scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Root-associated microbiota of decline-affected and asymptomatic *Pinus sylvestris* trees

Ana V. Lasa[1,3]✉, Antonio J. Pérez-Luque[2,3] & Manuel Fernández-López[1]✉

Forest decline is a worldwide phenomenon affecting many species such as *Pinus sylvestris*. Although it is driven by multiple stressors, the role of tree associated microorganisms remains still unclear. To reduce this knowledge gap we obtained amplicon sequences of the microbiota inhabiting the rhizosphere soil and root endosphere (bacterial 16S rRNA and fungal ITS2) of decline-affected and asymptomatic *P. sylvestris* trees in spring and summer. The dataset comprised a total of 384 samples from three mountainous areas which yielded an average of 59,592.3 ± 7,371 and 56,894.3 ± 12,983.5 (spring and summer) bacterial and 74,827.9 ± 12,095.4 and 85,363.9 ± 14,199.3 (spring and summer) fungal raw reads, resulting in 23,982.4 ± 11,312.4 (spring) and 17,921.8 ± 10,802.7 (summer) bacterial and 50,571.1 ± 10,499.5 (spring) and 49,509.4 ± 12,673.8 (summer) fungal quality-filtered sequences. These data and the corresponding metadata could be used to identify pine decline bioindicators, to develop novel diagnosis tools of specific microorganisms and could serve as reference against which to compare other microbial communities.

Background & Summary

In the last decades, forests are suffering a generalized deterioration due to decline episodes[1]. This phenomenon results in a reduced growth of the trees, commonly accompanied by the yellowing and defoliation of the crown, and can even lead to the tree death. High tree mortality rates have been reported worldwide, however, the cause of forest decline remains still an open question[1–3]. Forest decline is thought to be driven by multiple interacting factors, whose weight may vary along the time[4]. Consequently, understanding and predicting decline episodes remains a challenging task. Manion (1981)[5] proposed a model to describe this complex phenomenon: long-term stressors known as predisposing factors e.g., high tree density) affect negatively the trees. Subsequently, inciting factors such as droughts or other short-term extreme weather events further stress the trees, making them more vulnerable. Eventually, contributing factors like pathogens or pests overcome the resistance of affected hosts, potentially leading to tree mortality. Therefore, for a better understanding of tree mortality and forest decline, multidisciplinary approaches that integrate different and complementary spatiotemporal scales are required[6]. Such approaches will provide deeper insights into the processes driving forest decline.

Pinus sylvestris L. (Scots pine) is a widely distributed boreo-alpine species that is undergoing severe decline episodes[7–9], especially in the southeast of the Iberian Peninsula[10], its southernmost limit of distribution. So much so that 314.8 ha of the National Park of Sierra Nevada (Granada and Almeria, SE Spain) covered by *P. sylvestris* stands were affected by decline in 2017[11]. Populations located in the southernmost limit of their distribution, i.e. rear-edge populations are of great interest due to their ecological and evolutionary significance[12]. Studying the ecological dynamics at species range margins offers valuable insights into their potential to adapt and respond to climate change[13]. Rear-edge populations, located at the low-latitude extremes, often harbour unique genetic diversity essential for species conservation, management, and long-term resilience[12,14]. Climatic scenarios at rear-edge sites can act as natural laboratories for understanding how northern populations might respond under future climate conditions[15,16]. Therefore, examining the drivers of decline in *P. sylvestris* forests at the rear-edge in the southeast Iberian Peninsula could be helpful in developing preventive forestry tasks[17,18].

[1]Department of Soil and Plant Microbiology, Estación Experimental del Zaidín, CSIC, Profesor Albareda 1, 18008, Granada, Spain. [2]Institute of Forest Sciences ICIFOR, INIA-CSIC. Ctra. La Coruña km 7.5, 28040, Madrid, Spain. [3]These authors contributed equally: Ana V. Lasa, Antonio J. Pérez-Luque. ✉e-mail: ana.vicente@eez.csic.es; manuel.fernandez@eez.csic.es

Although the model proposed by Manion (1981)[5] included biotic stressors, very little is known about the host tree-associated microbiota. Most of the works have focused on the role of several pests such as the pinewood nematode[19] and its interaction with abiotic factors[20], pine processionary moth (*Thaumetopoea pityocampa*[21]), even parasitic plants such as mistletoe[22], demonstrating all of them the positive correlation between the presence of such stressors and decline symptoms. Notwithstanding the above, shifts in the plant host-associated microorganisms have been almost overlooked so far. It is well-known that plant microbiota deploys essential functions in plant nutrition and defense against (a)biotic stresses[23]. To the best to our knowledge, we addressed for the first time the diversity, composition and structure of the microbial communities (bacteria and fungi) associated to the roots of *P. sylvestris* trees affected by forest decline[11]. Metabarcoding approaches based on amplicons of the bacterial 16S rRNA and fungal ITS2 sequences were employed in that study to analyze both rhizosphere and root endosphere microbial populations of decline-affected and asymptomatic Scots pines just during the spring. These data were deposited and are publicly available in two repositories: NCBI SRA (National Center for Biotechnology Information Sequence Read Archive) and GBIF (Global Biodiversity Information Facility). Furthermore, here we describe for the first time the dataset corresponding to the microbial communities dwelling in the rhizosphere soil and root endosphere of exactly the same trees as the previous work[11], under summer conditions. Both spring and summer datasets could provide valuable information about potential microbial bioindicators of pine decline. In addition, this dataset can serve as a valuable resource for other researchers seeking to identify specific microorganisms, even employ the specific sequences to develop DNA-based diagnosis methods of specific microorganisms, for instance, by developing probes targeting specific microorganisms for TaqMan qPCR approaches. Beyond diagnostics, the dataset may support the identification of microbial indicators linked to forest health or stress, providing a foundation for early-warning systems. Moreover, the data can offer insights into potential interactions with biotic and abiotic components of forest ecosystems, contributing to a better understanding of ecosystem functioning. Furthermore, the data and the corresponding metadata could be useful in developing integrated forest management and ecological restoration strategies, especially in the context of environmental change.

Methods

Experimental design. Three mountainous regions were selected for sample collection: Sierra de Almijara, Tejeda and Alhama (Sierra de Almijara hereinafter), Sierra Nevada and Sierra the Baza, all of them located in Granada (Andalusia, southeast Spain) (Fig. 1). In Sierra de Almijara, 12 unaffected *P. sylvestris* trees were chosen and marked (ASH01-12), while just decline-affected Scots pines were found in Sierra de Baza, and 12 individuals were also selected (BSD01-12). In contrast, 12 asymptomatic and 12 affected *P. sylvestris* trees were considered for the analyses in Sierra Nevada, which were named NSH and NSD, respectively. Those trees showing stunted growth, yellowing of the crown and/or defoliation were classified as symptomatic. Samplings were performed in summer 2021 and spring of 2022 (Table 1), selecting exactly the same trees in both seasons.

In each experimental area, rhizosphere soil samples and young roots were collected from two different points of the root system of the same tree, and processed in parallel as described below. Both rhizosphere and root samples coming from the same trees were mixed in further steps, obtaining a total of 192 bacterial and 192 fungal libraries [12 replicates × 2 plant compartments × 4 plant conditions (NSH, NSD, ASH, BSD) × 2 seasons = 192 samples]. Table 1 summarizes the main characteristics of the samples, while the experimental design is depicted in Fig. 2.

Sample collection and processing. At a distance of less than 50 cm from the trunk of the trees, the topsoil (25–50 cm) was removed by digging, in order to avoid the soil horizon rich in organic matter degrading fungi and small roots of herbaceous plants. Then, the main roots of the selected tree were followed until young poorly suberized roots were found. The soil closely adhered to these roots was collected by rubbing them manually. These roots were also collected, and both rhizosphere soil and the roots were kept at 4 °C until they were processed in the laboratory. It should be marked that two samples of rhizosphere soil and roots were taken from two different parts of the same root system of each tree (2 × 48 trees). 500 g of the soil close to the roots of each tree was also collected just in spring in order measure soil physico-chemical parameters. Soil samples were analyzed by Laboratorio Analítico Bioclínico (Almería, Spain) following standardized procedures, applying electrometric, ICP-MS (Inductively Coupled Plasma Mass Spectrometry), ion chromatographic, volumetric, spectrophotometric or densimetric methods.

Once in the laboratory, the roots were surface sterilized. For that purpose, roots were washed twice with 20 mL NaCl solution 0.8% (w/v). Then, they were rinsed five times with distilled water so that soil particles closely attached to the roots were removed. Roots were immersed in 50 mL an ethanol solution 70% (v/v) for five minutes, and then immersed in a solution (50 mL) composed of NaClO (4%, w/v) and Tween20 (0.01%, v/v) for three minutes. Roots were rinsed twice with distilled sterilized water for one minute each, and 100 µL water from the last rinse were plated onto LB (Luria Bertani) and TSA (Trypto-Casein Soy Agar) media containing plates. They were incubated at 30 °C for 14 days and visually evaluated periodically in order to check the sterilization efficiency.

Sterilized roots were then lyophilized at −45 °C until they were completely dehydrated, by employing the lyophilizer Thermo Savant Modulyo D-230 (Waltham, United States). Subsequently, root material was ground in the grinder MM-400 (Retsch, Germany) for 1 min at 30 Hz with the help of sterilized grinding balls of 20 mm diameter.

DNA extraction and sequencing. Total DNA was extracted from 0.1 g of sterilized root material and 0.25 of rhizosphere soil by means of the DNeasy® Plant Pro and DNeasy® PowerSoil® Pro kits (Qiagen, Germany), respectively, following the manufacturer's instructions. DNA extracted from the same trees was then pooled into

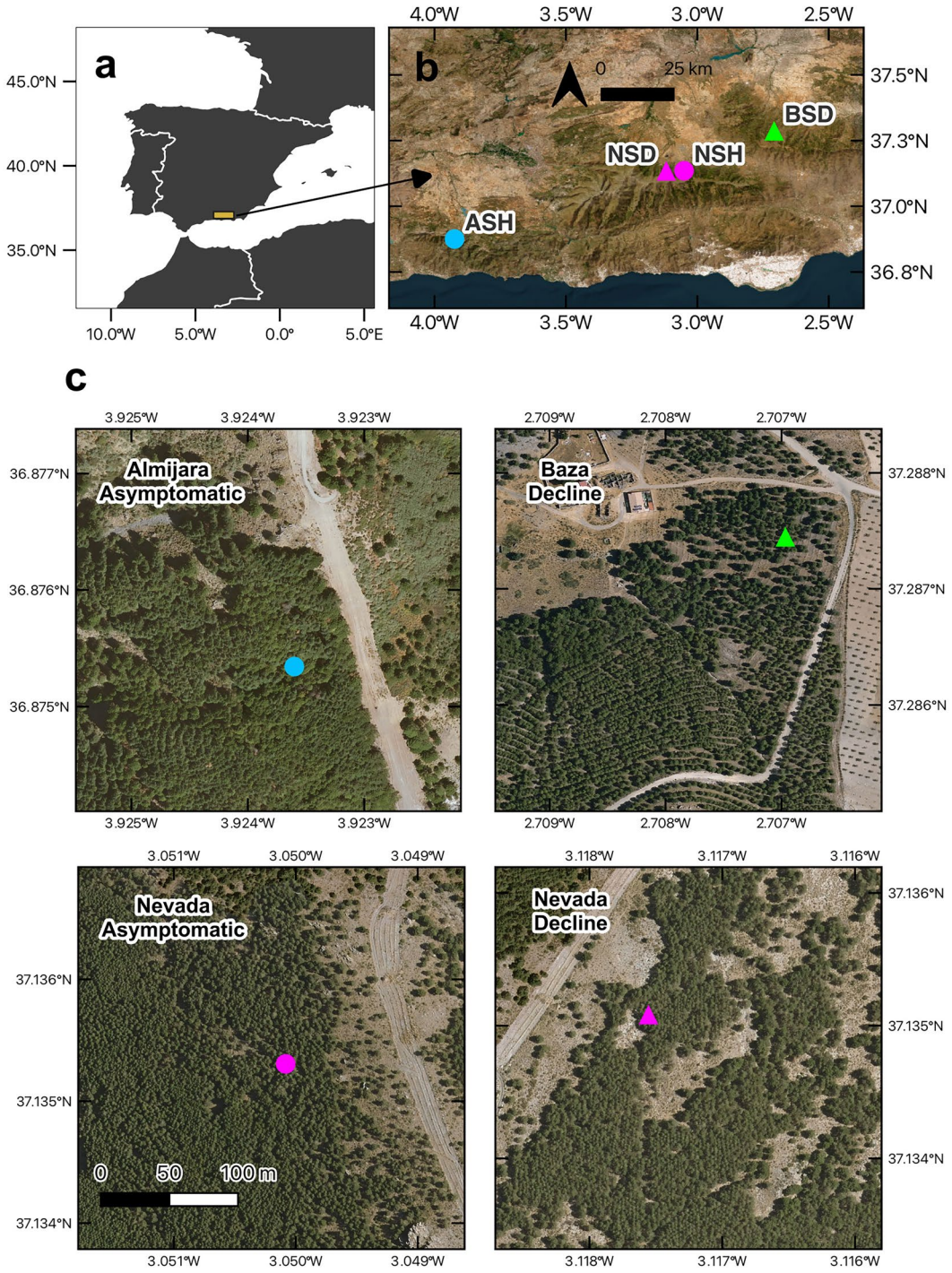

Fig. 1 General (**a,b**) and detailed (**c**) view of the study site's locations. Circles and triangles indicate asymptomatic and symptomatic, respectively, *P. sylvestris* plots. Green symbols indicate Sierra de Baza plot, purple corresponds with Sierra Nevada plots, and blue represents Sierra de Almijara plot.

a composite sample, for both rhizosphere soil and root endosphere samples. Thus, a total of 48 rhizosphere soil and 48 root endosphere composite samples were obtained in each season. DNA yields were determined by means of the fluorometer Qubit 3.0 (Life Technologies, Unites States).

DNA from each individual sample was sequenced by Illumina MiSeq platform at the genomics service of the Institute of Parasitology and Biomedicine López-Neyra (CSIC, Granada, Spain). The hypervariable regions V3-V4 of the bacterial 16S rRNA were sequenced by using the Pro341F (CCTACGGGNBGCASCAG) and Pro805R (GACTACNVGGGTATCTAATCC) primers[24]. Prior to the sequencing step, amplicons obtained from root endosphere were treated with PNA PCR clamps[25] so that the amplification of plant plastidic and mitochondrial DNA was diminished. For fungal dataset, primers ITS4[26] (GTGARTCATCGAATCTTTG) and

	ASH (n = 12)[a]	BSD (n = 12)[a]	NSH (n = 12)[a]	NSD (n = 12)[a]
Experimental area	Sierra de Almijara	Sierra de Baza	Sierra Nevada	
Coordinates	36° 52′ 53″ N 03° 55′ 31.4″ W	37° 17′ 20.7″ N 02° 42′ 20.0″ W	37° 08′ 07.1″ N 03° 03′ 00.3″ W	37° 08′ 06.3″ N 03° 07′ 03.2″ W
Decline symptoms	No	Yes	No	Yes
Replicates	01–12			
Sampling date	Spring	April 20th, 2022	April 27th, 2022	May 11th, 2022
	Summer	July 12th, 2021	July 26th, 2021	July 19th, 2021
Plant compartments	Root endosphere and rhizosphere soil			
Microorganisms under study	Bacterial and fungal communities			

Table 1. Main characteristics of all the groups of samples analyzed. [a]12 trees were sampled each sampling campaign (12 in summer and 12 in spring; the same trees were selected in both seasons).

fiTS7[27] (TCCTCCGCTTATTGATATGC) were employed for the sequencing of the fungal ITS2 region. In each sequencing run, three replicates of the mock community ZymoBIOMICS Microbial Community Standard II (logarithmic distribution; ZYMO Research, United States) were included as quality control of the sequencing process. It should be mentioned that 2 × 275 and 2 × 300 Paired-end (PE) sequencing strategy were followed (runs corresponding to spring and summer samples, respectively).

Bioinformatic processing of the sequencing data. Sequencing reads were processed in our previous work as described there[11]. Raw reads obtained by high-throughput sequencing were fully processed in R software v. 4.2.3.[28]. Unless otherwise stated, both fungal and bacterial reads were processed in the same way (Fig. 2). Firstly, Figaro software[29] was employed just for bacterial libraries with the aim of stablishing the best parameters of the following filtering step. The function *filterAndTrim* included in the package DADA2[30] was computed in order to remove low quality reads and those with ambiguities, by indicating the parameters selected previously with Figaro tool. Both primers were removed from the reads by employing Cutadapt bioinformatics tool[31], and in the case of the fungal library, the filter and trimming step was performed at this point. For both bacterial and fungal datasets, modelling of parametric errors and their correction was performed my means of the function *learnErrors* of package DADA2. Sample inference was then performed (function *dada* of the same package). In both cases, forward and reverse reads were merged (function *mergePairs* of the package DADA2, keeping the default parameters). Then, ASV (Amplicon Sequence Variant) tables were constructed for each sequencing run, and ASV tables corresponding to the runs of spring and summer samplings were merged in this work by using the function *mergeSequenceTables*. Chimeric sequences were then removed from the dataset by running the function *removeBimeraDenovo*, package DADA2. High-quality bacterial and fungal sequences were taxonomically classified by comparing them with those held in the Ribosomal Database Project (training set v. 18[32] and UNITE v. 7.2[33] databases, respectively. For that purpose, the function *assignTaxonomy* of the package DADA2 was implemented. Once sequences were classified, the sequencing detection limit was stablished. In this case, we checked the taxonomical composition of the samples corresponding to the mock communities. We found ASVs in that samples that were not part of the commercial mock community (whose composition is already known), and the relative abundance of these ASVs was considered the detection limit (here, 0.0018%). Hence, those ASVs which accounted for less than the established detection limit were removed from the dataset, as they may be potential biases of the sequencing process. Those bacterial sequences identified as Cyanobacteria/Chloroplast were compared to those held in GenBank database by BLASTn and removed in case there were identified as eukaryotic. Eventually, ASVs classified as mitochondria, chloroplasts, not classified at Kingdom level or identified as Eukaryota (just in the bacterial libraries) were not retained for further analyses.

Data Records

Data obtained from high-throughput sequencing platform were deposited and are publicly available at the NCBI SRA repository[34]. Compressed files which include raw forward and reverse reads (.fastq.gz) were deposited for each tree replicate, both for root endosphere and rhizosphere soil samples, both for fungi and bacteria, both for spring and summer sampling points. Thus, 792 fastq files are publicly available under the BioProject ID PRJNA993625. Supplementary Table 1 specifies the accessions that are unique to this work and those which were already reported in our previous work[11], and the specific metadata.

Integration in the global biodiversity information facility. The datasets were also accommodated to fulfill the Darwin Core Standard[35–37] with "*occurrence*" and "*measurement or Facts*" extensions (GBIF 2018). We used Darwin Core Archive Validator tool (http://tools.gbif.org/dwca-validator/) to check whether the dataset met Darwin Core specifications. Then, the dataset was deposited and published through the Integrated Publishing Toolkit (IPT v2.3.6)[38] of the Spanish node of the Global Biodiversity Information Facility (GBIF) (http://ipt.gbif.es). A version of the dataset is available at GBIF[39].

Structure of the darwin core archive (DwC-A). The structure of the *Darwin Core Archive (DwC-A)* encompasses comprehensive sampling event data[40] consisting of "*event-type*" data, "*occurrence*" data and "*measurement or fact*" type data (Supplementary Figure 1). This structure is organised based on a hierarchical framework for sampling events. The event file contains a total of 288 records distributed as follows: 48 events (*parent events*) that describe the spatial coverage of the sampled trees (see Table 1); 48 events corresponding to the soil

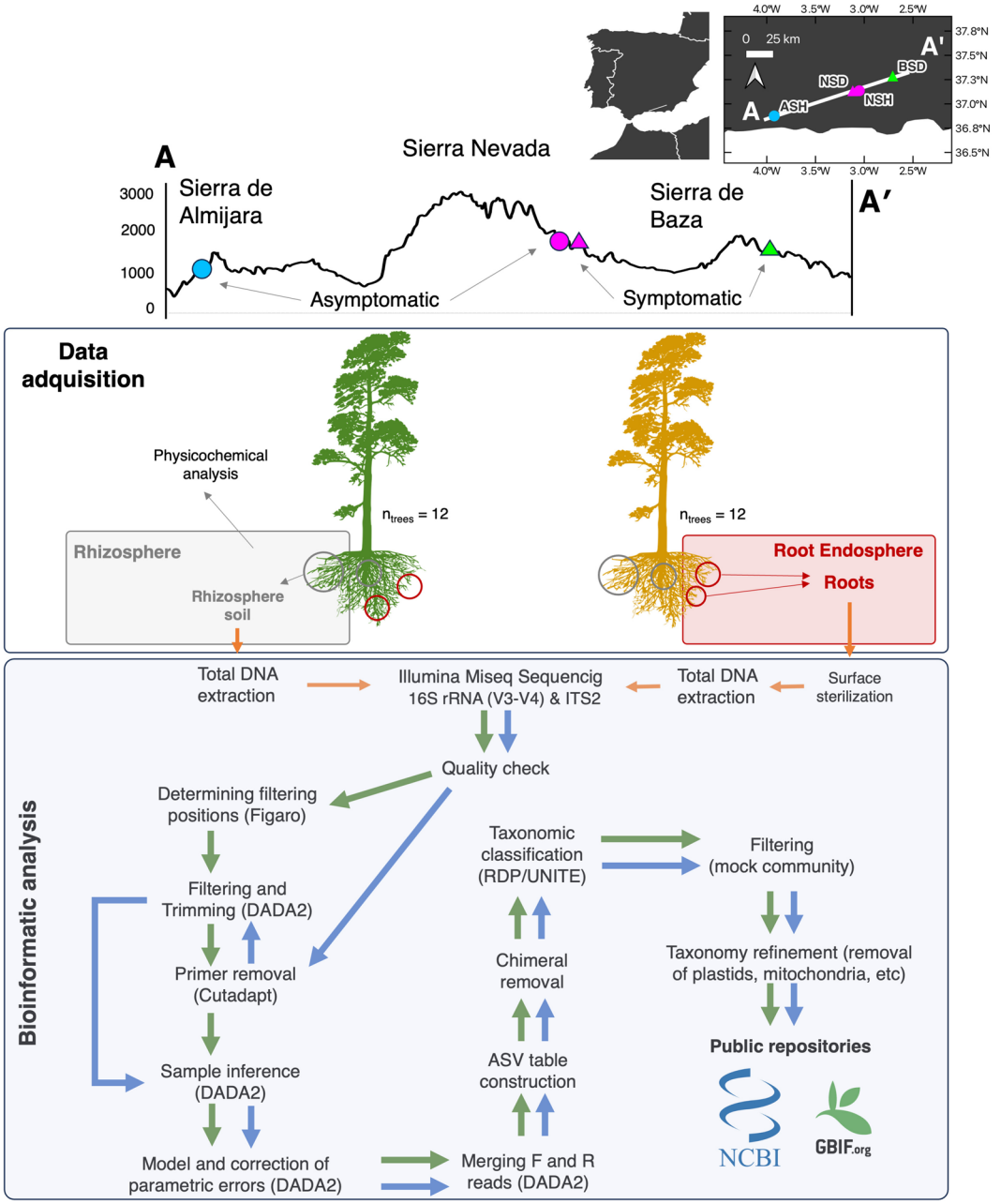

Fig. 2 Experimental workflow. Orange arrows indicate steps common to fungal and bacterial data acquisition. Green and blue arrows indicate steps that were solely performed for bacterial and fungal data, respectively. Green symbol indicates Sierra de Baza plot, purple corresponds with Sierra Nevada plots, and blue represents Sierra de Almijara plot. ASV, *Amplicon Sequence Variants*, Figaro, DADA2 and Cutadapt are the names of the bioinformatics tools employed in the corresponding steps (see the text for more details). "A" letter denotes the altitude measured in meters above sea level.

sampling in a specific date (those events are related to the parent events); 192 events corresponding at the root endosphere (n = 96) and the rhizosphere (n = 96) samplings for spring and summer. This structure allows to append new records in future sampling campaign. The soil, root endosphere and rhizosphere events contain a *parentEventID* field, linking them to their respective trees. The occurrence file comprises 79640 records, while the "*measurement or fact*" file encompasses 672 records of various soil measurements associated with each of the trees in a determined date.

Taxonomic Coverage

The taxonomic depth achieved in this dataset was considerable, covering two major kingdoms: Bacteria and Fungi. Bacteria (29.4% of the total sequences), are represented by 16 classified phyla, with the most abundant being *Proteobacteria* in the rhizosphere (22.6–36.4%, see Fig. 3a) and *Tenericutes* (0–90%) or *Actinobacteria* in the root endosphere (6.1–56.7%). Fungi (70.6% of the total sequences) are represented by 8 taxonomically

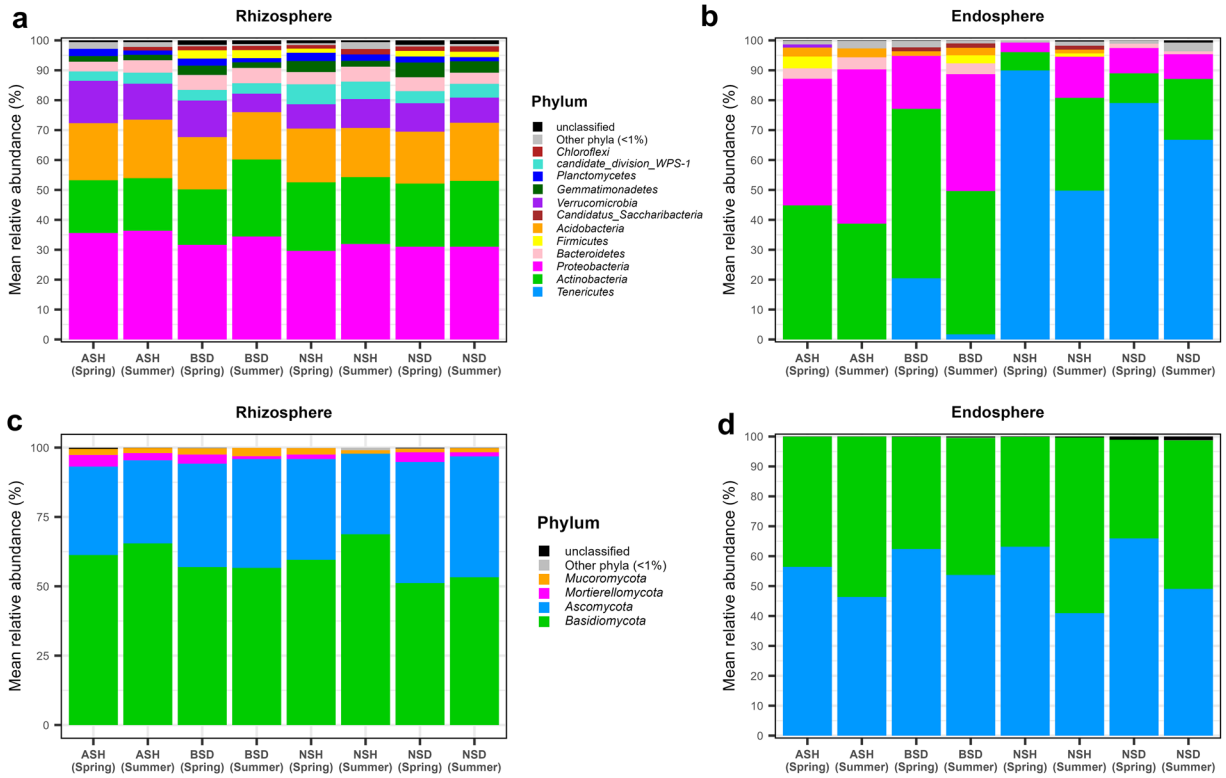

Fig. 3 Mean relative abundance of the main bacterial phyla detected in the rhizosphere soil (**a**) and root endosphere (**b**), and main fungal phyla inhabiting the rhizosphere soil and root endosphere (**c,d**, respectively). The artificial group "Other phyla (<1%)" emcompasses those phyla accounting less than 1% in at least one group of samples. ASH: Almijara samples (Asymptomatic), BSD: Baza samples (Symptomatic), NSD: Sierra Nevada samples (Symptomatic), NSH: Sierra Nevada samples (Asymptomatic).

classified phyla, primarily dominated by *Basidiomycota* in the rhizosphere (51.2–68.7%, see Fig. 3c) and *Ascomycota* (41–65.9%) and *Basidiomycota* (33–58.7%). The dataset covers a total of 74 microbial classes, 156 orders, 309 families, and 580 genera. Figure 4 summarizes main bacterial and fungal genera.

Technical Validation

We registered the total amount of raw sequences per sample as quality control of the sequencing process, being the minimum number 41580 and 190 sequences in the dataset corresponding to spring and summer samples, respectively (Supplementary Table 2). All the reads with at least one ambiguity, and with an associated quality score per nucleotide below 2 were removed from the analysis, as well as those shorter than 50 bp.

After the overlapping of forward and reverse reads and chimera removal step, the taxonomy of the ASVs detected in the mock community samples was checked. In these samples, sequences corresponding to all the taxa included in the mock community were detected with the exception of *Lactobacillus*, the yeast *Saccharomyces* and the fungus *Cryptococcus* (the mock community was just sequenced with the primers amplifying the bacterial 16S rRNA). It should be mentioned that two ASVs belonging to the genus *Limosilactobacillus* and ASVs assigned to *Bacteroides*, *Dialister* and *Zea* (none of them included in the commercial mock community) were also detected in the mock community samples. Notwithstanding, the sequences corresponding to *Limosilactobacillus* showed 100% similarity to both *Limosilactobacillus* and *Lactobacillus* sequences held in the GenBank database included in the NCBI, and 100% similarity to *Limosilactobacillus* sequences included in the EzBioCloud database[41]. Taking into account that some species of *Lactobacillus* were reclassified into *Limosilactobacillus* genus in 2020[42], we considered that the ASVs belonging to the former genus were members of the mock community. Moreover, ASVs belonging to *Limosilactobacillus*, *Bacteroides* and *Dialister* were just detected in the mock community but not in the rhizosphere and root endosphere samples, thus, they were considered as potential biases included in the samples corresponding to the commercial mock community. Sequences corresponding to *Zea* (Eukaryotic, Plantae) were not considered for further analyses. Samples corresponding to root endosphere of trees ASH10, BSD10, BSD11 and NSH11 (spring) and those corresponding to the root endosphere of trees ASH01, ASH10, ASH12, BSD03, BSD04, BSD10, NSD01 and NSH11 (summer) were removed from the bacterial dataset since they did not harbor enough quality reads.

In order to verify that the sampling and sequencing efforts were enough to obtain high quality sequences, we computed the rarefaction curves depicted in Supplementary Figure 2 with the help of the function *rarecurve* (package vegan[43]). Practically all the curves reached to the asymptote, indicating that almost all the microbial diversity was covered and the samples were representative of ecosystems under study.

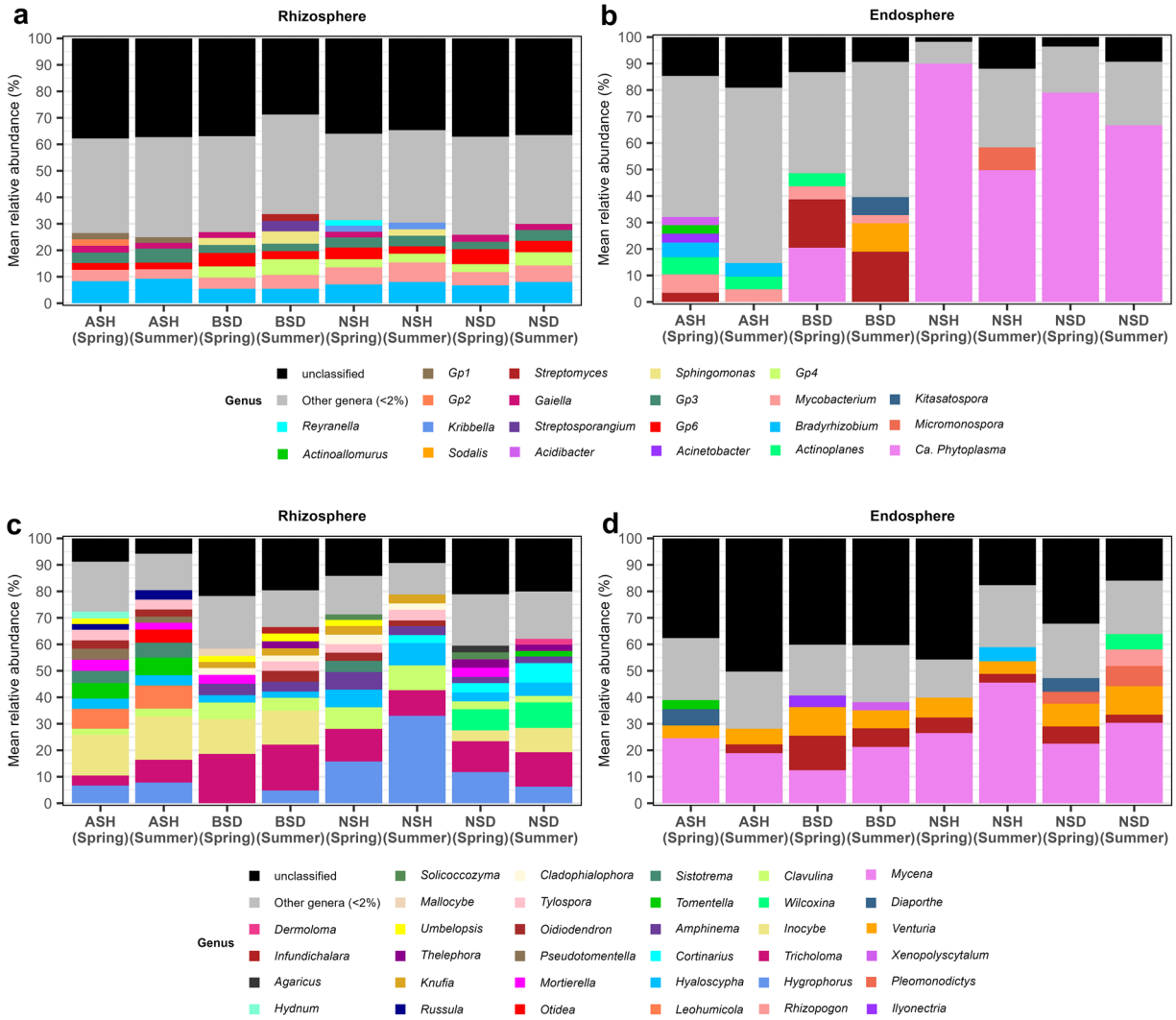

Fig. 4 Mean relative abundance of the main bacterial genera detected in the rhizosphere soil (**a**) and root endosphere (**b**), and main fungal genera inhabiting the rhizosphere soil and root endosphere (**c,d**, respectively). The artificial group "Other genera (<2%)" emcompasses those genera accounting less than 2% in at least one group of samples. ASH: Almijara samples (Asymptomatic), BSD: Baza samples (Symptomatic), NSD: Sierra Nevada samples (Symptomatic), NSH: Sierra Nevada samples (Asymptomatic).

Code availability

All the scripts used for the processing and analyses of the sequencing data are openly available to ensure full reproducity of the workflow. This includes scripts for the quality control and filtering of Illumina (MiSeq) reads, processing of ASV tables for both bacterial and fungal datasets, and formatting of data into DarwinCore Archive structure for integration into GBIF. The full workflow is documented and hosted in a GitHub repository (https://anitalasa.github.io/Psylvestris_DataPaper/), and permanently archived in Zenodo[44]. By providing a fully documented and reproducible workflow, this dataset complies with FAIR principles[45] and facilitates its application in microbial ecology, forest health monitoring, and biodiversity research. Furthermore, all the repositories are under license CC BY.

Received: 19 September 2024; Accepted: 15 April 2025;
Published online: 28 May 2025

References

1. Allen, C. D. *et al.* A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For. Ecol. Manage.* **259**, 660–684 (2010).
2. Hammond, W. M. *et al.* Global field observations of tree die-off reveal hotter-drought fingerprint for Earth's forests. *Nat. Commun.* **13**, 1761 (2022).
3. Hartmann, H. *et al.* Climate change cisks to clobal forest health: emergence of unexpected events of elevated tree mortality worldwide. *Annu. Rev. Plant Biol.* **73**, 673–702 (2022).

4. Anderegg, W. R. L. Tree mortality from drought, insects, and their interactions in a changing climate. *New Phytol.* **208**, 674–683 (2015).
5. Manion, P.D. *Tree Disease Concepts* 2nd edn (Prentice Hall Inc., 1991)
6. Hartmann, H. *et al.* Monitoring global tree mortality patterns and trends. Report from the VW symposium 'Crossing scales and disciplines to identify global trends of tree mortality as indicators of forest health'. *New Phytol.* **217**, 984–987 (2018).
7. Galiano, L., Martinez-Vilalta, J. & Lloret, F. Drought-induced multifactor decline of scots pine in the Pyrenees and potential vegetation change by the expansion of co-occurring oak species. *Ecosystems.* **13**, 978–991 (2010).
8. Vilà-Cabrera, A., Martinez-Vilalta, J., Galiano, L. & Retana, J. Patterns of Forest decline and regeneration across scots pine populations. *Ecosystems.* **16**, 323–335 (2013).
9. Gonthier, P., Giordano, L. & Nicolotti, G. Further observations on sudden diebacks of Scots pine in the European Alps. *The Forestry Chronicle.* **86**, 110–117 (2010).
10. Navarro-Cerrillo, R. M. *et al.* Growth decline assessment in Pinus sylvestris L. and Pinus nigra Arnold. forests by using 3-PG model. *For. Syst.* **25**, e068 (2016).
11. Lasa, A. V. *et al.* Mediterranean pine forest decline: a matter of root-associated microbiota and climate change. *Sci. Total. Environ.* **926**, 171858 (2024).
12. Hampe, A. & Petit, R. J. Conserving biodiversity under climate change: The rear edge matters. *Ecol. Lett.* **8**, 461–467 (2005).
13. Holt, R. B. & Kleitt, T. H. Species' borders: a unifying theme in ecology. *Oikos.* **108**, 3–6 (2005).
14. Jump, A. S., Calvin, L. & Hunter, P. D. Monitoring and managing responses to climate change at the retreating range edge of forest trees. *J. Environ. Monit.* **12**, 1791–1798 (2010).
15. Matías, L. & Jump, A. S. Interactions between growth, demography and biotic interactions in determining species range limits in a warming world: The case of *Pinus sylvestris*. *For. Ecol. Manage.* **282**, 10–22 (2012).
16. Martínez-Vilalta, J. *et al.* Iberian scots pine populations under climate change: some don't like it hot. *Ecosistemas.* **21**, 15–21 (2012).
17. Benavides, R. *et al.* Direct and Indirect Effects of Climate on Demography and Early Growth of *Pinus sylvestris* at the Rear Edge: Changing Roles of Biotic and Abiotic Factors. *PLoS One.* **8**, e59824 (2013).
18. Herrero, A., Rigling, A., Zamora, R. Varying climate sensitivity at the dry distribution edge of *Pinus sylvestris* and *P. nigra*. *For. Ecol. Manage.* **308**, 40–61.
19. Proença, D. N. *et al.* The microbiome of endophytic, wood colonizing bacteria from pine trees as affected by pine wilt disease. *Sci. Rep.* **7**, 4205 (2017).
20. Calvão, T., Duarte, C. M. & Pimentel, C. S. Climate and landscape patterns of pine forest decline after invasion by the pinewood nematode. *For. Ecol. Manage.* **433**, 43–51 (2019).
21. Azcárate, F. M., Seoane, J. & Silvestre, M. Factors affecting pine processionary moth (Thaumetopoea pityocampa) incidence in Mediterranean pine stands: a multiscale approach. *For. Ecol. Mange.* **529**, 120728 (2023).
22. Gea-Izquierdo, G. *et al.* Synergistic abiotic and biotic stressors explain widespread decline of *Pinus pinaster* in a mixed forest. *Sci. Total. Environ.* **685**, 963–975 (2019).
23. Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T. & Singh, B. K. Plant-microbiome interactions: from community assembly to plant health. *Nat. Rev.* **18**, 607–621 (2020).
24. Takahashi, S., Tomita, J., Nishioka, K., Hisada, T. & Nishijima, M. Development of a prokaryotic universal primer for simultaneous analysis of bacteria and archaea using next-generation sequencing. *PloS One.* **9**, 8 (2014).
25. Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D. & Dangl, J. L. Practical innovations for high-throughput amplicon sequencing. *Nat. Methods.* **10**, 999–1002 (2013).
26. White, T. J., Bruns, T. D., Lee, S. B. & Taylor, J. W. in PCR *Protocols: A Guide to Methods and Applications*. (eds. Innis, M. A., Gelfand, D. H., Sninsky, J. F. & White, T. J.) **Ch**. **38** (Academics Press, 1990).
27. Ihrmark, K. *et al.* New primers to amplify the fungal ITS2 region – evaluation by 454- sequencing of artificial and natural communities. *FEMS Microbiol. Ecol.* **82**, 666–677 (2012).
28. R Core Team. R: A language and environment for statistical computing, version 4.2.2. https://www.R-project.org/ (2022).
29. Sasada, R., Weinstein, M., Prem, A., Jin, M. & Bhasin, J. FIGARO: an efficient and objective tool for optimizing microbiome rRNA gene trimming parameters. *J. Biomol. Tech.* **31**, S2 (2020).
30. Callahan, B. J. *et al.* DADA2: high-resolution simple inference from Illumina amplicon data. *Nat. Methods.* **13**, 581–583 (2016).
31. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
32. Cole, J. R. *et al.* Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42**, 633–642 (2014).
33. Abarenkov, K., *et al.* UNITE general FASTA release for eukaryotes, version 7.2. https://unite.ut.ee/ (2022).
34. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP448600 (2022).
35. Wieczorek, J. *et al.* Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE.* **7**, e29715 (2012).
36. Darwin Core Maintenance Group List of Darwin Core terms. Biodiversity Information Standards (TDWG). http://rs.tdwg.org/dwc/doc/list/ (2021a).
37. Darwin Core Maintenance Group Darwin Core quick reference guide. Biodiversity Information Standards (TDWG). https://dwc.tdwg.org/terms/ (2021b).
38. Robertson, T. *et al.* The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. *PLoS ONE.* **9**, e102623 (2014).
39. Lasa, A. V., Pérez-Luque, A. J. & Fernández-López, M. Root-associated microbiota of decline-affected and asymptomatic *Pinus sylvestris* trees. *GBIF* https://doi.org/10.15470/dardca (2024).
40. GBIF Best Practices in Publishing Sampling-event data, version 2.2. *Copenhagen: GBIF Secretariat* https://ipt.gbif.org/manual/en/ipt/3.0/best-practices-sampling-event-data (2018).
41. Chalita, M. *et al.* EzBioCloud: a genome-driven database and platform for microbiome identification and discovery. *Int J Syst Evol Microbiol.* **74**, 006421 (2024).
42. Zheng, J. *et al.* A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol.* **70**, 2782–2858 (2020).
43. Oksanen J. *et al. vegan: Community Ecology Package* https://CRAN.R-project.org/package=vegan (2022).
44. Lasa, A. V. & Pérez-Luque, A. J. Code repository of the DataPaper: Root-associated microbiota of decline-affected and asymptomatic *Pinus sylvestris* trees. *Zenodo* https://doi.org/10.5281/zenodo.13711241 (2024).
45. Wilkinson, M. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).

Acknowledgements

This work was supported by the MICINN through European Regional Development Fund [SUMHAL, LIFEWATCH-2019-09-CSIC-13, POPE 2014-2020]. AVL is hired under the Generation D initiative, promoted by Red.es, an organisation attached to the Ministry for Digital Transformation and the Civil Service, for the attraction and retention of talent through grants and training contracts, financed by the Recovery, Transformation and Resilience Plan through the European Union's Next Generation funds. AJPL is currently funded by MCIN/AEI/10.13039/ 501100011033 and by "European Union NextGenerationEU/PRTR" with a "Juan de la Cierva"

fellowship programme (Grant JDC2022- 050056-I). We also thank to Katia Cezón (Spanish GBIF node-CSIC) for technical support in the data integration in GBIF.

Author contributions

A.V.L., A.J.P.L., M.F.L. designed the experiment; A.V.L. and A.J.P.L. analysed the data; A.V.L. and A.J.P.L. validated the data; A.V.L. and A.J.P.L. deposited the data in public repositories; A.V.L. and A.J.P.L. wrote the manuscript; M.F.L. was the responsible for acquiring the funding and for the management of the project. A.V.L., A.J.P.L. and M.F.L. reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-05013-9.

Correspondence and requests for materials should be addressed to A.V.L. or M.F.-L.