in Microbiology* 2012;163:142–150.

Redford AJ, Bowers RM, Knight R *et al.* The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environmental microbiology* 2010;12:2885-2893.

Saaroni H, Amorim JH, Hiemstra JA *et al.* Urban Green Infrastructure as a tool for urban heat mitigation: Survey of research methodologies and findings across different climatic regions. *Urban Climate* 2018;24:94–110.

Satola B, Wübbeler JH, Steinbüchel A. Metabolic characteristics of the species Variovorax paradoxus. *Applied Microbiology and Biotechnology* 2013;97:541–560.

Schmidt TR, Scott EJ, Dyer DW Whole-genome phylogenies of the family Bacillaceae and expansion of the sigma factor gene family in the Bacillus cereus species-group. *BMC Genomics* 2011;12.

Shepherd T, Griffiths DW. The effects of stress on plant cuticular waxes. *New Phytologist* 2006;171:469–499.

Smets W, Wuyts K, Oerlemans E *et al.* Impact of urban land use on the bacterial phyllosphere of ivy (Hedera sp.). *Atmospheric Environment* 2016;147:376–383.

Sobrino JA, Jiménez-Munoz JC; Paolini L. Land surface temperature retrieval from LANDSAT TM-5. *Remote Sensing of Environment* 2004;90:434–440.

Timedanddate (2022) Available at: timeanddate.com (13 December 2021, date last accessed)

Thissen D, Steinberg L, Kuang D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics* 2002;27:77–83.

Truchado P, Gil MI, Moreno-Candel M *et al.* Impact of weather conditions, leaf age and irrigation water disinfection on the major epiphytic bacterial genera of baby spinach grown in an open field. *Food microbiology* 2019;78:46-52.

United Nations, Department of Economic and Social Affairs, Population Division World Urbanization Prospects: The 2014 Revision Highlights 2014 UN

United Nations Publications. Resolutions and Decisions of the Economic and Social Council. Vol. 1. United Nations Publications, 2006.

Urban Atlas. 2012. http://land.copernicus.eu/

Vacher C, Hampe A, Porté AJ *et al.* The Phyllosphere: Microbial Jungle at the Plant-Climate Interface. *Annual Review of Ecology, Evolution, and Systematics* 2016;47:1–24.

Van Mensel A, Wuyts K, Pihno P *et al.* Magnetic signal of urban trees' trunk bark catches the variation in particulate matter exposure within and across six European cities. Unpublished 2022.

Vlahov, D, Galea S. Urbanization, urbanicity, and health. Journal of Urban Health 2002;79:1–12.

Vorholt JA. Microbial life in the phyllosphere. *Nature Reviews Microbiology* 2012;10:828–840.

Wei X. et al. Phylloremediation of air pollutants: exploiting the potential of plant leaves and leaf-associated microbes. *Frontiers in plant science* 2017;8-1318.

Weyens N. et al. The role of plant–microbe interactions and their exploitation for phytoremediation of air pollutants. *International journal of molecular sciences* 2015;16.10: 25576-25604.

Wuyts K, Smets W, Lebeer S *et al*. Green infrastructure and atmospheric pollution shape diversity and composition of phyllosphere bacterial communities in an urban landscape. *FEMS Microbiology Ecology* 2020;96:1–14.

Zhou ZX, Jiang H, Yang C *et al.* (2010). Microbial community on healthy and diseased leaves of an invasive plant Eupatorium adenophorum in Southwest China. *Journal of Microbiology* 2010 ;48 :139–145.

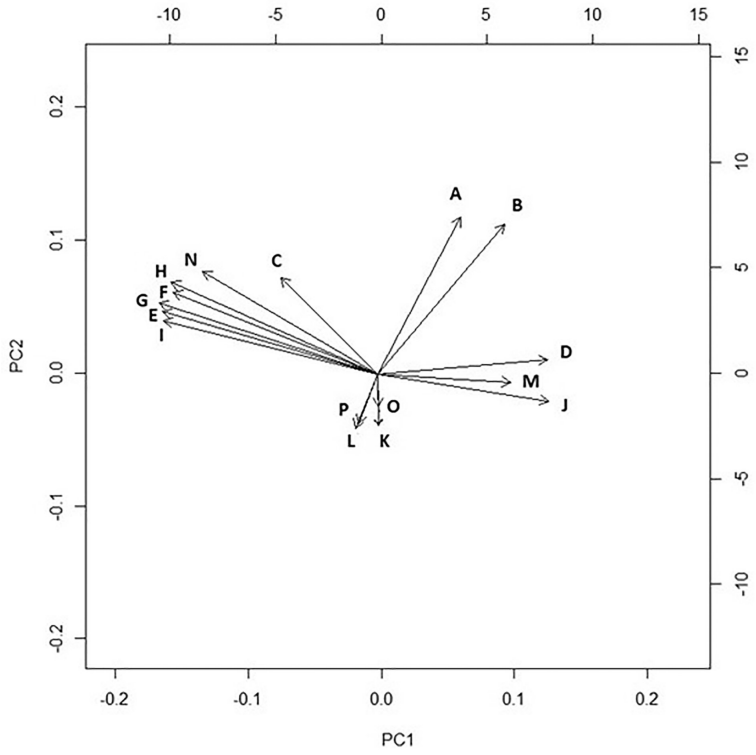

Figure 1 Biplot of the sample scores along the first two principal components of the environmental variables taking into account the data from the two sampling campaigns. The first principle component (PC1) explained 47.6% and the second principle component (PC2) explained 9.3% of all variation. A: SIRM values for campaign 1, B: SIRM values for campaign 2, C: park area, D: land surface temperature in, E: normalized difference vegetation index, F: green normalized difference vegetation index, G: fraction of the area covered by vegetation H: leaf chlorophyll content, I: canopy water content, J: area covered by urban high, K: area covered by urban low, L: area covered by large structures, M: area covered by roads, N: area covered by vegetation, O: area covered by sports areas and P: area covered by water. Each variable is measured or calculated individually in a 100 m circular area around the sampled tree.

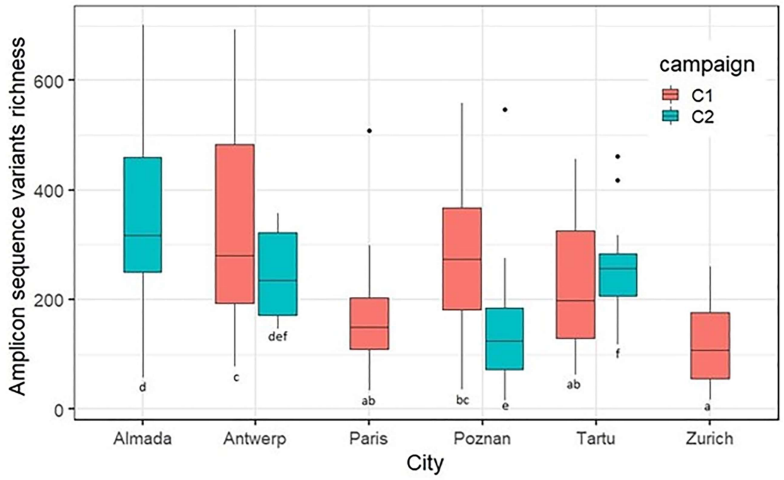

Figure 2 Amplicon sequence variants richness of the phyllosphere bacterial communities of Platanus x acerifolia and Acer pseudoplatanus trees for each city in campaign 1 (red) and campaign 2 (green). Letters indicate groups of significance. Indicated are the maximum observation, third quartile, median, first quartile and minimum observation.

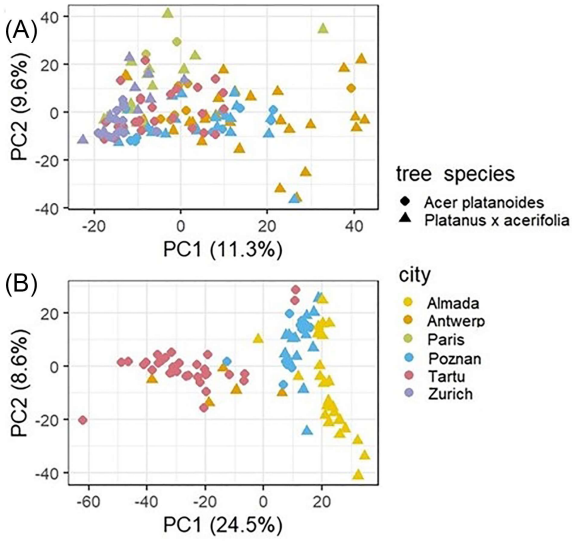

Figure 3 Biplot of the sample scores along the first two principal components after principal component analysis on the clr-abundances of the phyllosphere bacterial communities of campaign 1 (panel A) and 2 (panel B). Colors represent different cities, squares represent Acer pseudoplatanus and triangles Platanus x acerifolia.

Table 1 Geographic location, mean annual concentration of fine particulate matter($PM_{2.5}$, in µg/m^3) for 2019 and 2020 (EEA 2021, Van Mensel et al. (2020)), mean minimum and maximum temperature (T_{min} and T_{max}, in °C), total precipitation (in mm), average relative air humidity (RH in %), population density (inhabitants per km^2 in 2018 (Eurostat, 2021)), climate type following the classification of Köppen-Geiger (Cfb= Temperate maritime climate, Csa = warm Mediterranean climate and Dfb = Moderate continental climate with precipitation throughout the year of each city included in this study (Kottek et al. (2006)). Source meteorological data: www.timeanddate.com; T_{min}, T_{max}, total precipitation and RH were averaged for the months when the sampling campaigns took place (respectively June to September).

City	Location	$PM_{2.5}$	T_{min}	T_{max}	Precipitation	RH	Pop. dens.	Climate
Antwerp	51° 13' N 4° 24' E	12.5	7.0 ± 4.7	32.0 ± 4.9	182.0	74.5 ± 3.1	6500	Cfb
Almada	38° 42' N 9° 11' W	9.5	14.0 ± 1.6	36.5 ± 6.5	12.1	66.7 ± 1.7	6458	Csa
Paris	48° 52' N 2° 22' E	10.5	9.2 ± 3.6	32.2 ± 3.4	214.8	69.0 ± 2.9	21 616	Cfb
Poznan	52° 24' N 16° 55' E	17.2	5.5 ± 2.5	32.7 ± 2.2	136.6	72.0 ± 4.1	2031	Dfb
Tartu	58° 42' N 26° 45' E	5.2	5.2 ± 2.6	28.7 ± 2.6	284.3	75.0 ± 5.7	619	Dfb
Zurich	47° 23' N 8° 32' E	8.6	8.0 ± 3.7	30.2 ± 3.7	445.0	72.8 ± 4.5	4700	Cfb

Table 2 Overview of the number of samples taken in each sampling campaign, per city and per tree species. NA: not sampled.

	Campaign 1		Campaign 2	
	A. Pseudoplatanus	***P. x acerifolia***	***A. Pseudoplatanus***	***P. x acerifolia***
Almada	NA	NA	0	27
Antwerp	7	29	0	5
Paris	2	18	NA	NA
Poznan	11	20	11	18
Tartu	26	0	32	0
Zurich	13	12	NA	NA

Table 3 Overview of the used land use/land cover (LU/LC) classes following the Urban Atlas (EEA 2012) and the land cover characteristics derived from remote sensing data

Land use/ land cover data		
Abbreviation	**Definition**	**Combined classes in UA (2012)**
Urban high	continuous and discontinuous dense urban area	11210, 11100
Urban low	discontinuous medium, to very low density urban area	11220, 11230, 11240
Roads	roads and associated land	12210, 12220
Large structures	industrial, commercial, public, military and private units	12100
Sports areas	sports and leisure areas	14200
Green	green areas including forests, herbaceous vegetation and pastures	23000, 31000, 32000
Water	water bodies	50000
Remote sensing data		
Abbreviation	**Definition**	**Unit**
LST	Land surface temperature, i.e. radiative skin temperature of the Earth's surface	°C
NDVI	Normalized Difference Vegetation index, indicator of greenness of vegetation	-
GNDVI	Green Normalized Difference Vegetation index, more sensitive for dense vegetation than NDVI	-
FCOVER	Proportion of surface covered by vegetation	%
Cab	leaf chlorophyll content (chlorophyll a and b)	µg/cm² of leaf surface
CW	Indicator of leaf water content	µg water/cm² of leaf surface

Table 4 Overview of the Kendall's τ and p values according to the Mantel tests used for testing correlations between on the one hand pairwise beta diversity measures and on the other hand the spatial distance and the pairwise differences in leaf SIRM, in Green and land surface temperature (LST) in the surrounding and in UGA size, for campaign 1 and 2. Numbers in bold are statistically significant.

	Campaign 1				**Campaign 2**			
	Bray Curtis		**Sørensen**		**Bray Curtis**		**Sørensen**	
	Kendall's τ	**p value**	**Kendall's τ**	**p value**	**Kendall's τ**	**p value**	**Kendall's .τ**	**p value**
spatial distance	0.026	0.146	**0.128**	**0.001**	**0.239**	**0.001**	**0.493**	**0.001**
Δ leaf SIRM	**0.034**	**0.023**	0.013	0.633	**0.048**	**0.022**	**0.281**	**0.001**
Δ Green	-0.016	0.815	0.002	0.480	0.019	0.131	**0.113**	**0.005**
ΔLST	-0.029	0.952	0.034	0.198	**0.045**	**0.003**	**0.359**	**0.001**
Δ UGA size	-0.019	0.844	0.044	0.268	0.025	0.067	**0.432**	**0.001**

Table 5 Overview of the p values according to the outcome of the mixed model explaining the variation in samples scores along the $PC1_{clr}$ and $PC2_{clr}$ in campaign 1 and 2, by the explanatory variables city, tree species and the sample scores along PC_{env}. Columns show the dependent variables and the rows represent the explanatory variables.

	Campaign 1		Campaign 2	
	PC1clr	**PC2clr**	**PC1clr**	**PC2clr**
city	<0.001	<0.001	<0.001	<0.001
tree species	0.08	0.66	0.55	0.91
PC_{env}	0.21	0.56	0.015	0.016

Table 6 Correlations of clr-abundance of the most abundant families with $PC1_{env}$ in Campaign 1 and Campaign 2. Green squares mean significant correlation to the more green variables and grey squares mean significant correlation to the more urban variables. Empty cells were not significantly correlated to $PC1_{env}$ for that specific campaign.

Phylum	Class	Order	Family	Campaign 1	Campaign 2
Acidobacteria	Acidobacteriia	Acidobacteriales	Acidobacteriaceae		-0.274
Acidobacteria	CP015136_c	CP015136_o	CP015136_f		-0.393
Actinobacteria	Actinobacteria_c	Kineosporiales	Kineosporiaceae		0.395
Actinobacteria	Rubrobacteria	Gaiellales	Gaiellaceae		0.512
Armatimonadetes	AB374370_c	AB374370_o	AB374370_f		-0.351
Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae		-0.561
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae		-0.429
Chloroflexi	Thermomicrobia	DQ129389_o	DQ129389_f		0.379
Cyanobacteria	Chroobacteria	Pleurocapsales	Chroococcidiopsis_f		0.411
Cyanobacteria	Hormogoneae	Nostocales	Nostocaceae		0.312
Deinococcus-Thermus	Deinococci	Deinococcales	Deinococcaceae	-0.3339	0.408
Firmicutes	Bacilli	Bacillales	Bacillaceae		-0.565
Firmicutes	Bacilli	Bacillales	Planococcaceae		0.594
Firmicutes	Clostridia	Clostridiales	Clostridiaceae		-0.358
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae		0.291
Planctomycetes	Phycisphaerae	Phycisphaerales	Tepidisphaeraceae		-0.407
Planctomycetes	Planctomycetia	Planctomycetales	Planctomycetaceae		0.651
Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae		-0.290
Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	0.3583	
Proteobacteria	Alphaproteobacteria	Rhizobiales	Devosia_f		0.558
Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae		0.676
Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae		0.477
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae		-0.431
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae		0.490
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae		0.590
Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae		0.386
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae		-0.296
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae		-0.308
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae		-0.526
Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae		0.512