Università degli Studi di Torino
Department of Life Sciences and Systems Biology

PhD Programme in Biological Sciences and Applied Biotechnologies

Soils associated to herbarium plants: a resource to address the temporal evolution of plant-associated microbiomes

Gianluca Grasso

Academic Years: 2021-2024

Tutors: **Dr.ssa Valeria Bianciotto, Dr. Roland Marmeisse, Prof.ssa Silvia Perotto**

Muséum National d'Histoire Naturelle
Institut de Systématique, Évolution, Biodiversité

PhD Programme Sciences de la Nature et de l'Homme: Évolution et Écologie

Soils associated to herbarium plants: a resource to address the temporal evolution of plant-associated microbiomes

PhD thesis of the Università degli Studi di Torino and the Muséum National d'Histoire Naturelle in Paris.
Turin, 2024

Cover Figure :
Herbarium specimens of *Avena sativa* of Herbier Roland Bonaparte at the University of Lyon 1 collected on 28 August 1909. Image available on ReCOLNAT (https://explore.recolnat.org/) with the accession number: LY0642300

UNIVERSITÀ
DI TORINO

Università degli Studi di Torino

PhD Programme in Biological Sciencesand Applied Biotechnologies

XXXVII° Cycle: 2021/2024

Muséum National d'Histoire Naturelle

PhD Programme Sciences de la Nature et de l'Homme: Évolution et Écologie

Soils associated to herbarium plants: a resource to address the temporal evolution of plant-associated microbiomes

Tutors:
Dr.ssa Valeria Bianciotto
Dr. Roland Marmeisse
Prof.ssa Silvia Perotto

Candidate:
Gianluca Grasso

Members of Thesis Board:

President :	Prof. Guido LINGUA	Professor, Università del Piemonte Orientale, Italy
Rapporteaurs :	Prof.ssa Laura PARDUCCI	Professor, Università Roma Sapienza, Italy
	Prof. Christophe MOUGEL	Research Director, INRAe Rennes, France
Examinateurs :	Dr.ssa Valeria BIANCIOTTO - Supervisor	Research Director, CNR-IPSP Torino, Italy
	Dr. Roland MARMEISSE – Supervisor	Research Director, CNRS-MNHN Paris, France
	Dr.ssa Beatrice DEMARCHI	Professor, University of Torino, Italy
	Dr. Eric CHENIN	Research engineer, IRD, Bondy, France

Index

Abstracts

English version

The current biodiversity crisis has been essentially documented for plants and animals, thanks to the availability of historical data their populations and communities, in some cases dating back several centuries. What about microorganisms, and more specifically soil microorganisms? In the absence of archives from the past, the only available data on the impacts of current environmental changes on microbial diversity come mainly from short-term impact studies comparing soils subjected to different management practices.

We observed that plant roots preserved in herbariums are occasionally surrounded by a sheath of rhizospheric soil that may have preserved traces of past microbial communities. In this thesis, we investigated which plants are best suited for preserving soil in herbarium collections and how these can be used to study past microbiomes and understand the effects of climate change and the intensification of agricultural practices that have occurred over the past two centuries.

Through a citizen science project (*Les Herbonautes*), we created a curated database of herbarium specimens of major crop plants (more than 2000 specimens in 14 species) cultivated in France. It contains information on the presence or absence of roots and soil in herbaria. It revealed that several species, particularly in the *Poaceae*, are better suited to preserve soil compared to plants with simpler root systems (like *Fabaceae*) or large plants (like *Zea mays* or *Cannabis*). DNA extraction performed on 30 herbarium soil samples associated to *Triticum* and *Hordeum* plants (collected between 1820 and 1980) allowed us to identify chemical and physical soil parameters that seem to affect soil DNA preservation.

Thus using herbarium soils, we developed a molecular paleomicrobiological approach to study the diversity of ancient soil microbial communities. It is based on the extraction of ancient DNA from this material and on its systematic sequencing. Annotation of the resulting sequences illustrated the taxonomic of past rhizospheric microbial communities associated to several crop species (*Avena, Secale, Triticum, Lactuca*). Our results do suggest that herbarium soil seem to preserve traces of their original soil microbiome, which are very similar (in composition, identity of the most abundant taxa, alpha and beta diversity) to modern soil microbiomes. This approach should thus allow us to assess the long-term impact of global changes on these communities such as the intensification of farming practices and climate change.

French version

La crise actuelle de la biodiversité a été largement documentée pour les plantes et les animaux, grâce à la disponibilité de données historiques sur leurs populations et communautés, datant pour certaines de plus d'un siècle. Qu'en est-il des micro-organismes, et plus spécifiquement des micro-organismes du sol ? En l'absence d'archives, les seules données disponibles sur l'impact des changements environnementaux sur la diversité microbienne proviennent d'études d'impact à court terme comparant des sols soumis à différentes pratiques et modes de gestion.

Nous avons observé que les racines des plantes conservées dans les herbiers sont parfois entourées de sol rhizosphérique qui pourrait avoir préservé des traces des communautés microbiennes passées. Au cours de cette thèse, nous avons étudié quelles étaient les plantes les plus adaptées à la conservation du sol dans les collections d'herbiers et comment celles-ci pouvaient être utilisées pour étudier les microbiomes du passé et ainsi comprendre les effets du changement climatique et de l'intensification des pratiques agricoles sur ces communautés.

Au travers d'un projet de science participative (*Les Herbonautes*), il a été créé une base de données des principales plantes cultivées en France (plus de 2000 spécimens de 17 espèces) mentionnant la présence ou l'absence de racines et de sol dans les exemplaires d'herbiers. Ces annotations montrent que certaines espèces, notamment dans les *Poaceae*, retiennent plus fréquemment du sol par rapport aux plantes avec des systèmes racinaires plus simples (comme les *Fabaceae*) ou aux plantes de grandes tailles (*Zea mays* ou *Cannabis*). L'extraction d'ADN réalisée sur 30 échantillons de sol d'herbiers associés à des *Triticum* et *Hordeum* (collectés entre 1820 et 1980) nous a permis d'illustrer quels facteurs physico-chimiques du sol impactent sur la conservation de l'ADN dans les sols.

Nous avons ainsi développé une approche paléomicrobiologique moléculaire, basée sur l'extraction et le séquençage de l'ADN ancien des sols d'herbier. L'annotation des séquences a révélé la diversité taxonomique des communautés microbiennes rhizosphériques passées associées à quatre espèces cultivées (*Avena, Secale, Triticum, Lactuca*). Nos résultats suggèrent que les sols d'herbier semblent avoir préservés les traces de leurs microbiomes originels qui sont très similaires (en composition, identité des taxa majeurs, diversités alpha et béta) des microbiomes de sols modernes. Cette approche nous permettra ainsi d'évaluer l'impact sur le long terme des changements globaux, tels que l'intensification des pratiques agricoles et le changement climatique, sur ces communautés microbiennes.

Italian version

L'attuale crisi della biodiversità è stata essenzialmente documentata per quanto riguarda piante e animali, grazie alla disponibilità di dati storici sulle popolazioni e sulle comunità, in alcuni casi risalenti a più di un secolo fa. Tuttavia, ciò che è accaduto ai microrganismi, in particolare al microbioma del suolo, negli ultimi decenni rimane in gran parte sconosciuto. In assenza di dati storici sui microrganismi, le uniche informazioni disponibili sull'impatto degli attuali cambiamenti climatici sulla diversità microbica provengono da studi di breve durata, che confrontano suoli sottoposti a diverse pratiche di gestione agricola.

Abbiamo osservato che le radici delle piante conservate negli erbari sono talvolta circondate da suolo che potrebbe aver conservato tracce molecolari delle comunità microbiche del passato. Nel corso di questa tesi, abbiamo studiato quali piante siano più adatte a conservare il suolo nelle collezioni d'erbario e come queste possano essere utilizzate per studiare i microbiomi del passato.

Grazie a un progetto di citizen science (*Les Herbonautes*), è stato possibile creare un database delle principali piante coltivate in Francia (più di 2000 campioni d'erbario di 14 specie), contenente informazioni sulla presenza o meno di radici e suolo nei campioni d'erbario. Questo ci ha permesso di osservare che piante come le *Poaceae* sono in grado di conservare meglio il suolo rispetto a piante con un apparato radicale più semplice (come le *Fabaceae*) o a piante di grandi dimensioni (*Zea mays* o *Cannabis*). Inoltre, l'estrazione del DNA da 30 campioni d'erbario di suoli associati a due specie di Poaceae (*Triticum* e *Hordeum* raccolti tra il 1820 e il 1980) ci ha permesso di studiare quali fattori chimici e fisici del suolo sembrano avere la maggiore influenza sulla conservazione del DNA.

Parallelamente, utilizzando i suoli d'erbario, è stato sviluppato un approccio paleomicrobiologico molecolare basato sull'estrazione del DNA e sul sequenziamento metagenomico. L'annotazione delle sequenze ha permesso di investigare la composizione tassonomica delle comunità microbiche rizosferiche del passato associate a diverse piante coltivate (*Avena, Secale, Triticum e Lactuca*). È stato possibile dimostrare che i microbiomi degli erbari sembrano conservare tracce del microbioma originario del suolo e sono molto simili (per composizione, alfa e beta diversità) ai microbiomi moderni del suolo.

In futuro, il suolo conservato negli erbari ci permetterà di valutare l'impatto a lungo termine dei cambiamenti globali, come l'intensificazione delle pratiche agricole e i cambiamenti climatici, sulle comunità microbiche del suolo.

Short abstracts

English version

The current biodiversity crisis is well documented thanks to historical data showing how human activities have significantly affected plants and animals. But what do we know about soil microorganisms? Unfortunately, very little; it remains largely a mystery. Without archives from the past, information on the effects of environmental change on microbial diversity comes mainly from short-term studies and fails to describe what has really happened to the billions of microorganisms living in the soil over the last 200 years.

To fill this knowledge gap, we discovered that plants preserved in herbaria can reveal traces of past microbial communities thanks to the study of soil trapped in their root systems. Through a citizen science project called "*Les Herbonautes*", we have created a database of over 2000 herbarium plants cultivated in France and collected information on the presence of roots and soil. We found that plants such as wheat (*Poaceae*) with complex root systems conserve soil better than plants with simpler roots such as legumes (*Fabaceae*) or plant of larger size such as maize or cannabis. Furthermore, by extracting DNA from herbarium soil samples collected between 1820 and 1980, we were able to investigate the main factors influencing DNA conservation in this ancient material. Finally, we developed a method to study these ancient microorganisms showing that those preserved in herbarium soils are very similar to today's once.

The study of herbarium soils will allow us to assess the long-term impact of global changes on soil microorganisms and help us understand the changes they have undergone over the past centuries.

French version

La crise actuelle de la biodiversité est bien documentée grâce à la présence de données historiques qui montrent comment l'activité humaine a eu un fort impact sur les plantes et les animaux. Mais que savons-nous des microorganismes du sol? Malheureusement, très peu ; cela reste en grande partie un mystère. Sans archives du passé, les informations sur les effets des changements environnementaux sur la diversité microbienne proviennent principalement d'études à court terme et ne parviennent pas à décrire ce qui est réellement arrivé au cours des 200 dernières années aux milliards de microorganismes qui vivent dans le sol.

Pour résoudre ce mystère, nous avons découvert que les plantes conservées dans les herbiers peuvent révéler des traces des communautés microbiennes du passé grâce à l'étude du sol qui peut rester piégé dans leurs systèmes racinaires. Grâce à un projet de science participative appelé "*Les Herbonautes*", nous avons créé une base de données de plus de 2000 plantes cultivées en France, recueillant des informations sur la présence de racines et de sol dans les herbiers. Nous avons observé que des plantes comme le blé (*Poaceae*), avec des racines complexes, sont plus aptes à préserver le sol par rapport à des plantes avec des racines plus simples comme les légumineuses (*Fabaceae*) ou des plantes de plus grande taille comme le maïs ou le cannabis. De plus, grâce à l'extraction de l'ADN à partir d'échantillons de sol d'herbier recueillis entre 1820 et 1980, il a été possible d'étudier les principaux facteurs qui influencent la conservation de l'ADN dans ce matériau ancien. Enfin, nous avons développé une méthode pour étudier les microorganismes du passé, montrant que ceux conservés dans les herbiers sont très similaires à ceux des sols modernes.

L'utilisation du sol des herbiers nous permettra d'évaluer l'impact à long terme des changements globaux sur les microorganismes du sol, nous permettant de comprendre les changements qu'ils ont subis au cours des derniers siècles.

Italian version

L’attuale crisi della biodiversità è ben documentata grazie alla presenza di dati storici che mostrano come l’attività umana abbia avuto un forte impatto sulle piante e gli animali. Ma cosa sappiamo dei microrganismi del suolo? Sfortunatamente poco, in gran parte resta un mistero. Senza archivi dal passato, le informazioni sugli effetti dei cambiamenti ambientali sulla diversità microbica provengono principalmente da studi a breve termine e non riescono a descrivere cosa è realmente successo ai miliardi di microrganismi che vivono nel suolo negli ultimi 200 anni.

Per risolvere questo mistero, abbiamo scoperto che le piante conservate negli erbari possono rivelare tracce delle comunità microbiche del passato attraverso lo studio del suolo che può rimanere intrappolato nelll’apparato radicale. Grazie a un progetto di citizen science chiamato "Les Herbonautes", abbiamo creato un database di oltre 2000 piante coltivate in Francia, raccogliendo informazioni sulla presenza di radici e suolo negli erbari. Abbiamo scoperto che piante come l grano (Poaceae), con un’architettura radicale complessa, sono più adatte a preservare il suolo rispetto a piante con radici più semplici come i legumi (Fabaceae) o piante di maggiori dimensioni come il mais o la Cannabis. Inoltre, attraverso l’estrazione del DNA da campioni di suolo d'erbario raccolti tra il 1820 e il 1980 è stato possibile studiare i principali fattori che influenzano la conservazione del DNA all’interno di questo materiale antico. Infine, abbiamo sviluppato un metodo per studiare i microganismi del passato, mostrando che quelli conservati negli erbari sono molto simili a quelli moderni.

L’utilizzo del suolo degli erbari ci permetterà di valutare l'impatto a lungo termine dei cambiamenti globali sui microganismi del suolo, permettendoci di comprendere i cambiamenti che hanno subito nel corso degli ultimi secoli.

Context and Objectives of the thesis

Plant microbiomes in the anthropocene era

Plant-associated microorganisms (the "plant microbiome") encompass a wide range of harmful (pathogens), beneficial (e.g. mutualistic symbionts) and apparently "neutral" (e.g. asymptomatic endophytes) taxa belonging essentially, but not exclusively, to the Viruses, Bacteria and Fungi. While traditionally studied individually in research projects focusing on one plant and one symbiont or one pathogen, it is now widely accepted that plant health and fitness cannot be fully appreciated without considering its interaction with its whole microbiome (Busby et al., 2017). Plant-microbiome association results in a dense network of relationships and is currently described with the term 'holobiont', that considers this complex host-microbiome system as a single biological entity that can express extended phenotypes that the plant alone could not express (Chialva & Bonfante, 2018). Systemic acquired resistance, where plants acquire resistance to pathogens following stimulation by other interacting microorganisms represents a classical example of an extended phenotype (Selosse et al., 2014). More recent studies however demonstrate that plant-microbiome interactions cannot be restricted to the field of plant disease. These interactions can also affect agronomically important traits as diverse as flowering time or even the expression of heterosis in maize hybrid lines (Lu et al., 2018; Wagner et al., 2021).

In this framework, research projects have emerged to understand how plant recruit their microbiomes and how this process is affected by the environment including local microbial diversity and to identify key microbial taxa affecting, the most, plant phenotypes (Busby et al., 2017). In the context of plant cultivation, it is expected that manipulation of the plant microbiome could ultimately represent a mean to manipulate plant health and to optimize agronomic characters while minimizing the use of inputs, either pesticides or mineral fertilizers (Agoussar & Yergeau, 2021). Recently it was also demonstrated that agricultural practices strongly impact the communities of plant-associated microorganisms, not only from a purely taxonomic point of view, but also from a functional one. For example, in the case of soil fungi, agricultural intensification has been shown to reduce the complexity of symbiont networks (Banerjee et al., 2019) and the balance between pathogens and mutualists (Lekberg et al., 2021). In the case of soil bacteria, a recent study has revealed that repeated herbicide applications result in an increase in the prevalence of antibiotic-resistance genes (Liao et al., 2021).

Intensification of agricultural practices represents one component of the many global changes that characterize the Anthropocene. Several other global changes such as global warming, widespread pollution, biodiversity losses and biological invasions are all likely to have impacted both the local and global diversity of plant-associated microorganisms. However, these impacts have largely gone unseen and have not been evaluated in the absence of historical records of microbial diversity. Exceptions to this assertion regard the historical spread of specific (crop) plant diseases (Yoshida et al., 2013) or the intentional introduction of symbionts or biological control agents (Jack et al., 2021). Consequently, most studies that monitor the impact of environmental changes on plant/soil microbial diversity are short-term synchronic ones based on the comparison of present-day situations, rather than diachronic studies based on the study of medium/long-term time series of environmental (plant/soil) samples.

Objectives of the thesis

The long-term ("distal") scientific objective of the research project underlying this PhD thesis is to evaluate the putative changes in communities of (crop) plant-associated microbes that took place in the last 100-150 years, a period that corresponds to the emergence and establishment of the Anthropocene. One recent characteristics of the Anthropocene is farming intensification characterized by mechanization, introduction of high yield crop varieties and the "massive" use of mineral fertilizers and of pesticides.

This research project, called "Holoherbarium", is based on the study of "past" plant and soil microbiomes associated to herbarium plant specimens. Plants in these collections (including cultivated ones) have been collected in the period of interest (~100-150 years – today) and indeed contain the molecular (DNA) signature of their associated microbial communities (either bacteria, leaf fungal endophytes or root fungal symbionts) as demonstrated by the preliminary studies of Bieker et al. (2020); Daru et al. (2021); Heberling & Burke (2019) or Weiß et al. (2020). We would like to propose rhizospheric soil, occasionally (and involuntarily) collected and preserved around the roots of herbarium plants (especially herbaceous taxa such as cereals), as historical soil samples giving access to historical soil microbiomes.

Thus, the molecular study of herbarium plants (and soils) collected "before", and "during" the exponential "acceleration" of global changes should inform us on how and to what extent these changes have durably affected the plant microbiome, widely recognized as an essential component of plant health.

Considering the almost absence of published reports in this field, the short-term ("proximal") objectives of the PhD project were, more specifically, to (1) validate herbaria as repositories of ancient cultivated soils that could represent a source of ancient soil microbiomes, (2) establish and validate laboratory protocols and analytical "pipelines" to study the diversity of herbarium-associated microbial communities and (3) demonstrate that these communities do reflect the original "living" communities associated to the plants at the time of their sampling.

Identification of these communities was performed based on the analysis of taxonomically informative historical DNA of microbial origin directly extracted from herbarium soil plant root samples. As described above, while a similar strategy has been already followed to trace individual microbial species (pathogens) associated to herbarium plants, to one exception that regarded leaves (Bieker et al. 2020), it has not been developed for entire root/soil communities. This strategy has however already been validated in the case of ancient microbial communities from sediments (Pearman et al., 2022; Siano et al., 2021) or animal/human remains such as paleofeces or dental calculus (Fellow Yates et al, 2021; Ottoni et al., 2021; Maixner et al. 2021, Wibowo et al.,2021).

The research work is presented in different chapters of the manuscript. **Chapter 1** focuses on paleomicrobiology, an emerging discipline that allows the study of past microorganisms and has been utilized throughout my thesis work. **Chapter 2** addresses the topic of herbaria as an invaluable resource for scientific research. **Chapter 3** addresses the methodological challenges, that regard the creation of a specific protocol for the extraction of DNA from root and soil samples originating from herbarium specimens and the bioinformatic processing of the resulting DNA sequences. **Chapter 4** explores the possibility of extending this analysis to different genera of plants and describes how the analysis of herbarium soils can be generalized to many different types of soils collected over a period encompassing most of the 19th and 20th centuries. **Chapter 5** explores root and soil microbial communities associated to four crop species specimens collected in the early 20th century. It demonstrates that herbarium root and soil microbiomes are similar to and obey to similar assembly rules as extant ones. Finally, **Chapter 6** outlines future perspectives, particularly the initiation of a new project on the study of agricultural soil microbial community's evolution over the past 200 years through the comparison of herbarium and extant soil samples collected all over mainland France.

References

Agoussar, A., & Yergeau, E. (2021). Engineering the plant microbiota in the context of the theory of ecological communities. Current Opinion in Biotechnology, 70, 220–225. https://doi.org/10.1016/j.copbio.2021.06.009

Banerjee, S., Walder, F., Büchi, L., Meyer, M., Held, A. Y., Gattinger, A., Keller, T., Charles, R., & van der Heijden, M. G. A. (2019). Agricultural intensification reduces microbial network complexity and the abundance of keystone taxa in roots. The ISME Journal, 13(7), 1722–1736. https://doi.org/10.1038/s41396-019-0383-2

Bieker, V. C., Barreiro, F. S., Rasmussen, J. A., Brunier, M., Wales, N., & Martin, M. D. (2020). Metagenomic analysis of historical herbarium specimens reveals a postmortem microbial community. Molecular Ecology Resources, 20(5), 1206–1219. https://doi.org/10.1111/1755-0998.13174

Busby, P. E., Soman, C., Wagner, M. R., Friesen, M. L., Kremer, J., Bennett, A., Morsy, M., Eisen, J. A., Leach, J. E., & Dangl, J. L. (2017). Research priorities for harnessing plant microbiomes in sustainable agriculture. PLOS Biology, 15(3), e2001793. https://doi.org/10.1371/journal.pbio.2001793

Chialva, M., & Bonfante, P. (2018). Plant microbiota: From model plants to Mediterranean crops. In *Phytopathologia Mediterranea* (Vol. 57, Issue 1, pp. 123–145). Firenze University Press. https://doi.org/10.14601/Phytopathol_Mediterr-22052

Daru, B. H., Davies, T. J., Willis, C. G., Meineke, E. K., Ronk, A., Zobel, M., Pärtel, M., Antonelli, A., & Davis, C. C. (2021). Widespread homogenization of plant communities in the Anthropocene. Nature Communications 2021 12:1, 12(1), 1–10. https://doi.org/10.1038/s41467-021-27186-8

Fellows Yates, J. A., Velsko, I. M., Aron, F., Posth, C., Hofman, C. A., Austin, R. M., Parker, C. E., Mann, A. E., Nagele, K., Arthur, K. W., Arthur, J. W., Bauer, C. C., Crevecoeur, I., Cupillard, C., Curtis, M. C., Dalen, L., Bonilla, M. D. Z., Fernandez-Lomana, J. C. D., Drucker, D. G., … Warinner, C. (2021). The evolution and changing ecology of the African hominid oral microbiome. Proceedings of the National Academy of Sciences of the United States of America, 118(20), e2021655118. https://doi.org/https://www.pnas.org/doi/abs/10.1073/pnas.2021655118

Heberling, J. M., & Burke, D. J. (2019). Utilizing herbarium specimens to quantify historical mycorrhizal communities. Applications in Plant Sciences, 7(4). https://doi.org/10.1002/APS3.1223

Jack, C. N., Petipas, R. H., Cheeke, T. E., Rowland, J. L., & Friesen, M. L. (2021). Microbial Inoculants: Silver Bullet or Microbial Jurassic Park? Trends in Microbiology, 29(4), 299–308. https://doi.org/10.1016/j.tim.2020.11.006

Lekberg, Y., Arnillas, C. A., Borer, E. T., Bullington, L. S., Fierer, N., Kennedy, P. G., Leff, J. W., Luis, A. D., Seabloom, E. W., & Henning, J. A. (2021). Nitrogen and phosphorus fertilization consistently favor pathogenic over mutualistic fungi in grassland soils. Nature Communications, 12(1), 3484. https://doi.org/10.1038/s41467-021-23605-y

Liao, H., Li, X., Yang, Q., Bai, Y., Cui, P., Wen, C., Liu, C., Chen, Z., Tang, J., Che, J., Yu, Z., Geisen, S., Zhou, S., Friman, V.-P., & Zhu, Y.-G. (2021). Herbicide Selection Promotes Antibiotic Resistance in Soil Microbiomes. Molecular Biology and Evolution, 38(6), 2337–2350. https://doi.org/10.1093/molbev/msab029

Lu, T., Ke, M., Lavoie, M., Jin, Y., Fan, X., Zhang, Z., Fu, Z., Sun, L., Gillings, M., Peñuelas, J., Qian, H., & Zhu, Y.-G. (2018). Rhizosphere microorganisms can influence the timing of plant flowering. Microbiome, 6(1), 231. https://doi.org/10.1186/s40168-018-0615-0

Maixner, F., Sarhan, M. S., Huang, K. D., Tett, A., Schoenafinger, A., Zingale, S., Blanco-Míguez, A., Manghi, P., Cemper-Kiesslich, J., Rosendahl, W., Kusebauch, U., Morrone, S.

R., Hoopmann, M. R., Rota-Stabelli, O., Rattei, T., Moritz, R. L., Oeggl, K., Segata, N., Zink, A., … Kowarik, K. (2021). Hallstatt miners consumed blue cheese and beer during the Iron Age and retained a non-Westernized gut microbiome until the Baroque period. Current Biology, 31(23), 5149-5162.e6. https://doi.org/10.1016/j.cub.2021.09.031

Ottoni, C., Boric, D., Cheronet, O., Sparacello, V., Dori, I., Coppa, A., Antonovic, D., Vujevic, D., Price, T. D., Pinhasi, R., & Cristiani, E. (2021). Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus. Proceedings of the National Academy of Sciences of the United States of America, 118(32), e2102116118. https://doi.org/https://www.pnas.org/doi/abs/10.1073/pnas.2102116118

Pearman, J. K., Biessy, L., Howarth, J. D., Vandergoes, M. J., Rees, A., & Wood, S. A. (2022). Deciphering the molecular signal from past and alive bacterial communities in aquatic sedimentary archives. *Molecular Ecology Resources*, *22*(3), 877–890. https://doi.org/10.1111/1755-0998.13515

Selosse, M.-A., Bessis, A., & Pozo, M. J. (2014). Microbial priming of plant and animal immunity: Symbionts as developmental signals. Trends in Microbiology, 22(11), 607–613. https://doi.org/10.1016/j.tim.2014.07.003

Siano, R., Lassudrie, M., Cuzin, P., Briant, N., Loizeau, V., Schmidt, S., Ehrhold, A., Mertens, K. N., Lambert, C., Quintric, L., Noël, C., Latimier, M., Quéré, J., Durand, P., & Penaud, A. (2021). Sediment archives reveal irreversible shifts in plankton communities after World War II and agricultural pollution. *Current Biology*, *31*(12), 2682-2689.e7. https://doi.org/10.1016/J.CUB.2021.03.079

Wagner, M. R., Tang, C., Salvato, F., Clouse, K. M., Bartlett, A., Vintila, S., Phillips, L., Sermons, S., Hoffmann, M., Balint-Kurti, P. J., & Kleiner, M. (2021). Microbe-dependent heterosis in maize. Proceedings of the National Academy of Sciences, 118(30), e2021965118. https://doi.org/10.1073/pnas.2021965118

Weiß CL, Schuenemann VJ, Devos J, Shirsekar G, Reiter E, Gould BA, Stinchcombe JR, Krause J, Burbano HA. 2016 Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens.R. Soc. open sci.3: 160239. http://dx.doi.org/10.1098/rsos.160239

Wibowo, M. C., Yang, Z., Borry, M., Hübner, A., Huang, K. D., Tierney, B. T., Zimmerman, S., Barajas-Olmos, F., Contreras-Cubas, C., García-Ortiz, H., Martínez-Hernández, A., Luber, J. M., Kirstahler, P., Blohm, T., Smiley, F. E., Arnold, R., Ballal, S. A., Pamp, S. J., Russ, J., … Kostic, A. D. (2021). Reconstruction of ancient microbial genomes from the human gut. Nature, 594(7862), 234–239. https://doi.org/10.1038/s41586-021-03532-0

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., Thines, M., Weigel, D., & Burbano, H. A. (2013). The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. *ELife*, *2*(2), 731. https://doi.org/10.7554/ELIFE.00731

Chapter1. Paleomicrobiology: Tracking the past microbial life from single species to entire microbial communities

This introductory chapter consists of a review paper published in Microbial Biotechnology entitled *'Tracking the past microbial life from single species to entire microbial communities'*. A shorter, popular science version of this article appeared in the online "The Microbiologist" journal: G. Grasso, V. Bianciotto and R. Marmeisse (2024) '*Shining a light on microbes from the past with molecular paleomicrobiology*' (https://www.the-microbiologist.com/features/shining-a-light-on-microbes-from-the-past-with-molecular-paleomicrobiology/2115.article).

This chapter describes and discusses an emerging discipline, exploited in my research work, known as molecular paleomicrobiology. This discipline makes use of degraded ancient DNA (or RNA) to highlight the occurrence of microorganisms and their genetic makeup directly in ancient samples. The field is expanding rapidly, propelled by technologies like very-high-throughput sequencing, enabling researchers to explore periods ranging from decades to hundreds of thousands of years before present.

Molecular paleomicrobiology is at the heart of my PhD thesis work that aims at highlighting the diversity and evolution of ancient plant-associated microbial communities associated to plant specimens stored in herbarium collections. To better approach this objective, it is essential to be familiar with the latest developments in this field of research, of the latest technical developments and scientific breakthrough.

Thus far, molecular paleomicrobiology has essentially concentrated on the evolution and diffusion of pathogens. Among these, *Yersinia pestis*, the causal agent of plague, has been the focus of the highest number of published studies. However, several other pathogens have been targeted, not only bacteria (e.g., *Mycobacterium* sp.), but also viruses (e.g., Hepatitis B virus), or eukaryotes (e.g., the Oomycete *Phytophthora infestans*), infecting either animal/human or plant species.

Paleomicrobiology extends however beyond single individual microbial species to encompass entire microbial communities associated with macro-organisms such as humans, animals, or plants. These host-associated microbial communities are influenced by various factors including the host's identity (species, genotype), lifestyle (diet, health status), and local environment. Several studies in molecular paleomicrobiology have tried to understand how factors like

the host's characteristics, where they lived, their environment, and their interactions with other organisms affected the evolution of their microbiomes over time. This is an important issue since microbiomes are known to affect the health and survival of their hosts. Finally, paleomicrobiology can also investigate free-living microbial communities trapped in environmental matrices as diverse as freshwater or marine sediments, ancient soils, permafrost or ice cores. These microbial communities make it possible to reconstruct the history of ecosystems over geological timescales and reveal shifts and changes in microbial communities that have occurred in response to natural or anthropogenic events.

Received: 28 July 2023 | Accepted: 10 December 2023

DOI: 10.1111/1751-7915.14390

MINI REVIEW

MICROBIAL BIOTECHNOLOGY Applied Microbiology International

Paleomicrobiology: Tracking the past microbial life from single species to entire microbial communities

Gianluca Grasso[1,2,3] | **Valeria Bianciotto**[3] | **Roland Marmeisse**[2,3]

[1]Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università degli Studi of Turin, Turin, Italy

[2]Institut Systématique Evolution, Biodiversité (ISYEB: UMR7205 CNRS-MNHN-Sorbonne Université-EPHE-UA), Muséum National d'Histoire Naturelle, Paris, France

[3]Institute for Sustainable Plant Protection (IPSP), SS, National Research Council (CNR), Turin, Italy

Correspondence
Gianluca Grasso and Roland Marmeisse, Institut Systématique Evolution, Biodiversité (ISYEB: UMR7205 CNRS-MNHN-Sorbonne Université-EPHE-UA), Muséum National d'Histoire Naturelle, 12 rue Buffon, Paris F-75005, France.
Email: gianluca.grasso@unito.it and roland.marmeisse@mnhn.fr

Funding information
Alliance Sorbonne Université; Muséum National d'Histoire Naturelle, Grant/Award Number: ATM 2021; Università Italo Francese, Grant/Award Number: Bando Da Vinci 2022

Abstract
By deciphering information encoded in degraded ancient DNA extracted from up to million-years-old samples, molecular paleomicrobiology enables to objectively retrace the temporal evolution of microbial species and communities. Assembly of full-length genomes of ancient pathogen lineages allows not only to follow historical epidemics in space and time but also to identify the acquisition of genetic features that represent landmarks in the evolution of the host–microbe interaction. Analysis of microbial community DNA extracted from essentially human paleo-artefacts (paleofeces, dental calculi) evaluates the relative contribution of diet, lifestyle and geography on the taxonomic and functional diversity of these guilds in which have been identified species that may have gone extinct in today's human microbiome. As for non-host-associated environmental samples, such as stratified sediment cores, analysis of their DNA illustrates how and at which pace microbial communities are affected by local or widespread environmental disturbance. Description of pre-disturbance microbial diversity patterns can aid in evaluating the relevance and effectiveness of remediation policies. We finally discuss how recent achievements in paleomicrobiology could contribute to microbial biotechnology in the fields of medical microbiology and food science to trace the domestication of microorganisms used in food processing or to illustrate the historic evolution of food processing microbial consortia.

Introduction

Archaea, bacteria and unicellular eukaryotes are certainly the oldest forms of life on Earth. They originated about 3.5 Gyr ago and participated in the formation of the original biosphere. They first inhabited the earth's anoxic environment and then enabled its oxygenation (Falkowski et al., 2008; Fischer et al., 2016).

Although they played essential roles in ancient and modern environments, temporal patterns of evolution and diversification of microorganisms are poorly known because of the scarcity of fossil records whose taxonomic assignments are often problematic. With few notable exceptions *(e.g.* foraminifera, diatoms), most microbial cells do not produce mineralized structures and their often indistinctive shapes and structures preclude precise identification of fossilized microbial taxa based on morphological examination (Xie & Kershaw, 2012). Furthermore, even for taxa well represented in the fossil records, such as the Foraminifera, the existence of clades comprising naked unfossilized species may prevent a correct interpretation of their patterns of evolution over time when exclusively based on fossil data (Pawlowski et al., 2003). Thus, as opposed to ancient fauna and flora (McElwain & Punyasena, 2007; Raup, 1982; Signor, 1994), the rarity of taxonomically assignable paleontological specimens illuminates only a small sliver of the real past microbial diversity and makes it difficult to investigate microbial evolution, diversification and functional significance across the different geological eras.

For these reasons, evolution and diversification of microorganisms have been largely inferred from molecular phylogenetic reconstructions that make use of DNA/protein sequences of extant species (Nee, Holmes, et al., 1994; Nee, May, et al., 1994; Michael J., 1996; Morlon, 2014; Stilianos Louca et al., 2018). However, these phylogenies are difficult to time calibrate, precisely because of the absence, gaps or imprecision of microbial fossil records that could be used as landmarks. Molecular phylogenies also hardly predict the potential existence and functions (ecological roles) of extinct clades that are thought to exceed extant ones in number (Stilianos Louca et al., 2018; Tricou et al., 2022). To date, a global microbial life pattern of evolution remains largely unresolved and only few studies focus on their past diversification patterns (Morlon et al., 2012; Lorén et al., 2014; Gubry-Rangin et al., 2015; Lebreton et al., 2017; Marin et al., 2017; Stilianos Louca et al., 2018).

Furthermore, until a very recent past, most of diversity and global phylogenetic studies also suffered from the lack of a global view regarding the magnitude of

extant microbial diversity not only in terms of absolute number of taxa whose estimates vary by several orders of magnitude in the case of Bacteria and Archaea (Loucaid et al., 2019; Lennon & Locey, 2020), but also in terms of phylogenetic diversity. Regarding this last aspect, the gap is rapidly closing thanks to the multiplication of metagenomics studies that give access to the genome sequences (so-called Metagenome assembled genomes, MAGs) of species belonging to, thus far, overlooked microbial clades (Hug et al., 2016; Nayfach et al., 2021).

While the previous paragraphs exposed the obstacles to which one is confronted when addressing microbial evolution, past diversity and contribution to ecosystem processes in a "distant past", these obstacles partially vanish when referring to a "recent past". This is made possible thanks to the emergence of paleogenetics/paleogenomics that analyses microbial DNA preserved in ancient environmental samples as diverse as sediment cores, archaeological artefacts, long buried animal/human skeletons, or items preserved in natural history collections (**Fig. 1**). Indeed, ancient DNA (aDNA, but also RNA or proteins) that survives through time to the death of any organism can be regarded as a fossil trace of the corresponding organism and can be interrogated to retrace the "recent" evolution of the corresponding species or group of species (Arning & Wilson, 2020; Kistler et al., 2020; Orlando, 2020; Orlando et al., 2021; Raxworthy & Smith, 2021; Siano et al., 2021). Besides specific taxa, aDNA studies also make the exploration of entire communities of (micro)organisms from the past possible, providing a comprehensive vision of their diversity and functional roles in their original ecosystems.

The aim of this review is to highlight, through the description of selected examples, the different facets and main achievements of molecular paleomicrobiology. In this article, which does not claim to be exhaustive, technical and theoretical issues of paleogenetics/paleogenomics that have been reviewed many times and that are not at the heart of this article, will only be briefly mentioned (Warinner et al., 2017; Brunson & Reich, 2019). We separately illustrate and discuss studies that focus (i) on individual microbial taxa and those that probe (ii) entire microbial communities whatever the environment they originate from. Finally, we will conclude by suggesting what could be the potential contribution of molecular paleomicrobiology to the field of microbial biotechnology.

Molecular paleomicrobiology as a discipline

Time frame

Molecular paleomicrobiology, based on the analysis of degraded ancient DNA (or eventually RNA) is a discipline that presents several specificities. It is based on a diachronic approach that documents the occurrence of microorganisms and their original genetic make-up directly in ancient samples as diverse as environmental *(e.g.* soils or sediments) or archaeological ones to compare the ancient samples to those of today. It is thus different from phylogenetic-based approaches that reconstruct past evolutionary events *in silico*, using information obtained on extant organisms. Although the prefix "paleo" commonly refers to a "distant" past, it seems preferable not to impose a minimum temporal threshold to molecular paleomicrobiology that studies organisms with often very short generation times that can acquire adaptive mutations in far shorter time spans compared to most "macroorganisms". Thus, a study published in 2016 (Worobey et al., 2016) that investigated the initial events of the AIDS epidemics in North America in the 1970s based on the extraction and sequencing of degraded RNA molecules from circa 40-year-old archived blood samples to reconstruct HIV genomes clearly belongs to the research field from both a technical and scientific point of view.

Ancient DNA (aDNA)

Indeed, a second specificity of the discipline, detailed in several technical reviews (Pedersen et al., 2015; Rivera-Perez et al., 2016; Warinner et al., 2017; Afouda et al., 2020; Orlando et al., 2021), is to deal with aDNA, a material often difficult to access, limiting in quantity and degraded. Its extraction and manipulation require the implementation, by specifically-trained scientists, of particular protocols in dedicated laboratories (clean rooms, Cooper & Poinar, 2000; Orlando et al., 2021). As in 2023, the oldest DNA sample ever analyzed was extracted from two-million-year-old frozen sediments in Greenland and allowed reconstructing the entire ecosystem that shaped this region at that time (Kjær et al., 2022).

Ancient DNA is more complex to analyze compared to modern one due to the presence of postmortem damages (PMDs) resulting in DNA fragmentation and base modifications (Dabney et al., 2013, **Fig. 2**). These PMDs are essentially the result of depurination and deamination. Depurination, *i.e.* the loss of adenine and guanine bases, is the consequence of cleavage of β-N-glycosidic bonds and is at

the origin of DNA fragmentation producing very short fragments of mostly less than 100 bp in length. The rate of DNA fragmentation is however environment-dependent. While aDNA extracted from one hundred-year-old plant herbarium specimens is mostly made of fragments in the range 40-150 bp (Staats et al., 2011; Yoshida et al., 2013), aDNA extracted from more than 10,000 years-old lake sediments can still be mostly made of fragments larger than 100 bp (Talas et al., 2021). Consequently, studies carried out on sedimentary DNA can implement metabarcoding approaches based on the amplification of barcode DNA sequences sometime larger than 200 bp. This approach would be inoperative for highly degraded DNA samples.

As for "spontaneous" cytosine deamination, it leads to the conversion of cytosine to uracil and results in the incorporation of adenine on the complementary strand instead of guanine during in vitro DNA synthesis. Cytosine deamination occurs at a higher frequency in single-stranded ends of degraded DNA molecules. While age of the specimen represents one of the main factors that controls the magnitude of PMDs, other factors like local temperature, depositional conditions, post-

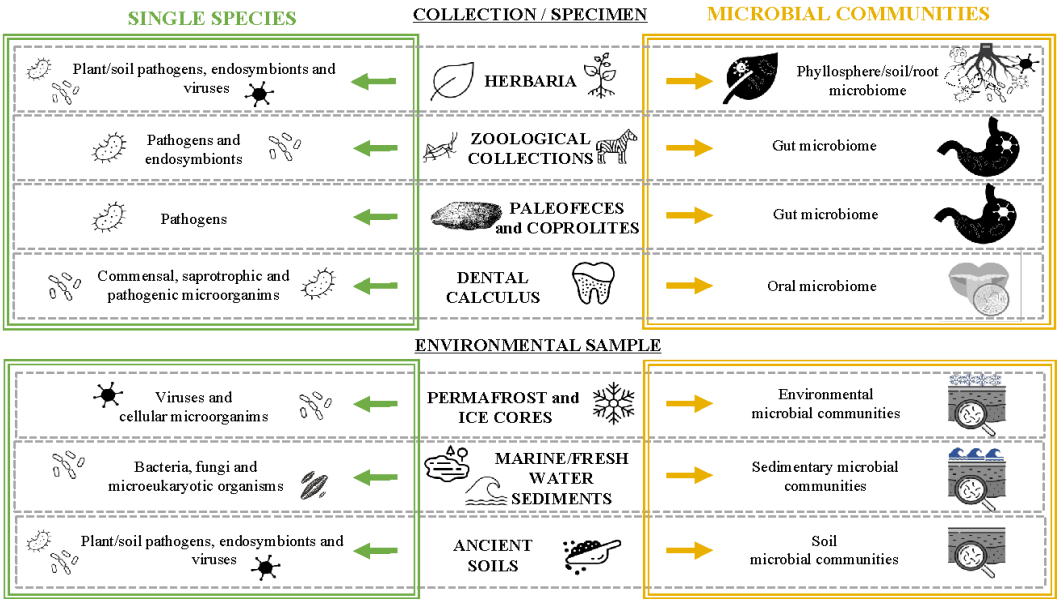


Figure 1. Main types of historical samples used as sources of ancient microbial DNA. These samples can be extremely ancient as Bronze Age teeth or dental calculi or much more recent such as often less than 100-year-old herbarium plants. These samples have been used to study either specific (often pathogenic) microbial species or entire communities.

excavation handling, specimen treatment and source tissue also influence aDNA conservation (Arning & Wilson, 2020).

Thus, besides the "wet lab" manipulation of degraded DNA, paleogenetics, including molecular paleomicrobiology, relies on a suite of adapted bioinformatics tools to extract relevant information from very short DNA sequences and to quantify chemical alterations that represent signatures of the ancient origin of the studied nucleic acids. Detection of these chemical modifications is performed by specific software like PMDTools (Skoglund et al., 2014) or mapDamage (Jónsson et al., 2013) which quantify C-T and G-A transitions at the ends of the aDNA fragments.

A dialog between disciplines

Finally, a third specificity of paleomicrobiology is to frequently require close collaborations between microbiologists and specialists of other disciplines sometimes from humanities, such as historians, archaeologists, paleoanthropologists, paleopathologists, paleoclimatologists, geologists, or curators of collections of natural history or of anthropology. These collaborations are necessary to localize, identify, date and access relevant historical samples with the highest probability to contain usable DNA traces of the studied microbial species or microbial communities and also to "contextualize" these samples in their original, historical environment (Kuyl, 2022).

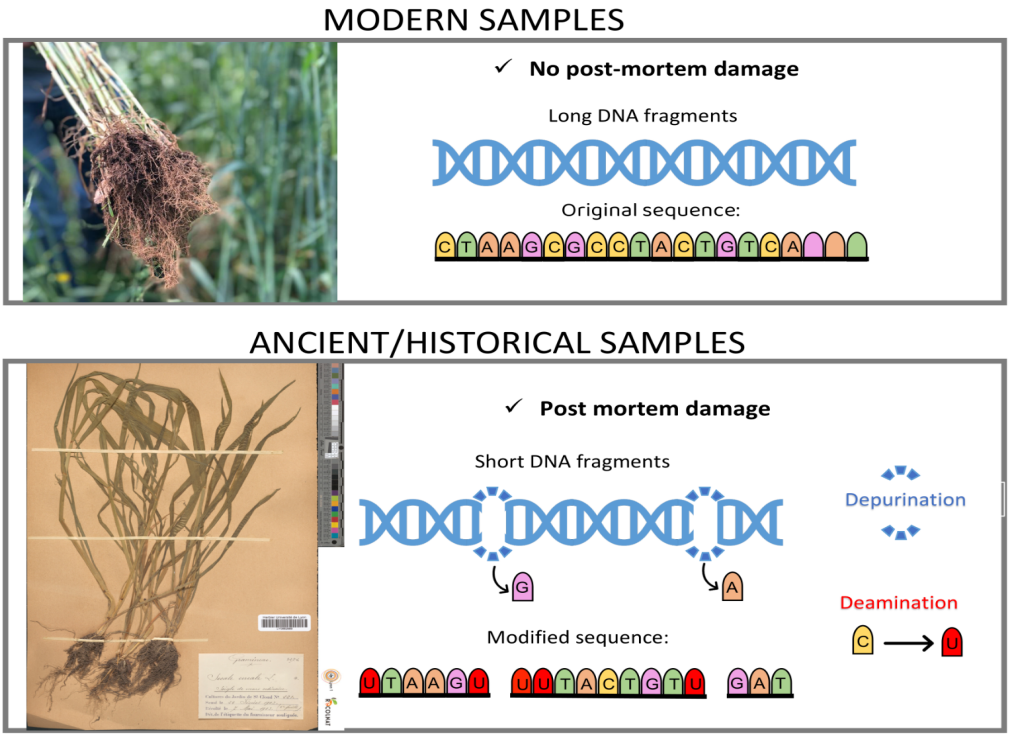


Figure 2. Main characteristics of ancient DNA. While DNA extracted from living organisms is made of long molecules, often exceeding 1 kbp in length and perfectly representativeof the original organism's genomic sequence, ancient DNA extracted from historical samples has experienced post-mortem damages. They lead to DNA fragmentation (fragments often shorter than 100bp) and base modifications altering the original sequence. Thus, cytosine deamination produces uracil, preferentially at the ends of the molecules. Photo credits: modern plant samples: ©Herbier LY, FR-CERESE, UCBLyon1 Recolnat portal (https://explore.recolnat.org/) under accessions LY0662689.

Single species approach

Numerous molecular paleomicrobiology studies focus on a specific microbial taxon. Most of them are pathogens, primarily human ones, bacteria or viruses. The field of investigation is likely to widen rapidly, with the emergence of studies in the field of plant pathology as well as the availability of genomes of non-pathogenic, often uncultivable species reconstructed *in silico* (so-called MAGs, for metagenome-assembled genomes) from massive sequencing data of "microbial paleo-communities" (Marx, 2016; Bos et al., 2019; Granehäll et al., 2021; Wibowo et al., 2021). This section of the manuscript takes as a main example the studies on the human plague that illustrate *a posteriori* the specificity of the discipline presented in the introductory section.

The history of the plague bacterium

Before paleogenomic studies of plague, there were a number of certainties about the nature of this disease, but also a number of uncertainties that could be not addressed using currently circulating bacterial strains. First of all, it is caused by *Yersinia pestis*, which was first isolated in China in 1894 during the last great plague epidemic (the so-called 3rd pandemic). The main reservoir of this disease is wild rodents, and it is essentially transmitted from animals to humans through the bites of infected fleas. This form of transmission leads to the bubonic form of the disease, which takes its name from the swelling of lymph nodes. The disease, in rarer cases, can be transmitted directly from humans to humans through the respiratory tract where it leads to a pulmonary infection that is also usually fatal. In the absence of appropriate treatment, the infection leads to very high mortality rates during epidemic episodes. *Y. pestis*, whose genome has been sequenced many times (more than 600 genomes available in GenBank in 2023), has been the subject of numerous experimental studies which have identified many genes involved in the virulence and aggressiveness of this bacterium in humans as well as in the insect vector (Demeure et al., 2019; Hinnebusch et al., 2021). Some of these genes are carried by plasmids.

None of the studies conducted on modern strains of *Y. pestis* could however confirm that the so-called "plague" epidemics, prior to the late 19th century one, were caused by the same pathogen despite the similarity of symptoms reported in historical texts and illustrations. The pandemic that broke out in China in the second half of the 19th century is referred to as the 3rd pandemic. It followed a first one called the Justinian pandemic (6th-8th century AD) and a second one

(14th-18th century) that peaked in Europe as the Black Death in 1347-1351, decimating up to 60% of the population locally. After a phase of decline, each of these pandemics gave rise to more localized epidemic episodes.

In this context, the paleomicrobiological approach is *a priori* the only one that could answer a set of questions, common to many other infectious diseases. (i) Are the historical cases of the disease (here the 1st and 2nd pandemics) attributable to the same infectious agent? (ii) Did several distinct strains/evolutionary lineages of the pathogen circulate during a single pandemic? (iii) Are successive pandemics/epidemics due to the re-emergence of the evolutionary lineage that predominated during the previous pandemic? (iv) Can infections be documented in humans at earlier dates for which documentary sources do not exist? And (v), what can the analysis of ancient strains teach us about the temporal evolution of the pathogen's virulence?

If certain diseases lead to bone alterations such as tuberculosis (bone tuberculosis) or leprosy, thus allowing human remains to be targeted for paleomicrobiological studies, this is not the case for plague. Nevertheless, the high mortality induced by the disease is known to lead to a modification of funerary practices characterized by the burial of bodies not in individual, but in collective graves. aDNA of *Y. pestis* has thus been successfully extracted essentially from the dental pulp of teeth, a densely vascularized tissue, taken from skeletons found in collective graves of the different pandemics (Harbeck et al., 2013; Bos et al., 2016; Spyrou et al., 2016; Susat et al., 2020). Several research groups have independently been able to reconstitute entire genomes of *Y. pestis* either directly from systematic sequencing of the extracted aDNA, or after enrichment in *Y. pestis* DNA by sequence capture (Bos et al., 2016; Spyrou et al., 2016). This first set of results validated the hypothesis that the first two plague pandemics (6th-8th century and 14th-18th century) were indeed due to *Y. pestis* found in numerous sites covering a vast territory in Eurasia and several decades (Harbeck et al., 2013; Bos et al., 2016; Susat et al., 2020; Bramanti et al., 2021; Spyrou et al., 2022).

Integration of ancient and modern genomic sequences into a global phylogeny of the *Y. pestis* species showed that strains from the same pandemic tend to cluster together to form distinct lineages, suggesting that major pandemics do not result from the re-emergence of strains that produced the previous ones (Bramanti et al., 2021; Hinnebusch et al., 2021). The earliest evidence of the evolutionary lineage of the second pandemic dates from 1338-1339 and was found in

Kyrgyzstan, suggesting that it originated in Central Asia (Spyrou et al., 2022). It was always members of this lineage that caused epidemic rebounds more than 3-4 centuries later in Europe (Bos et al., 2016; Bramanti et al., 2021; Spyrou et al., 2022).

To address the diffusion of *Y. pestis* before the first documented pandemics, a systematic screening for the presence of the bacterium sequences in DNA extracted from human remains buried individually or collectively was necessary. This more tedious approach led to the reconstruction of 17 ancient *Y. pestis* genomes among a set of 252 human skeletons dating from 5000 to 2500 y BP and over a territory ranging from the Iberian Peninsula in the West to Mongolia in the East (Valtueña et al., 2022). The ability to infect humans thus predates the recorded three pandemics (Valtueña et al., 2017, 2022; Spyrou et al., 2018; Rascovan et al., 2019). The fact that many infected bodies were buried individually suggested that these cases of historical infections may not have led to mass mortality.

We thus now have a substantial number of genomes of historical *Y. pestis* strains covering a period of time of more than 5000 years. As we have just summarized, the comparison of these genomes allows the construction of time-calibrated molecular phylogenies and traces "objectively" the course of epidemics and movements of these pathogens on different time and spatial scales (Spyrou et al., 2018; Rascovan et al., 2019; Valtueña et al., 2022). Besides phylogenetic reconstructions, confrontation of the gene content of ancient genomes with our current knowledge of the molecular basis of the modes of infection and virulence of the pathogen offers a unique opportunity to address the evolution of the modes of transmission and of the virulence of the pathogen over time.

In the case of *Y. pestis*, several of the key genes necessary for this bacterium to infect fleas and make them effective in transmitting the pathogen to humans and causing the bubonic form of the disease are known (Hinnebusch et al., 2021). The ability to infect these insects requires for instance the presence of the *ymt* gene and a mutation in the *ure2* gene that inhibits the production of a functional urease. The key *ymt* gene was absent in most strains found in the Late Neolithic-Early Bronze age (5000-3500 BP) which also seem to be able to produce a functional urease (Valtueña et al., 2017; Spyrou et al., 2018; Vågene et al., 2022). During this period the plague could have thus been predominantly transmitted from human to human without an intermediary insect host. However, the mode of transmission via flea bites is by far the most efficient and determinant for a

massive diffusion of the pathogen. While early studies on Bronze Age strains suggested that acquisition of the *ymt* gene occurred later, two recent studies independently found two *Y. pestis* strains carrying this gene as well as mutations in the *ure2* gene and other genes promoting flea infection. These two observations were made on skeletons excavated, one in Spain (3200 years BP) (Valtueña et al., 2022) and the other in Russia (3800 years BP) (Spyrou et al., 2018). These strains potentially capable of being transmitted by fleas may therefore have spread unnoticed over a large geographical area at a time when other forms of pathogen transmission were predominant.

A second piece of information that may shed light on the aggressiveness on mammals of historical plague strains concerns the *pla* gene, a major virulence factor carried on the *Y. pestis* pPCP1 plasmid. Independent studies on different isolates from different geographical origins reported of a specific "erosion" of this gene in several strains of the second pandemics, after the end of the Black Death episode (after the 15th century, Susat et al., 2020; Bramanti et al., 2021). This “erosion” corresponds to a lower sequence coverage of this gene compared to the rest of the pPCP1 plasmid sequence. Thus, only a small percentage of plasmids may have carried the *pla* gene. This could indicate that the strains at the end of the epidemic peak were affected in their virulence, although this remains to be experimentally demonstrated.

Beyond plague, other pathosystems

While plague has, thus far, certainly been the focus of the largest number of studies in the field of paleomicrobiology; similar scientific questioning has been elaborated on other pathogens. They belong to the bacteria *(e.g. Mycobacterium sp*.) but also to the viruses *(e.g.* the Hepatitis B virus) or the Eukarya *(e.g.* the Oomycete *Phytophthora infestans*) and infect either animal/humans or plant species. In different instances the results reported contrast with those obtained for plague.

In the case of bone tuberculosis, essentially resulting nowadays from infections by *Mycobacterium tuberculosis*, it has been demonstrated that the symptoms observed on human skeletons from 950 to 1550 CE along the coasts of South America, prior to the European invasions that may have brought *M. tuberculosis* to this region, were the result of infections by the species *M. pinnipedii* (Vågene et al., 2022). This species preferentially infects Pinnipeds, such as fur seals abundant in this region. The presence of this infection in human communities

along the coast can be explained in by recurrent transmissions from animals to humans resulting from the proximity between these two species, the former of which being hunted for its meat. However, the presence of such infections on skeletons found further inland raises questions about possible human-to-human transmission by one or more lines of *M. pinnipedii* that would have adapted to this new host and that may have gone extinct (Vågene et al., 2022).

Regarding the progression of "successive waves" of different genotypes of a pathogen in the same geographic area, it has been reported for other microbial species belonging to different taxonomic groups in different time periods. Thus, the systematic screening of more than 130 human remains from Eurasia and America covering a period from 10,500 BP to 400 BP has allowed to trace the evolutionary history of the hepatitis B virus whose infection does not leave any visible trace on skeletons and does not lead to mass mortality (Kocher et al., 2021). In Western Europe, at least 5 evolutionary lineages of the virus followed one another over this period. One of them in particular, called WENBA, prevailed for nearly 4000 years from about 7500 to 3500 years ago, i.e. a period straddling the Neolithic and the Bronze Age, before being apparently eliminated and replaced by a new genotype that is still found today in Europe. The dissemination of the WENBA lineage found on several Early European Farmers skeletons coincides with the Hunter-Gatherers/Farmers transition in Europe. It is interesting to note that a descendant of the WENBA lineage, not found for nearly 3500 years, has recently reappeared in today's human populations, often in association with HIV carriers. The source of this re-emergence of a lineage thought to be extinct remains to be identified (Kocher et al., 2021).

Another example of historical epidemiological monitoring concerns the Oomycete *Phytophthora infestans*, agent of the Potato blight. Its introduction from America to Europe in the first half of the 19th century led to the destruction of this crop and triggered the great famine in Ireland, which caused more than a million deaths and resulted in a wave of emigration from Europe to North America. Genome sequencing of this species was performed using (degraded) aDNA extracted from potato leaves, stored in herbaria, which displayed typical leaf lesions (Martin et al., 2013; Yoshida et al., 2013). These specimens were collected mainly in Western Europe but also in North America between 1845, the probable date of the first introduction of the pathogen into Europe, and the end of the 19th century. All isolates collected along the 19th century had genomes very similar to each other but distinct from the genomes of strains currently

circulating in Europe. Thus, it appears that the initial introduction into Europe was from a single strain or from closely related strains that circulated throughout the 19th century only to be replaced by new, genetically distinct isolates (Yoshida et al., 2013).

As in the case of *Y. pestis*, functional information was also deduced from the analysis of *P. infestans* genomes that circulated in Europe in the 19th century. They possessed a functional AVR3KI avirulence gene, whereas modern strains possess the AVR3EM allele of this gene (Martin et al., 2013; Yoshida et al., 2013). When a *P. infestans* strain carries a AVR3KI allele it cannot infect a potato cultivar carrying the cognate *R3a* resistance gene (Yoshida et al., 2013; Bieker et al., 2020). This *R3a* resistance gene was however absent in potato lines grown in Europe in the 19th century. Its introgression into modern potato cultivars to fight *P. infestans* led however to the emergence of strains of the pathogen carrying the new AVR3EM allele that bypassed the resistance conferred by the *R3a* gene.

Ancient microbial communities - specific issues

Unlike studies targeting a single species, the analyses of ancient microbial communities must take several factors into account that may have biased the relative abundance of the different species within the pool of ancient DNA extracted from the samples under study.

A first factor is the contamination of ancient material by an external source, which may itself be ancient and therefore characterized by the presence of DNA with post-mortem damages. For example, in the case of buried animal/human remains, the surrounding substrate (soil) may have contaminated these remains. These contaminations can be evaluated and removed *a posteriori* from the sequence datasets using bioinformatics approaches. Thus, to validate microbiome preservation in different ancient dental biofilm samples, Fellow Yates at al., 2021, developed a multistep procedure that included (i), metagenomic binning of the data to the NCBI nucleotide database for a taxonomic assignation and, (ii) subsequent validation of the identified microbial taxa by comparison with oral and non-oral reference microbiomes. Another popular approach is the use of the SourceTracker software (Kinghts et al. 2011), a tool based on Bayesian methods that compares the microbiome dataset under study with datasets of published microbiomes from the same and different environments. Similarly, the R decontam package allows for the removal of laboratory and environmental contaminants prior to subsequent analysis of microbiomes (Davis et al., 2018).

In certain environments, such as lake or marine sediments, active or dormant microorganisms may also naturally cohabit with ancient nucleic acids from microorganisms that have disappeared. In paleomicrobiology, it is not easy to distinguish between these different sources and this can lead to erroneous conclusions about the temporal evolution of microbial communities. In environmental microbiology, it is traditionally considered that DNA persists longer in the environment after the death of cells than RNA, which is a more labile molecule whose presence would indicate the presence of active living cells. Thus the detection within an environmental archive of RNA associated with a specific taxon should prompt its exclusion from downstream analyses. However, recently published results suggest that RNAs may themselves persist in the environment for several decades (Pearman et al., 2022). Therefore, as a precautionary measure, paleoenvironmental analyses, particularly on sediment cores, should primarily focus on microbial taxonomic groups whose short- to medium-term survival in sediments is unlikely. This is the case, for example, for

many strictly photoautotrophic organisms such as cyanobacteria or unicellular eukaryotic algae.

Contamination of ancient material may also be endogenous, due to the secondary development of microorganisms during the conservation or fossilization process. This was suggested in the study of herbarium samples of *Ambrosia* and *Arabidopsis* plants (Bieker et al., 2020). Most of the plants preserved in herbaria contained sequences attributed to the ascomycete fungus *Alternaria alternata* that were never found in modern samples of plants of these species. This observation led the authors to suggest that this fungus colonized the plants in the herbaria after they had been collected.

Another factor that is difficult to assess is the differential conservation (of DNA) of the different species in the microbial communities, leading to artifactual changes in their relative abundance. These changes could occur at different stages in the conservation process. In aquatic ecosystems, in the case of lacustrine cyanobacterial communities, differential sedimentation of planktonic cells has been reported. Nwosu et al., 2021 observed an over-representation in sediment traps, placed at the bottom of a lake, of species producing aggregates or colonies of cells compared with other species, which, although abundant in the water column, are characterized by small individual cells producing gas vesicles. These latter species are not only less likely to sediment rapidly but are also more vulnerable to predation and to the rapid degradation of their DNA (Nwosu et al., 2021). In addition to this direct evidence made on extant samples, observations made on ancient material also suggest differential conservation of the genetic material of certain taxa after their death. This was reported for a desiccated microbial mat dated around 1000 yr BP collected in Antarctica. Analysis of the DNA and proteins extracted from this mat revealed very different taxonomic profiles of the microbial communities (Lezcano et al., 2022). Among the DNA sequences, the authors observed a high prevalence of Clostridiales and Actinomycetales and a virtual absence of cyanobacteria. The taxonomic affiliation of the extracted proteins, on the other hand, reveals a high abundance of cyanobacteria and a quasi-absence of Clostridiales and Actinomycetales. This latter observation reflected better the presumed nature of the studied material. It was hypothesized that certain microbial taxa, particularly those producing spores such as the Clostridiales, are characterized by a slower postmortem degradation of their DNA and are therefore over-represented in the archives. It should however be noted that this study was conducted using a metabarcoding approach,

which only allows analysis of the fraction of extracted fragments whose size exceeds the size of the amplified marker. In that way, DNA less degraded is favored over the very short fragments that may be more representative of ancient communities.

To minimize the complex problem of artefactual modification of the relative abundance of taxa over time, it is possible to discuss the results only in terms of the presence or absence of taxa within the communities. However, this only makes sense in cases where these taxa play a very specific role, as is the case for pathogenic species. Thus, Bonczarowska et al., (2022) systematically searched for human pathogen sequences (bacteria and viruses) in DNA extracted from the teeth of 70 individuals buried in the same German village during the Merovingian period (fifth-eighth century CE). On twenty-two individuals (31%) were found at least one of the following 4 pathogens: Hepatitis B virus, Smallpox virus or Parvovirus B and the leprosy agent, *Mycobacterium leprae*. Seven cases of double infection and one case of triple infection suggested that several of the pathogens were endemic within the population and that overall this village community had a poor health status.

Despite these warnings, molecular paleomicrobiology remains the only approach for probing the diversity of past microbial communities, whose members, sometimes all of them, have left no fossil record. Nevertheless, this assertion is also true in the case of the very few microbial groups that are widely studied in micropaleontology, such as the Foraminifera. Thus, a metabarcoding study targeting this taxon in a marine sediment core covering more than 1000 years of sedimentary deposits identified nine times more molecular taxa than morphological ones producing hard shells preserved in sediment (Pawłowska et al., 2014). This 'excess' of molecular taxa can be explained by the existence of cryptic species producing morphologically similar hard shells, as well as by the existence of numerous taxa that do not produce these fossilized structures.

In this section dedicated to communities, we will distinguish studies exploring free-living microbial communities from studies of “host-associated” microbial communities. This distinction is justified by the fact that to date, studies of free-living communities have mainly involved the analysis of DNA extracted from sediment cores using metabarcoding approaches. As for studies on “host-associated” communities, besides exploring more diverse sample types (**Fig. 1**), they mainly follow a systematic high throughput DNA sequencing approach.

This approach allows *de novo* reconstruction of microbial genomes and detailed functional analysis of the role of microorganisms in ancient ecosystems.

Free-living ancient microbial communities

Free-living ancient microbiomes, or "non-host associated microbiomes", encompass microbial communities entrapped in environmental matrices as diverse as freshwater or marine sediments, ancient soils, permafrost, or ice cores (**Fig. 1**). The study of aDNA from these samples, often referred to as sedimentary DNA (sedaDNA), makes it possible to reconstruct the history of ecosystems over geologic times and to reveal shifts and changes in microbial communities that have occurred in response to natural or anthropogenic constraints. At present, most studies have been performed on sediments and few data are available on ancient soils (Clark et al., 2008) and permafrost (see below "Future perspectives").

Freshwater and marine sediments

Freshwater *(e.g.* lake), and marine sediments archive DNA not only from aquatic benthic and pelagic (micro)organisms, but also from terrestrial ones encapsulated in wind dispersed propagules *(e.g.* fungal spores) or that are transported horizontally by rivers to their estuaries where they sediment. Therefore, study of sedimentary DNA (sedaDNA) composition does not only illustrate past aquatic biodiversity but also the global diversity of surrounding terrestrial ecosystems, including their fauna and flora (Wang et al., 2021; Kjær et al., 2022). Thus, the analysis of sedaDNA extracted from a Latvian lake sediment core, allowed Talas et al. (2021) to retrace the evolution of both aquatic (23% of the molecular taxa) and terrestrial (40%) fungal communities over a period of more than 10,000 years encompassing the Holocene. Besides taxonomic assignation, functional assignation to different trophic modes *(e.g.* saprotrophs, animal of plant pathogens or symbionts) provided indirect information regarding the occurrence of plant genera (*Alnus*, *Salix*,...) known to be specifically associated to particular fungal pathogenic or mutualistic species.

Studies focusing on restricted geographic areas allowed monitoring on a fine temporal scale how human activities strongly, and maybe irreversibly, impacted coastal microbial, and more specifically microeukaryotic communities. In the bay of Brest, on the French Atlantic coast, after a long period of global stability since the Middle Ages, communities of eukaryotic microorganisms "suddenly" changed from the Second World War onwards and since then never turned back to their original composition (Siano et al., 2021). Dinoflagellates and Stamenopiles were the most affected groups with the almost complete

disappearance of taxa that durably dominated the communities since the middle age and their replacement by other taxa such as the potentially harmful, toxin-producing *Alexandrium* algae. Such changes parallel and can certainly be attributed to simultaneous abrupt changes in local human activities that also left identifiable traces in sediments in the form of accumulation of inorganic and organic pollutants or changes in sediments accumulation patterns attributable to agricultural practices. Similar dramatic changes in the composition of coastal microbial communities have also been reported by Barrenechea Angeles et al. (2023) who studied the temporal dynamics of bacterial, eukaryotic and Foraminiferal communities in a sediment core sampled in the bay of Pozzoli (Mediterranean sea, SW Italy). Changes in microbial communities recapitulated the different phases of the heavy industrial development of the surrounding area in the period 1851-1992 that left geochemical signatures in the core layers. Both these studies (Siano et al., 2021; Barrenechea Angeles et al., 2023) highlight the initial status of microbial communities before human impact. This status could be considered as a baseline value that represents the target to reach in a restoration operation of the polluted sites.

Besides studies targeting a single geographic site that recapitulates its local history, multisite studies allow evaluating the impact of more widespread environmental changes. By studying sediment cores sampled in *ca* 50 lakes in France along a ca 2000 m elevation gradient (Keck et al., 2020; Barouillet et al., 2022) aimed at assessing the impact of the so-called Anthropocene "great acceleration" on freshwater ecosystems. For each core, DNA was extracted from one layer that deposited in the 19th century and from a second more modern one representative of extant microbial communities. Metabarcoding analyses targeting either the whole eukaryotic microbial communities (Keck et al., 2020) or more specifically the ciliates (Barouillet et al., 2022), gave convergent results. In a global comparative analysis, it emerged that modern eukaryotic communities of the 50 lakes displayed greater similarity between them than their 19th century counterparts. This spatial homogenization was stronger for lakes located below 1400m above sea level than for those located above. In France, this altitude globally marks the upper limit of permanent human settlements and territories above this line are unlikely to be directly impacted by intensive farming practices that may represent the main causes of diversity changes. Functional assignation of molecular taxa also highlighted pervasive shifts in the functional profile of lake eukaryotic microbial communities. Modern communities were significantly

enriched in mixotrophic and photosynthetic taxa at the expense of primary consumers, parasitic and saprotrophic species that were possibly counter-selected by eutrophication of the lacustrine ecosystems indicated by the higher organic carbon concentrations found in the most recent sediment layers.

Host-associated ancient microbial communities

Host-associated microbial communities can be affected by the characteristics of their host (species, genotype), its lifestyle *(e.g.* its diet, its health status) and the environment in which it evolves and where part or all of its microbiome is recruited horizontally. Several studies have examined the relative importance of the host, geography, ecology and environment in the evolution of these communities, which contribute to the fitness and health of their host and partly determine its selective value.

Diversity of source materials

Dental calculus. As in the case of studies targeting a specific microbial species, many paleomicrobiology studies of host-associated communities regard humans and related species (hominids, monkeys). The most numerous studies focus on the oral ecosystem where mineralization of dental plaque leads to the formation of dental calculus (**Fig. 1, Fig. 2**). It protects the microorganisms that are present in this environment, whose DNA remains preserved over very long periodsof time (Ozga & Ottoni, 2023). Interestingly, the ancient DNA of dental calculi from skulls preserved in natural history collections is now being studied not in a historical context, but simply because it allows studying oral microbial communities associated with rare animal species in danger of extinction, whose extant populations have become difficult to sample. An illustration of this approach is the study carried out on three extant species/sub-species of Gorillas, whose skulls were preserved in various natural history collections (Moraitou et al., 2022). Sequencing of degraded DNA from the corresponding dental calculi indicated that the nature of the Gorilla's diet (ecology), more than the phylogenetic proximity of Gorilla species, determined the composition of the oral microbiome.

Paleofeces and coprolites. These remains represent dried and mineralized fossilized feces respectively. A paleomicrobiological analysis of coprolites has already provided information about a variety of organisms, including micro-eukaryotes, bacteria, and archaea, present in this material, thus enhancing our understanding of ancient human diet, gut microbiota, and intestinal and systemic diseases (Appelt et al., 2016). Feces are more rarely preserved with their original intestinal microbiome, but exceptional cases of preservation do exist, such as in salt-rich environments (salt mines) (Maixner et al., 2021) or dry desert ones (Wibowo et al., 2021a) that allow for a rapid desiccation of the samples.

Zoological collections. In the animal kingdom, studies are however not limited to mammals, several papers have for instance reported the composition of the gut microbiome of specimens of other animal taxa preserved in alcohol in natural history collections. This is the case of related Mexican *Herichthys* endemic fish species that; for several specimens; were stored for about 50 years in alcohol (Mejía et al., 2022), or for terrestrial snails, several of which collected 98 years before analysis (Chalifour et al., 2022). As in the case of dental calculus, mineralized structures are however more likely to preserve DNA for longer periods. Thus, Scott et al. (2022) successfully extracted and sequenced DNA from millennia-old corals that revealed their original microbiomes which showed similarities to those from today. However, they identified very few sequences that could be affiliated to Symbiodiniaceae, that are essential eukaryotic phototrophic symbionts of corals.

Herbarium collections. In the case of plants, both aerial and underground (roots) organs of dried herbarium plants have been examined for their microbiome (Heberling & Burke, 2019; Bieker et al., 2020). As most herbaria have been constituted in the 19th and 20th centuries with very few specimens from the 17th and 18th centuries, other sources of plant material have to be looked for to address plant microbiome evolution across longer periods. Ancient plant DNA has been extracted from up to ca 10,000-year-old woods preserved in waterlogged environments; however, their associated microbiomes probably correspond to "post-mortem" communities recruited from the surrounding sediments, and not to the original endophytic ones (Wagner et al., 2018). DNA has also been extracted from seeds or inflorescence of cultivated plants collected in archaeological sites (Kistler et al., 2020; Trucchi et al., 2021), but their microbiome has, thus far, not been specifically reported.

Monitoring host-associated-microbiome taxonomic and functional shifts

In this section, using selected examples, we describe studies on past host-associated microbial communities that investigate either their taxonomic or functional diversity. Both approaches illustrate how microbial communities evolve over time and how environmental factors can influence the microbiome composition, its metabolism and ultimately its activities and roles in its original ecosystem.

Evolution of animal/human microbiomes and taxonomic diversity. Focusing on human dental calculus, evolution of its microbiome has been addressed at different time and geographic scales. Deep in time, Fellows Yates et al. (2021) compared calculus microbiomes of different extant (Gorillas, Chimpanzees, *Homo sapiens*) and extinct (*H. neanderthalensis*) Hominids to delineate the set of microbial species shared between the different species (i.e. the Hominid calculus core microbiome) and the set of taxa specific, or more frequent, in one or more species.

As opposed to metabarcoding, thus far commonly used to explore non-host associated paleo communities, systematic sequencing allows either the de novo assembly of entire microbial genomes (MAGs for Metagenome Assembled Genomes) or the mapping of sequencing reads against already known microbial genomes. Identification of different genomes of the same microbial taxon in different samples allows delineating intraspecific lineages whose distribution in time and space can be studied. This latter approach was presented by Fellows Yates et al. (2021) for three bacterial species found in all studied Hominid species. Phylogenetic analyses based on Single Nucleotide Polymorphisms (SNPs) identified in whole genome sequence data highlighted a greater proximity between human associated lineages (from either modern or ancient individuals) that were distinct from Gorilla and Chimpanzee ones that grouped together. Thus, although these three bacterial species belonged to the core Hominid oral microbiome, genome-level analyses split each of these taxa in different host-specific groups. A number of potential confounding factors can however contribute to these observed associations between hosts. One of them is geography as almost all human samples were of European origin while Gorillas and Chimpanzees co-occur in Central Africa.

Data interpretation in paleomicrobiology, especially when referring to microbial communities, is indeed often subject to caution since available archaeological samples *(e.g.* human remains) are often rare and not evenly distributed (and therefore available for analysis) across time, space and ecological gradients. It is thus often difficult to disentangle the relative contribution of each of these factors and others on the genetic make-up, diversity and evolution of past host-associated microbial communities. This may explain some conflicting results in studies addressing for example the consequences of the Neolithic transition to agriculture in Europe on the human oral (dental calculus as a proxy) microbiome. Because this progressive transition lasted thousands of years, involved human migrations,

episodes of cohabitation between different human groups and did not probably proceed at the same pace in different geographic areas, differences in sample selection may lead to discrepancies across studies. Thus, while Ottoni et al. (2021) concluded that the introduction of farming did not significantly alter the oral microbiome present in ancient foragers, Quagliariello et al. (2022) concluded the opposite. Both studies followed the same approach based on the systematic sequencing of aDNA and the assembly of ancient bacterial genomes. They shared nevertheless a number of common observations, such as the higher frequency of specific taxa *(e.g. Olsenella sp. Oral taxon 807*) in Neolithic farmers compared to earlier hunter-gatherers.

At the intraspecific level, as exposed above for different Hominid species (Fellows Yates et al., 2021), association between specific bacterial lineages and specific human groups has also been reported. In the case of commensal *Anaerolineaceae oral taxon 438* specific lineages each associated to a specific geographical and chronological group (Mesolithic-Neolithic) of Humans in Europe were identified (Ottoni et al., 2021) as well as others, specific to Japanese Jomon hunter gatherers (-3000 y BP), Japanese Edo agriculturalists (400-140 BP) (Eisenhofer et al., 2020) or native Wichita north Americans (1250-1450 CE) (Honap et al., 2023).

This last observation was part of a study that aimed at understanding the impact of colonization by Europeans of North-America on the native north-American populations' microbiomes (Honap et al., 2023). Since this historical event is far better documented and took place over a shorter time span (a few centuries) compared to the Mesolithic-Neolithic transition, it could represent a more appropriate framework to disentangle the relative contribution of geography, nutrition and host (Human) genetics on the Human microbiome.

Besides dental calculi, reconstruction of nearly 500 MAGs has been reported for aDNA extracted from exceptionally well-preserved 1000-2000 years-old Native Americans human feces found in desert areas of S-W USA and nearby Mexico (Wibowo et al., 2021a). In multivariate meta-analysis these ancient microbiomes grouped with modern ones from individuals living in "non-industrialized societies", while microbiomes from individuals from "industrialized societies" grouped together in a distinct cluster. This observation lends support to the hypothesis that the microbiome of extant humans of "non-industrialized societies" represents an ancestral state that may be explained by similarities in diets between these extant human populations and ancestral ones. Although

ancient microbiomes shared similarities with extant ones, the authors reported that 39% of the reconstructed genome sequences corresponded to microbial taxa that had not been reported previously although several thousands of reference genomes are available for the human microbiome (Wibowo et al., 2021a). While these microbial taxa may no longer be associated to extant humans, they may still be present, but associated to underexplored human populations (Almeida et al., 2021).

Beyond taxonomic diversity, functional diversity. As opposed to metabarcoding, commonly used to explore non-host associated paleo communities, systematic sequencing of aDNA offers the opportunity of exploring not only the taxonomic, but also the functional and metabolic diversity of ancient microbial communities. Thus, in their study of the oral microbiome of Hominids, Fellows Yates et al. (2021) identified the putative acquisition of a "salivary amylase-binding capability" by oral streptococci as a potential functional marker that distinguishes oral *Homo sp*. Microbiomes from other Hominid ones. Acquisition of this marker could be explained by the adoption of a starch-rich alimentation by *Homo* species. Similarly, regarding the gut microbiome, paleofeces and feces from non-industrialized extant humans are enriched in genes encoding enzymes degrading starch and glycogen (Wibowo et al., 2021a). This difference with "industrialized humans" could result from a larger intake of food products enriched in complex carbohydrates by ancestral and non-industrialized populations.

Among gene categories that have been scrutinized in both oral (dental calculi) and intestinal (feces) microbiomes, a special attention has been paid to those coding for antibiotic-resistance. In accordance with the hypothesis that the prevalence and diversity of these categories increased as a result of post-World War II antibiotic massive diffusion, it was observed that tetracycline resistance gene categories were the most enriched ones in comparisons between extant human (from both industrialized and non-industrialized societies) and ancestral one paleofeces (Wibowo et al., 2021a). Similarly, Ottoni et al. (2021) reported the exclusive presence of three antibiotic resistance gene categories in modern dental calculus, the increase in frequency of a fourth one between ancient and modern calculus, but also the disappearance of a fifth one (coding for vancomycin resistance) in the modern calculus.

All studies cited thus far used high-throughput sequence data to probe both the taxonomic and the functional diversity of ancient microbiomes, it is however

conceivable to focus exclusively on functional analyses to address a specific scientific issue. Thus, Brealey et al. (2021) addressed the historical impact of antibiotic use by humans and domestic animals on the distribution and prevalence of antibiotic resistance genes (ARGs) among bacterial communities associated to non-target wild animal species. This was achieved through the specific identification and annotation of these genes in DNA samples extracted from dental calculi of Brown Bears whose skulls were deposited in natural history collections. These samples, all from Sweden, covered a period of time ranging from 1842 until 2016. It encompassed several phases of antibiotic use in this country. While prior to 1951 antibiotics were not yet available, the 1951-1985 period corresponded to their diffusion and massive use in human and livestock. After that date, measures were implemented to reduce their use, including their ban as growth promoters. The authors observed that the prevalence, but also diversity of ARGs significantly changed over time, seemingly reflecting both the different phases of antibiotic use in Sweden and their widespread impact even on species remotely associated to humans. Schematically, ARGs increased in abundance during the 1951-1985 period when compared to the pre-1951 one, to then regress, especially in bear specimens who died after the year 2000 and spent their entire life after antibiotic restrictive measures entered into force. This study represents an additional example of the strength of molecular paleomicrobiology that provides a dynamic view of past events and can be used to evaluate the impact and validity of environmental policies.

Future prospects

As illustrated in the manuscript, molecular paleomicrobiology is revolutionizing several aspects of microbiology by inserting microorganisms in a historical framework that predates the development of microbiology as a scientific discipline (**Fig. 3**). The large number of studies on human-associated microbial species and communities identified both long-term *(e.g. Y. pestis* or Hepatitis B virus) and occasional *(e.g. M. pinnipedii*) human "companion" species. Besides revealing the long history of several human/microbe associations, by giving access to historical microbial genomes, paleomicrobiology highlights the temporal population dynamics of major pathogens and the successive acquisition of key functional attributes that are susceptible to affect the outcome of these associations. However, future research should clarify whether these temporal and functional changes are due to co-evolutionary processes between host and microbe or to competition between microbial lineages that have inadvertently entered into contact as a result of host migration.

Several studies also identified microbial species or species lineages (in the case of pathogens) that have never been reported in studies on modern samples (Granehäll et al., 2021; Wibowo et al., 2021a). These observations are likely to fuel the debate on species extinction in the microbial world and to stimulate systematic surveys to identify potential refuges where microbial species could "hide" for long periods of time. This is particularly relevant in the case of pathogens as illustrated by the apparent reemergence of the WENBA genotype

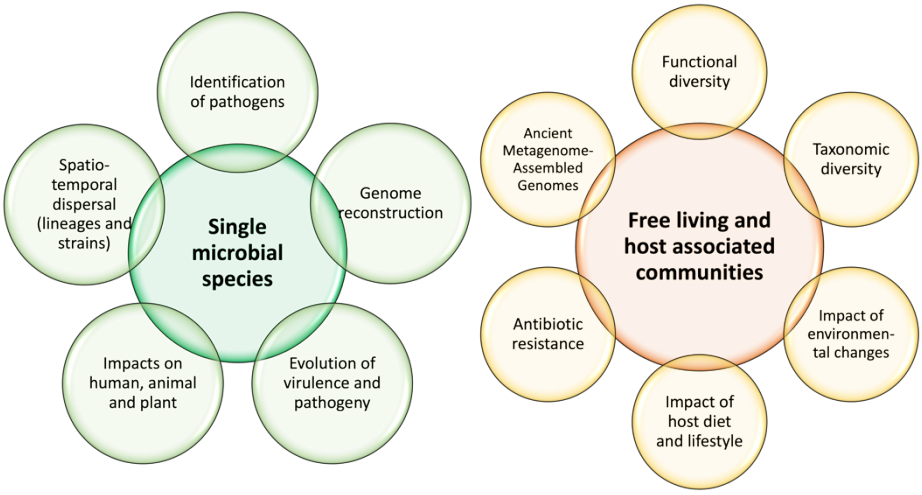


Figure 3. Main achievements and research fields explored by molecular paleomicrobiology.

of the hepatitis B virus that prevailed in Bronze Age human populations to then seemingly vanished (Kocher et al., 2021).

Several studies suggest that permafrost, which cover 11% of Earth's land surface (Obu, 2021), could represent a reservoir for “ancient”, but still alive microorganisms, including pathogens (Miner et al., 2021). For example, Liang et al. (2021) assembled similar MAGs from both intracellular iDNA and extracellular eDNA extracted from Siberian permafrost samples up to 43,000 years-old (Pleistocene-Holocene). This observation lends support to the persistence of intact bacterial cells for very long periods of time in this environment from which a variety of viruses and bacteria have been identified and put into culture (Liang et al., 2019, 2021). Thus, a recent outbreak of anthrax caused by the spore-forming *Bacillus anthracis*, which severely affected reindeer herds in Siberia, is probably connected with the thawing of the permafrost that is intensifying in the Arctic (Stella et al., 2020).

With the rapid accumulation of data, molecular paleomicrobiology could certainly benefit from the set-up of specific data repositories that would facilitate comparisons across studies and meta-analyses. Such a scientific endeavor has been undertaken in the case of ancient human genomes as illustrated by the ancient mtDNA database (amtdb, https://amtdb.org/, (Ehler et al., 2019), the Poseidon framework (https://www.poseidon-adna.org/#/) or the Allen Ancient DNA Resource (AADR), a curated version of the world's published ancient and modern human DNA data (Mallick et al., 2023). In the case of paleomicrobiology, the AncientMetagenomeDir of SPAAM-community represents a first initiative to collect published ancient metagenomics data (Fellows Yates et al., 2021).

Another perspective in molecular paleomicrobiology is to go beyond the analysis of ancient DNA extracted from ancient materials and to integrate additional genetic information deduced from the analysis of other biomolecules such as proteins or RNA. Thus, mass spectrometry analysis of proteins extracted from ancient human bones and dental calculi has identified signature peptide sequences of viruses as well as of pathogenic bacteria, such as Hepatitis virus B (Krause-Kyora et al., 2018), *Y. pestis* (Barbieri et al., 2017) or *Mycobacterium leprae* (Fotakis et al., 2020). Regarding RNA, usually considered as a highly labile molecule rapidly degraded upon cell death, it has been shown that plant virus small RNAs are surprisingly more stable than long RNAs and DNA molecules (Smith et al., 2014; Hartung et al., 2015; Rieux et al., 2021). This

property was used by Rieux et al. (2021) to assemble a genome of Cassava mosaic virus (ACMV) using small RNAs extracted from a 90 years-old Cassava herbarium plant herbarium samples. Analysis of small RNAs from ancient cellular microorganisms has not yet been reported.

A future contribution to microbial biotechnologies?

In addition, the study of ancient metabolites and natural compounds could provide new insights into the evolution and functions of ancient microbiomes through approaches that bring together paleomicrobiology and biochemistry (Velsko et al., 2017, Klapper et al., 2023). Paleobiotechnology is a recently developed discipline that enables the study of the evolution of natural products extracted directly from ancient materials, such as dental calculus (Velsko et al., 2017). By exploiting their potential biological activities, these molecules could represent new drugs to be used in medicine and pharmacology. Besides the direct extraction and characterization of metabolites from ancient samples, "paleometabolites" can also be produced by genetic engineering. This approach was followed by Klepper et al., 2023 who assembled high-quality MAGs from aDNA extracted from dental calculi of seven hominids (*H. neanderthalensis* and *H. sapiens*) who lived between the Middle and Upper Paleolithic. In MAGs affiliated to the genus *Chlorobium* the authors identified novel putative biosynthetic gene clusters (BCGs). Heterologous co-expression in Pseudomonas protegens of two enzyme-coding genes (*plfA* and *plfB*) from one of these clusters resulted in the production of two novel furan-like molecules called "paleofurans". This proof-of-concept study, which makes use of ancient microbial DNA information to characterize experimentally the metabolites produced by the corresponding microbes represents a novel way for discovering new metabolites and to address the biology of past microbial communities.

For millennia, humans have used microorganisms (bacteria, fungi) to transform raw agricultural products *(e.g.* grains, flour, milk, meat or fruit juice) into elaborated, often more digestible, and tasty food products that can be kept for longer periods of time *(e.g.* bread, cheese, fermented beverages, processed meat). Extensive studies on several of these microorganisms, as in the case of *Saccharomyces cerevisiae* or *Penicillium spp.,* clearly show that strains participating to food transformation differ from wild ones (Legras et al., 2018; Peter et al., 2018; De Chiara et al., 2022; Ropars & Giraud, 2022). Furthermore, whole genome phylogenetic analyses also tend to cluster strains according to the

food product they have been isolated from (Legras et al., 2018; Peter et al., 2018; Ropars & Giraud, 2022), thus defining in the case of *S. cerevisiae* so called "wine", "sake", "cheese", or "beer and bread" lineages (Legras et al., 2018; Peter et al., 2018). Besides phylogeny, genome comparisons coupled to reverse genetics and large-scale phenotyping have identified several key genetic determinants and phenotypes that distinguish on the one hand food-adapted strains from wild ones, and, on the other hand, strains adapted to different foodstuffs (Legras et al., 2018; De Chiara et al., 2022). These observations support the idea that some sort of inadvertent domestication process has led to the selection of these very specific microbial lineages.

We suggest that a molecular paleomicrobiological approach could well be implemented to explore the history and main stages of these domestication processes. Among the main questions that could be addressed we can cite: (i), where and to which date can we trace the first signs of domestication? (ii), for a given microbial species and a specific transformation process, has domestication occurred several times independently? (iii), can we establish a time-calibrated chronology of the genome modifications that lead to extant domesticated lineages? (iv), Can we identify extinct domesticated lineages characterized by specific genome features? (v) Had other, unsuspected, microbial species been domesticated in the past for the transformation of a specific foodstuff?

Although these questions regard individual species, paleomicrobiological investigations could be extended to microbial communities that participate to the maturation of food products and sometime define their specificity. In this framework, paleomicrobiological investigations could be carried out on a few century/decade-old food residues to evaluate recent evolutions in the preparation of so-called "traditional" local food products.

Archaeological food remains, storage vessels or tools for their preparation could be obvious sources of aDNA from foodborne microorganisms. Human paleofeces have recently been shown to, sometimes, contain significant amounts of DNA from two of these microorganisms, *Penicillium spp*. and *S. cerevisiae*, that displayed typical post-mortem damage features of aDNA (Maixner et al., 2021). Mapping of aDNA reads to modern genomes affiliated the ancient yeast sequences to a "beer clade" and the *Penicillium* ones to "non-Roquefort" *P. roqueforti* strain used nowadays for blue cheese ripening. From a historical point of view, these results suggest that Iron-age European populations may have already consumed beer and blue cheese and therefore mastered to some extent

their preparation. From a technical point of view, this study demonstrates how molecular paleomicrobiology can indeed contribute to microbial biotechnology.

Acknowledgments

Work on ancient plant-associated microbiota at the Muséum National d'Histoire Naturelle (MNHN) was supported by grant ATM 2021 (HoloHerbier) and the Emergence program of Sorbonne Université. GG was supported by a PhD grant from the University of Turin and a mobility grant from the French-Italian University (program da Vinci 2022). VB was supported by the Short Term Mobility (STM) 2023 program of the National Council of Research (CNR - Italy). We would like to thank our colleague William Conrad Ledford for improving the quality of the manuscript.

References

Afouda, P., Dubourg, G., & Raoult, D. (2020). Archeomicrobiology applied to environmental samples. *Microbial Pathogenesis*, *143*(104140). https://doi.org/10.1016/j.micpath.2020.104140

Almeida, A., Nayfach, S., Boland, M. et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology* 39, 105–114 (2021). https://doi.org/10.1038/s41587-020-0603-3

Arning, N., & Wilson, D. J. (2020). The past, present and future of ancient bacterial DNA. *Microbial Genomics*, *6*(7), 1–19. https://doi.org/10.1099/mgen.0.000384

Appelt, S., Drancourt, M., Le Bailly, M. (2016). Human Coprolites as a Source for Paleomicrobiology. *Microbiology Spectrum*, 4(4). https://doi.org/10.1128/microbiolspec.poh-0002-2014

Barbieri, R., Mekni, R., Levasseur, A., Chabrière, E., Signoli, M., Tzortzis, S., Aboudharam, G., & Drancourt, M. (2017). Paleoproteomics of the dental pulp: The plague paradigm. *PLoS ONE*, *12*(7), e0180552. https://doi.org/10.1371/journal.pone.0180552

Barouillet, C., Vasselon, V., Keck, F., Millet, L., Etienne, D., Galop, D., Rius, D., & Domaizon, I. (2022). Paleoreconstructions of ciliate communities reveal long-term ecological changes in temperate lakes. *Scientific Reports*, *12*(1), 1–12. https://doi.org/10.1038/s41598-022-12041-7

Barrenechea Angeles, I., Romero-Martínez, M. L., Cavaliere, M., Varrella, S., Francescangeli, F., Piredda, R., Mazzocchi, M. G., Montresor, M., Schirone, A., Delbono, I., Margiotta, F., Corinaldesi, C., Chiavarini, S., Montereali, M. R., Rimauro, J., Parrella, L., Musco, L., Dell'Anno, A., Tangherlini, M., … Frontalini, F. (2023). Encapsulated in sediments: eDNA deciphers the ecosystem history of one of the most polluted European marine sites. *Environment International*, *172*, 107738. https://doi.org/10.1016/J.ENVINT.2023.107738

Bieker, V. C., Sánchez Barreiro, F., Rasmussen, J. A., Brunier, M., Wales, N., & Martin, M. D. (2020). Metagenomic analysis of historical herbarium specimens reveals a postmortem microbial community. *Molecular Ecology Resources*, *20*(5), 1206–1219. https://doi.org/10.1111/1755-0998.13174

Bonczarowska, J. H., Susat, J., Mühlemann, B., Jasch-Boley, I., Brather, S., Höke, B., Brather-Walter, S., Schoenenberg, V., Scheschkewitz, J., Graenert, G., Krausse, D., Francken, M., Jones, T. C., Wahl, J., Nebel, A., & Krause-Kyora, B. (2022). Pathogen genomics study of an early medieval community in Germany reveals extensive co-infections. *Genome Biology*, *23*(250). https://doi.org/10.1186/S13059-022-02806-8

Bos, K. I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S. A., Klunk, J., Schuenemann, V. J., Poinar, D., Kuch, M., Golding, G. B., Dutour, O., Keim, P., Wagner, D. M., Holmes, E. C., Krause, J., & Poinar, H. N. (2016). Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. *ELife*, *5*(e12994). https://doi.org/10.7554/ELIFE.12994

Bos, K. I., Kühnert, D., Herbig, A., Esquivel-Gomez, L. R., Andrades Valtueña, A., Barquera, R., Giffin, K., Lankapalli, A. K., Nelson, E. A., Sabin, S., Spyrou, M. A., & Krause, J. (2019). Paleomicrobiology: Diagnosis and Evolution of Ancient Pathogens. *Annual Review of Microbiology*, *73*, 639–666. https://doi.org/10.1146/annurev-micro-090817

Bramanti, B., Wu, Y., Yang, R., Cui, Y., & Stenseth, N. C. (2021). Assessing the origins of the european plagues following the black death: A synthesis of genomic, historical, and ecological information. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(36), e2101940118. https://doi.org/https://doi.org/10.1073/pnas.2101940118

Brealey, J. C., Leitão, H. G., Hofstede, T., Kalthoff, D. C., & Guschanski, K. (2021). The oral microbiota of wild bears in Sweden reflects the history of antibiotic use by humans. *Current Biology*, *31*(20), 4650-4658.e6. https://doi.org/10.1016/J.CUB.2021.08.010

Brunson, K., & Reich, D. (2019). The Promise of Paleogenomics Beyond Our Own Species. *Trends in Genetics*, *35*(5), 319–329. https://doi.org/10.1016/J.TIG.2019.02.006

Chalifour, B. N., Elder, L. E., & Li, J. (2022). Gut microbiome of century-old snail specimens stable across time in preservation. *Microbiome*, *10*(1), 1–16. https://doi.org/10.1186/S40168-022-01286-Z/

Clark, I., & Hirsch, P. (2008). Survival of bacterial DNA and culturable bacteria in archived soils from the Rothamsted broadbalk experiment. *Soil Biology and Biochemistry*, 40(5), 1090–1102. https://doi.org/10.1016/j.soilbio.2007.11.021

Cooper, A., & Poinar, H. N. (2000). Ancient DNA: Do It Right or Not at All. *Science*, *289*(5482), 1139–1139. https://doi.org/10.1126/SCIENCE.289.5482.1139B

Dabney, J., Meyer, M., & Pääbo, S. (2013). Ancient DNA Damage. *Cold Spring Harbor Perspectives in Biology*, *5*(7). https://doi.org/10.1101/CSHPERSPECT.A012567

Davis, N.M., Proctor, D.M., Holmes, S.P. et al. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 6, 226 (2018). https://doi.org/10.1186/s40168-018-0605-2

De Chiara, M., Barré, B. P., Persson, K., Irizar, A., Vischioni, C., Khaiwal, S., Stenberg, S., Amadi, O. C., Žun, G., Doberšek, K., Taccioli, C., Schacherer, J., Petrovič, U., Warringer, J., & Liti, G. (2022). Domestication reprogrammed the budding yeast life cycle. *Nature Ecology & Evolution* , *6*(4), 448–460. https://doi.org/10.1038/s41559-022-01671-9

Demeure, C. E., Dussurget, O., Guillem, F. M., Le Guern, A.-S., Savin, C., & Pizarro-Cerdá, J. (2019). Yersinia pestis and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination, and diagnostics. *Genes & Immunity*, *20*, 357–370. https://doi.org/10.1038/s41435-019-0065-0

Ehler, E., Novotńy, J., Juras, A., Chylénski, M., Moravčik, O., & Pačes, J. (2019). AmtDB: A database of ancient human mitochondrial genomes. *Nucleic Acids Research*, *47*(D1), D29–D432. https://doi.org/10.1093/nar/gky843

Eisenhofer, R., Kanzawa-Kiriyama, H., Shinoda, K. I., & Weyrich, L. S. (2020). Investigating the demographic history of Japan using ancient oral microbiota. *Philosophical Transactions of the Royal Society B*, *375*(1812). https://doi.org/10.1098/RSTB.2019.0578

Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). The microbial engines that drive earth's biogeochemical cycles. *Science*, *320*(5879), 1034–1039. https://doi.org/10.1126/science.1153213

Fellows Yates, J. A., Andrades Valtueña, A., Vågene, Å. J., Cribdon, B., Velsko, I. M., Borry, M., Bravo-Lopez, M. J., Fernandez-Guerra, A., Green, E. J., Ramachandran, S. L., Heintzman, P. D., Spyrou, M. A., Hübner, A., Gancz, A. S., Hider, J., Allshouse, A. F., Zaro, V., & Warinner, C. (2021). Community-curated and standardised metadata of published ancient metagenomic samples with AncientMetagenomeDir. *Scientific Data*, *8*(31). https://doi.org/10.1038/s41597-021-00816-y

Fellows Yates, J. A., Velsko, I. M., Aron, F., Posth, C., Hofman, C. A., Austin, R. M., Parker, C. E., Mann, A. E., Nagele, K., Arthur, K. W., Arthur, J. W., Bauer, C. C., Crevecoeur, I., Cupillard, C., Curtis, M. C., Dalen, L., Bonilla, M. D. Z., Fernandez-Lomana, J. C. D., Drucker, D. G., … Warinner, C. (2021). The evolution and changing ecology of the African hominid oral microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(20), e2021655118. https://doi.org/https://www.pnas.org/doi/abs/10.1073/pnas.2021655118

Fischer, W. W., Hemp, J., & Johnson, J. E. (2016). Evolution of Oxygenic Photosynthesis. *Annual Review of Earth and Planetary Sciences*, *44*, 647–683. https://doi.org/10.1146/annurev-earth-060313-054810

Fotakis, A. K., Denham, S. D., MacKie, M., Orbegozo, M. I., Mylopotamitaki, D., Gopalakrishnan, S., Sicheritz-Pontén, T., Olsen, J. V., Cappellini, E., Zhang, G.,

Christophersen, A., Gilbert, M. T. P., & Vågene, Å. J. (2020). Multi-omic detection of Mycobacterium leprae in archaeological human dental calculus: M. leprae from dental calculus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1812). https://doi.org/10.1098/rstb.2019.0584

Granehäll, L., Huang, K. D., Tett, A., Manghi, P., Paladin, A., O'Sullivan, N., Rota-Stabelli, O., Segata, N., Zink, A., & Maixner, F. (2021). Metagenomic analysis of ancient dental calculus reveals unexplored diversity of oral archaeal Methanobrevibacter. *Microbiome* , *9*(1), 1–18. https://doi.org/10.1186/S40168-021-01132-8

Gubry-Rangin, C., Kratsch, C., Williams, T. A., Mchardy, A. C., Martin Embley, T., Prosser, J. I., & Macqueen, D. J. (2015). Coupling of diversification and pH adaptation during the evolution of terrestrial Thaumarchaeota. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(30), 9370–9375. https://doi.org/10.5061/dryad.0nv00

Harbeck, M., Seifert, L., Hänsch, S., Wagner, D. M., Birdsell, D., Parise, K. L., Wiechmann, I., Grupe, G., Thomas, A., Keim, P., Zöller, L., Bramanti, B., Riehm, J. M., & Scholz, H. C. (2013). Yersinia pestis DNA from Skeletal Remains from the 6th Century AD Reveals Insights into Justinianic Plague. *PLOS Pathogens*, *9*(5), e1003349. https://doi.org/10.1371/JOURNAL.PPAT.1003349

Hartung, J. S., Roy, A., Fu, S., Shao, J., Schneider, W. L., & Brlansky, R. H. (2015). History and diversity of Citrus Leprosis virus recorded in herbarium specimens. *Phytopathology*, *105*(9), 1277–1284. https://doi.org/10.1094/PHYTO-03-15-0064-R

Heberling, J. M., & Burke, D. J. (2019). Utilizing herbarium specimens to quantify historical mycorrhizal communities. *Applications in Plant Sciences*, *7*(4). https://doi.org/10.1002/APS3.1223

Hinnebusch, B. J., Jarrett, C. O., & Bland, D. M. (2021). Molecular and genetic mechanisms that mediate transmission of yersinia pestis by fleas. *Biomolecules*, *11*(2), 1–13. https://doi.org/10.3390/BIOM11020210

Honap, T. P., Monroe, C. R., Johnson, S. J., Jacobson, D. K., Abin, C. A., Austin, R. M., Sandberg, P., Levine, M., Sankaranarayanan, K., & Lewis, C. M. (2023). Oral metagenomes from Native American Ancestors reveal distinct microbial lineages in the pre-contact era. *American Journal of Biological Anthropology*, 1–15. https://doi.org/10.1002/AJPA.24735

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hernsdorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C., & Banfield, J. F. (2016). *A new view of the tree of life*. https://doi.org/10.1038/NMICROBIOL.2016.48

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Keck, F., Millet, L., Debroas, D., Etienne, D., Galop, D., Rius, D., & Domaizon, I. (2020). Assessing the response of micro-eukaryotic diversity to the Great Acceleration using lake sedimentary DNA. *Nature Communication*, *11*(1), 1–8. https://doi.org/10.1038/s41467-020-17682-8

Kistler, L., Thakar, H. B., VanDerwarker, A. M., Domic, A., Bergström, A., George, R. J., Harper, T. K., Allaby, R. G., Hirth, K., & Kennett, D. J. (2020). Archaeological Central American maize genomes suggest ancient gene flow from South America. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(52), 33124–33129. https://doi.org/10.1073/PNAS.2015560117

Kjær, K. H., Winther Pedersen, M., De Sanctis, B., De Cahsan, B., Korneliussen, T. S., Michelsen, C. S., Sand, K. K., Jelavić, S., Ruter, A. H., Schmidt, A. M. A., Kjeldsen, K. K., Tesakov, A. S., Snowball, I., Gosse, J. C., Alsos, I. G., Wang, Y., Dockter, C., Rasmussen, M., Jørgensen, M. E., … Willerslev, E. (2022). A 2-million-year-old ecosystem in Greenland

uncovered by environmental DNA. *Nature*, *612*(7939), 283–291. https://doi.org/10.1038/s41586-022-05453-y

Klapper, M., Hübner, A., Ibrahim, A., Wasmuth, I., Borry, M., Haensch, V. G., Zhang, S., Al-Jammal, W. K., Suma, H., Fellows Yates, J. A., Frangenberg, J., Velsko, I. M., Chowdhury, S., Herbst, R., Bratovanov, E. V., Dahse, H.-M., Horch, T., Hertweck, C., González Morales, M. R., … Stallforth, P. (2023). Natural products from reconstructed bacterial genomes of the middle and Upper Paleolithic. *Science*, *380*(6645), 619–624. https://doi.org/10.1126/science.adf5300

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., Bushman, F. D., Knight, R., & Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nature Methods* , *8*(9), 761–763. https://doi.org/10.1038/nmeth.1650

Kocher, A., Papac, L., Barquera, R., Key, F. M., Spyrou, M. A., Hübler, R., Rohrlach2, A. B., Aron, F., Stahl, R., Wissgott, A., Van Bömme, F., Pfefferkorn, M., Alissa, M., Villalba-Mouco, V., Neumann, G. U., Rivollat, M., Van De Loosdrecht, M. S., Majander, K., Tukhbatova, R. I., … Kocher, A. (2021). Ten millennia of hepatitis B virus evolution. *Science*, *374*(6564), 132. https://doi.org/https://www.science.org/doi/10.1126/science.abi5658

Krause-Kyora, B., Susat, J., Key, F. M., Kühnert, D., Bosse, E., Immel, A., Rinne, C., Kornell, S. C., Yepes, D., Franzenburg, S., Heyne, H. O., Meier, T., Lösch, S., Meller, H., Friederich, S., Nicklisch, N., Alt, K. W., Schreiber, S., Tholey, A., … Krause, J. (2018). Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *ELife*, *7*. https://doi.org/10.7554/ELIFE.36666.001

Kuyl, A. C. van der. (2022). Historic and Prehistoric Epidemics: An Overview of Sources Available for the Study of Ancient Pathogens. *Epidemiologia*, *3*(4), 443–464. https://doi.org/10.3390/EPIDEMIOLOGIA3040034

Lebreton, F., Manson, A. L., Saavedra, J. T., Straub, T. J., Earl, A. M., & Gilmore, M. S. (2017). Tracing the Enterococci from Paleozoic Origins to the Hospital. *Cell*, *169*(5), 849-861.e13. https://doi.org/10.1016/j.cell.2017.04.027

Legras, J. L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., Sanchez, I., Couloux, A., Guy, J., Franco-Duarte, R., Marcet-Houben, M., Gabaldon, T., Schuller, D., Sampaio, J. P., & Dequin, S. (2018). Adaptation of S. Cerevisiae to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. *Molecular Biology and Evolution*, *35*(7), 1712–1727. https://doi.org/10.1093/molbev/msy066

Lennon, J. T., & Locey, K. J. (2020). More support for Earth's massive microbiome. *Biology Direct*, *15*(1), 1–6. https://doi.org/10.1186/S13062-020-00261-8/

Lezcano, M. Á., Sánchez-García, L., Quesada, A., Carrizo, D., Fernández-Martínez, M. Á., Cavalcante-Silva, E., & Parro, V. (2022). Comprehensive Metabolic and Taxonomic Reconstruction of an Ancient Microbial Mat From the McMurdo Ice Shelf (Antarctica) by Integrating Genetic, Metaproteomic and Lipid Biomarker Analyses. *Frontiers in Microbiology*, *13*, 799360. https://doi.org/10.3389/FMICB.2022.799360/BIBTEX

Liang, R., Lau, M., Vishnivetskaya, T., Lloyd, K. G., Wang, W., Wiggins, J., Miller, J., Pfiffner, S., Rivkina, E. M., Onstott, T. C., & Liang, C. R. (2019). Predominance of Anaerobic, Spore-Forming Bacteria in Metabolically Active Microbial Communities from Ancient Siberian Permafrost. *Applied Environmental Microbiology*, *85*(15), 00560–19. https://journals.asm.org/journal/aem

Liang, R., Li, Z., Lau Vetter, M. C. Y., Vishnivetskaya, T. A., Zanina, O. G., Lloyd, K. G., Pfiffner, S. M., Rivkina, E. M., Wang, W., Wiggins, J., Miller, J., Hettich, R. L., & Onstott, T. C. (2021). Genomic reconstruction of fossil and living microorganisms in ancient Siberian permafrost. *Microbiome*, *9*(110). https://doi.org/10.1186/s40168-021-01057-2

Lorén, J. G., Farfán, M., & Fusté, M. C. (2014). Molecular phylogenetics and temporal diversification in the genus Aeromonas based on the sequences of five housekeeping genes. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0088805

Loucaid, S., Mazelid, F., Doebeli, M., & Parfrey, L. W. (2019). A census-based estimate of Earth's bacterial and archaeal diversity. *PLOS Biology*, *17(2): e3000106.* https://doi.org/10.1371/journal.pbio.3000106

Maixner, F., Sarhan, M. S., Huang, K. D., Tett, A., Schoenafinger, A., Zingale, S., Blanco-Míguez, A., Manghi, P., Cemper-Kiesslich, J., Rosendahl, W., Kusebauch, U., Morrone, S. R., Hoopmann, M. R., Rota-Stabelli, O., Rattei, T., Moritz, R. L., Oeggl, K., Segata, N., Zink, A., … Kowarik, K. (2021). Hallstatt miners consumed blue cheese and beer during the Iron Age and retained a non-Westernized gut microbiome until the Baroque period. *Current Biology*, *31*(23), 5149-5162.e6. https://doi.org/10.1016/j.cub.2021.09.031

Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., Patterson, N., & Reich, D. (2023). The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes. *BioRxiv*, preprint. https://doi.org/10.1101/2023.04.06.535797

Marin, J., Battistuzzi, F. U., Brown, A. C., & Hedges, S. B. (2017). The Timetree of Prokaryotes: New Insights into Their Evolution and Speciation. *Molecular Biology and Evolution*, *34*(2), 437–446. https://doi.org/10.1093/molbev/msw245

Martin, M. D., Cappellini, E., Samaniego, J. A., Zepeda, M. L., Campos, P. F., Seguin-Orlando, A., Wales, N., Orlando, L., Ho, S. Y. W., Dietrich, F. S., Mieczkowski, P. A., Heitman, J., Willerslev, E., Krogh, A., Ristaino, J. B., & Gilbert, M. T. P. (2013). Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nature Communications*, *4*(1), 1–7. https://doi.org/10.1038/ncomms3172

Marx, V. (2016). Microbiology: The road to strain-level identification. *Nature Methods*, *13*(5), 401–404. https://doi.org/10.1038/nmeth.3837

McElwain, J. C., & Punyasena, S. W. (2007). Mass extinction events and the plant fossil record. *Trends in Ecology and Evolution*, *22*(10), 548–557. https://doi.org/10.1016/j.tree.2007.09.003

Mejía, O., Sánchez-Quinto, A., Gómez-Acata, E. S., Pérez-Miranda, F., & Falcón, L. I. (2022). Unraveling the Gut Microbiome of the Genus Herichthys (Pisces: Cichlidae): What Can We Learn from Museum Specimens? *Current Microbiology*, *79*(11), 346. https://doi.org/10.1007/S00284-022-03047-5

Michael J., S. and M. J. D. (1996). Reconstructing shifts in diversification rates on phylogenetic trees. *Trends in Ecology & Evolution*, *11*, 15–20.

Miner, K. R., D'Andrilli, J., Mackelprang, R., Edwards, A., Malaska, M. J., Waldrop, M. P., & Miller, C. E. (2021). Emergent biogeochemical risks from Arctic permafrost degradation. *Nature Climate Change*, *11*(10), 809–819. https://doi.org/10.1038/s41558-021-01162-y

Moraitou, M., Forsythe, A., Fellows Yates, J. A., Brealey, J. C., Warinner, C., & Guschanski, K. (2022). Ecology, Not Host Phylogeny, Shapes the Oral Microbiome in Closely Related Species. *Molecular Biology and Evolution*, *39*(12). https://doi.org/10.1093/MOLBEV/MSAC263

Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*, *17*(4), 508–525. https://doi.org/10.1111/ele.12251

Morlon, H., Kemps, B. D., Plotkin, J. B., & Brisson, D. (2012). Explosive radiation of a bacterial species group. *Evolution*, *66*(8), 2577–2586. https://doi.org/10.1111/j.1558-5646.2012.01598.x

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I.-M., Huntemann, M., Palaniappan, K., Ladau, J., Mukherjee, S., K Reddy, T. B., Nielsen, T., Kirton, E., Faria, J. P., Edirisinghe, J. N., Henry, C. S., … Eloe-Fadrosh, E. A. (2021). A genomic catalog of Earth's microbiomes. *Nature Biotechnology*, *39*, 499–509. https://doi.org/10.1038/s41587-020-0718-6

Nee, S., Holmes, E. C., May, R. M., & Harvey, P. H. (1994). Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *344*, 77–82. https://royalsocietypublishing.org/

Nee, S., May, R. M., & Harvey, P. H. (1994). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *344*, 305–311. https://royalsocietypublishing.org/

Nwosu, E. C., Roeser, P., Yang, S., Ganzert, L., Dellwig, O., Pinkerneil, S., Brauer, A., Dittmann, E., Wagner, D., & Liebner, S. (2021). From water into sediment—tracing freshwater Cyanobacteria via DNA analyses. *Microorganisms*, *9*(8), 1778. https://doi.org/10.3390/MICROORGANISMS9081778/S1

Obu, J. (2021). How Much of the Earth's Surface is Underlain by Permafrost? *Journal of Geophysical Research: Earth Surface*, *126*(5). https://doi.org/10.1029/2021JF006123

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., Krause, J., Willerslev, E., Stone, A. C., & Warinner, C. (2021). Ancient DNA analysis. In *Nature Reviews Methods Primers* (Vol. 1, Issue 1). Springer Nature. https://doi.org/10.1038/s43586-020-00011-0

Ottoni, C., Boric, D., Cheronet, O., Sparacello, V., Dori, I., Coppa, A., Antonovic, D., Vujevic, D., Price, T. D., Pinhasi, R., & Cristiani, E. (2021). Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(32), e2102116118. https://doi.org/https://www.pnas.org/doi/abs/10.1073/pnas.2102116118

Ozga, A. T., & Ottoni, C. (2023). Dental calculus as a proxy for animal microbiomes. *Quaternary International*, *653–654*, 47–52. https://doi.org/10.1016/J.QUAINT.2021.06.012

Pawłowska, J., Lejzerowicz, F., Esling, P., Szczuciński, W., Zajaogonekczkowski, M., & Pawlowski, J. (2014). Ancient DNA sheds new light on the Svalbard foraminiferal fossil record of the last millennium. *Geobiology*, *12*(4), 277–288. https://doi.org/10.1111/GBI.12087

Pawlowski, J., Holzmann, M., Berney, C., Fahrni, J., Gooday, A. J., Cedhagen, T., Habura, A., & Bowser, S. S. (2003). The evolution of early Foraminifera. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(20), 11494–11498. https://doi.org/https://www.pnas.org/doi/abs/10.1073/pnas.2035132100

Pearman, J. K., Biessy, L., Howarth, J. D., Vandergoes, M. J., Rees, A., & Wood, S. A. (2022). Deciphering the molecular signal from past and alive bacterial communities in aquatic sedimentary archives. *Molecular Ecology Resources*, *22*(3), 877–890. https://doi.org/10.1111/1755-0998.13515

Pedersen, M. W., Overballe-Petersen, S., Ermini, L., der Sarkissian, C., Haile, J., Hellstrom, M., Spens, J., Thomsen, P. F., Bohmann, K., Cappellini, E., Schnell, I. B., Wales, N. A., Carøe, C., Campos, P. F., Schmidt, A. M. Z., Gilbert, M. T. P., Hansen, A. J., Orlando, L., & Willerslev, E. (2015). Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1660). https://doi.org/10.1098/rstb.2013.0383

Peter, J., De Chiara, M., Friedrich, A., Yue, J. X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J. M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., … Schacherer, J. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature*, *556*(7701), 339–344. https://doi.org/10.1038/s41586-018-0030-5

Quagliariello, A., Modi, A., Innocenti, G., Zaro, V., Conati Barbaro, C., Ronchitelli, A., Boschin, F., Cavazzuti, C., Dellù, E., Radina, F., Sperduti, A., Bondioli, L., Ricci, S., Lognoli, M., Belcastro, M. G., Mariotti, V., Caramelli, D., Mariotti Lippi, M., Cristiani, E., … Lari, M. (2022). Ancient oral microbiomes support gradual Neolithic dietary shifts towards agriculture. *Nature Communications 2022 13:1*, *13*(1), 1–14. https://doi.org/10.1038/s41467-022-34416-0

Rascovan, N., Sjögren, K. G., Kristiansen, K., Nielsen, R., Willerslev, E., Desnues, C., & Rasmussen, S. (2019). Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline. *Cell*, *176*(1–2), 295-305.e10. https://doi.org/10.1016/J.CELL.2018.11.005

Raup, D. M. & S. J. J. (1982). Mass Extinctions in the Marine Fossil Record. *Science* , *215*(19), 1501–1503.

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology and Evolution*, *36*(11), 1049–1060. https://doi.org/10.1016/j.tree.2021.07.009

Rieux, A., Campos, P., Duvermy, A., Scussel, S., Martin, D., Gaudeul, M., Lefeuvre, P., Becker, N., & Lett, J. M. (2021). Contribution of historical herbarium small RNAs to the reconstruction of a cassava mosaic geminivirus evolutionary history. *Scientific Reports*, *11*(21280). https://doi.org/10.1038/s41598-021-00518-w

Rivera-Perez, J. I., Santiago-Rodriguez, T. M., & Toranzos, G. A. (2016). Paleomicrobiology: a Snapshot of Ancient Microbes and Approaches to Forensic Microbiology. *Microbiology Spectrum*, *4*(4). https://doi.org/10.1128/microbiolspec.emf-0006-2015

Ropars, J., & Giraud, T. (2022). Convergence in domesticated fungi used for cheese and dry-cured meat maturation: beneficial traits, genomic mechanisms, and degeneration. *Current Opinion in Microbiology*, *70*, 102236. https://doi.org/10.1016/J.MIB.2022.102236

Scott, C. B., Cárdenas, A., Mah, M., Narasimhan, V. M., Rohland, N., Toth, L. T., Voolstra, C. R., Reich, D., & Matz, M. V. (2022). Millennia-old coral holobiont DNA provides insight into future adaptive trajectories. *Molecular Ecology*, *31*(19), 4979–4990. https://doi.org/10.1111/MEC.16642

Siano, R., Lassudrie, M., Cuzin, P., Briant, N., Loizeau, V., Schmidt, S., Ehrhold, A., Mertens, K. N., Lambert, C., Quintric, L., Noël, C., Latimier, M., Quéré, J., Durand, P., & Penaud, A. (2021). Sediment archives reveal irreversible shifts in plankton communities after World War II and agricultural pollution. *Current Biology*, *31*(12), 2682-2689.e7. https://doi.org/10.1016/J.CUB.2021.03.079

Signor, P. W. (1994). Biodiversity in Geological Time 1. *American Zoology*, *34*, 23–32. https://academic.oup.com/icb/article/34/1/23/111698

Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J., & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(6), 2229–2234. https://doi.org/10.1073/pnas.1318934111

Smith, O., Clapham, A., Rose, P., Liu, Y., Wang, J., & Allaby, R. G. (2014). A complete ancient RNA genome: Identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. *Scientific Reports*, *4*(4003). https://doi.org/10.1038/srep04003

Spyrou, M. A., Musralina, L., Gnecchi Ruscone, G. A., Kocher, A., Borbone, P. G., Khartanovich, V. I., Buzhilova, A., Djansugurova, L., Bos, K. I., Kühnert, D., Haak, W., Slavin, P., & Krause, J. (2022). The source of the Black Death in fourteenth-century central Eurasia. *Nature* , *606*(7915), 718–724. https://doi.org/10.1038/s41586-022-04800-3

Spyrou, M. A., Tukhbatova, R. I., Feldman, M., Herbig, A., Bos, K. I., Krause, J. M., & De (A. H.). (2016). Historical Y. pestis Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. *Cell Host & Microbe*, *19*, 874–881. https://doi.org/10.1016/j.chom.2016.05.012

Spyrou, M. A., Tukhbatova, R. I., Wang, C. C., Valtueña, A. A., Lankapalli, A. K., Kondrashin, V. V., Tsybin, V. A., Khokhlov, A., Kühnert, D., Herbig, A., Bos, K. I., & Krause, J. (2018). Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague. *Nature Communications* , *9*(1), 1–10. https://doi.org/10.1038/s41467-018-04550-9

Staats, M., Cuenca, A., Richardson, J. E., Vrielink-Van Ginkel, R., & Petersen, G. (2011). DNA Damage in Plant Herbarium Tissue. *PLoS ONE*, *6*(12), 28448. https://doi.org/10.1371/journal.pone.0028448

Stella, E., Mari, L., Gabrieli, J., Barbante, C., & Bertuzzo, E. (2020). Permafrost dynamics and the risk of anthrax transmission: a modelling study. *Scientific Reports*, *10*(16460). https://doi.org/10.1038/s41598-020-72440-6

Stilianos Louca, Patrick M. Shi, Matthew W. Pennell, Woodward W. Fischer, Laura Wegener Parfrey, & Michael Doebeli. (2018). Bacterial diversification through geological time. *Nature Ecology and Evolution*, *2*, 1458–1467.

Susat, J., Bonczarowska, J. H., Pētersone-Gordina, E., Immel, A., Nebel, A., Gerhards, G., & Krause-Kyora, B. (2020). Yersinia pestis strains from Latvia show depletion of the pla virulence gene at the end of the second plague pandemic. *Scientific Reports* , *10*(1), 1–10. https://doi.org/10.1038/s41598-020-71530-9

Talas, L., Stivrins, N., Veski, S., Tedersoo, L., & Kisand, V. (2021). Sedimentary ancient DNA (Sedadna) reveals fungal diversity and environmental drivers of community changes throughout the holocene in the present boreal lake Lielais Svētin, U (eastern Latvia). *Microorganisms*, *9*(4), 719. https://doi.org/10.3390/MICROORGANISMS9040719/S1

Tricou, T., Tannier, E., & de Vienne, D. M. (2022). Ghost lineages can invalidate or even reverse findings regarding gene flow. *PLoS Biology*, *20*(9), e3001776. https://doi.org/10.1371/JOURNAL.PBIO.3001776

Trucchi, E., Benazzo, A., Lari, M., Iob, A., Vai, S., Nanni, L., Bellucci, E., Bitocchi, E., Raffini, F., Xu, C., Jackson, S. A., Lema, V., Babot, P., Oliszewski, N., Gil, A., Neme, G., Michieli, C. T., De Lorenzi, M., Calcagnile, L., … Bertorelle, G. (2021). Ancient genomes reveal early Andean farmers selected common beans while preserving diversity. *Nature Plants* , *7*(2), 123–128. https://doi.org/10.1038/s41477-021-00848-7

Vågene, Å. J., Honap, T. P., Harkins, K. M., Rosenberg, M. S., Giffin, K., Cárdenas-Arroyo, F., Leguizamón, L. P., Arnett, J., Buikstra, J. E., Herbig, A., Krause, J., Stone, A. C., & Bos, K. I. (2022). Geographically dispersed zoonotic tuberculosis in pre-contact South American human populations. *Nature Communications*, *13*(1). https://doi.org/10.1038/S41467-022-28562-8

Valtueña, A. A., Mittnik, A., Key, F. M., Haak, W., Allmäe, R., Belinskij, A., Daubaras, M., Feldman, M., Jankauskas, R., Janković, I., Massy, K., Novak, M., Pfrengle, S., Reinhold, S., Šlaus, M., Spyrou, M. A., Szécsényi-Nagy, A., Tõrv, M., Hansen, S., … Krause, J. (2017). The Stone Age Plague and Its Persistence in Eurasia. *Current Biology* , *27*(23), 3683-3691.e8. https://doi.org/10.1016/J.CUB.2017.10.025

Valtueña, A. A., Neumann, G. U., Spyrou, M. A., Musralina, L., Aron, F., Beisenov, A., Belinskiy, A. B., Bos, K. I., Buzhilova, A., Conrad, M., Djansugurova, L. B., Dobes, M., Ernee, M., Fernandez-Eraso, J., Frohlich, B., Furmanek, M., Hałuszko, A., Hansen, S., Harney, E., … Herbig, A. (2022). Stone Age Yersinia pestis genomes shed light on the early evolution, diversity, and ecology of plague. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(17), e2116722119. https://doi.org/https://www.pnas.org/doi/abs/10.1073/pnas.2116722119

Velsko, I.M., Overmyer, K.A., Speller, C. et al. The dental calculus metabolome in modern and historic samples. Metabolomics 13, 134 (2017). https://doi.org/10.1007/s11306-017-1270-3

Wagner, S., Lagane, F., Seguin-Orlando, A., Schubert, M., Leroy, T., Guichoux, E., Chancerel, E., Bech-Hebelstrup, I., Bernard, V., Billard, C., Billaud, Y., Bolliger, M., Croutsch, C., Čufar, K., Eynaud, F., Heussner, K. U., Köninger, J., Langenegger, F., Leroy, F., … Orlando, L. (2018). High-Throughput DNA sequencing of ancient wood. *Molecular Ecology*, *27*(5), 1138–1154. https://doi.org/10.1111/MEC.14514

Wang, Y., Pedersen, M. W., Alsos, I. G., De Sanctis, B., Racimo, F., Prohaska, A., Coissac, E., Owens, H. L., Merkel, M. K. F., Fernandez-Guerra, A., Rouillard, A., Lammers, Y., Alberti,

A., Denoeud, F., Money, D., Ruter, A. H., McColl, H., Larsen, N. K., Cherezova, A. A., … Willerslev, E. (2021). Late Quaternary dynamics of Arctic biota from ancient environmental genomics. *Nature* , *600*(7887), 86–92. https://doi.org/10.1038/s41586-021-04016-x

Warinner, C., Herbig, A., Mann, A., Fellows Yates, J. A., Weiß, C. L., Burbano, H. A., Orlando, L., & Krause, J. (2017). A Robust Framework for Microbial Archaeology. *Annual Review of Genomics and Human Genetics*, *18*, 321–356. https://doi.org/10.1146/annurev-genom-091416

Wibowo, M. C., Yang, Z., Borry, M., Hübner, A., Huang, K. D., Tierney, B. T., Zimmerman, S., Barajas-Olmos, F., Contreras-Cubas, C., García-Ortiz, H., Martínez-Hernández, A., Luber, J. M., Kirstahler, P., Blohm, T., Smiley, F. E., Arnold, R., Ballal, S. A., Pamp, S. J., Russ, J., … Kostic, A. D. (2021). Reconstruction of ancient microbial genomes from the human gut. *Nature*, *594*(7862), 234–239. https://doi.org/10.1038/s41586-021-03532-0

Worobey, M., Watts, T. D., McKay, R. A., Suchard, M. A., Granade, T., Teuwen, D. E., Koblin, B. A., Heneine, W., Lemey, P., & Jaffe, H. W. (2016). 1970s and 'Patient 0' HIV-1 genomes illuminate early HIV/AIDS history in North America. *Nature*, *539*(7627), 98–101. https://doi.org/10.1038/nature19827

Xie, S., & Kershaw, S. (2012). Microbes and paleoenvironments. *Geobiology*, *10*(1), 1–2. https://doi.org/10.1111/j.1472-4669.2011.00311.x

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., Thines, M., Weigel, D., & Burbano, H. A. (2013). The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. *ELife*, *2*(e00731). https://doi.org/10.7554/ELIFE.00731

Chapter 2. Herbaria: more than mere collections of historical plant specimens

My PhD thesis makes use of herbarium collections to study ancient plant-associated microbiomes and their evolution through time; this is one of the many innovative uses of natural history collections. Thus, what are herbaria, how are they organized, what do they contain, what are they used for and what are their contribution to todays science, what are their future? These are some of the arguments that I briefly synthesize in the current chapter.

The first botanical collections were established in Europe during the Renaissance period, driven by extensive plant gathering expeditions, and reached their modern form with Linnaeus' organization of specimens in the 18th century. (Meredith, 1996). Herbarium collections, as listed in the current (2024) version of the Index Herbariorum (https://sweetgum.nybg.org/science/ih/) represent a remarkable worldwide repository of global plant diversity. With over 396 million specimens preserved across 3567 herbaria worldwide, these collections represent an invaluable resource for botanical research and conservation efforts (Thiers, 2022, **Fig. 1-A**). The three largest herbaria are located at the Royal Botanical Gardens in London, housing 8.1 million specimens, the Muséum National d'Histoire Naturelle in Paris with 8.0 million, and the New York Botanical Garden housing 7.9 million specimens.

Over time, herbaria have been the subject of various types of research spanning science, but also art, or anthropology (**Fig.1-C**). In addition to their essential roles in activities related to plant taxonomy, systematics, and evolution, herbaria have been exploited in novel contexts in recent times (Davis, 2023). For instance, Park et al. (2023) have highlighted the colonial legacy of herbarium collections by demonstrating an inverse relationship between where plant biodiversity is highest in nature and where it is housed in herbaria. Specifically, herbaria in the United States and several nations in Western and Central Europe house over twice the number of species that occur in these nations, underscoring the international appropriation of large amounts of plant diversity. Other examples are connected to the work of evolutionary biologists who increasingly collaborate with anthropologists and archaeologists to sequence complete genomes of extant and herbarium plant specimens to shed light on the origins of domesticated species (Bieker & Martin, 2018; Jaenicke-Després et al., 2003; Kistler et al., 2020; Swarts et al., 2017; White et al., 2021). Herbaria also make it possible to study how plants have adapted to the environment and responded to climate change during the Anthropocene. Herbarium plants with known

collection places and dates indeed allow estimate changes over time in the distribution range and phenology of specific plant taxa (Macdougall et al., 1998; Marsico et al., 2020; Primack et al., 2004; Willis et al., 2017; Yost et al., 2018). Due to the urgency of the biodiversity crisis, large collaborative efforts are underway to synthesize data and enhance conservation planning. These efforts involve harmonizing occurrence data from herbarium specimens with global-change data to gain insights into past, present, and future biodiversity patterns, processes, and dynamics (Davis, 2023; Heberling et al., 2019; Lang et al., 2019; Lavoie, 2013; Meineke et al., 2018; Nualart et al., 2017). Using herbaria, it is possible to analyze various aspects of the global change crisis, from identifying drivers of diversification in specific areas, to assessing the effectiveness of protected areas in regions threatened by urbanization, or investigating human-mediated biological invasions (Daru et al., 2021; Ding et al., 2020; Mishler et al., 2020; Simkin et al., 2022; Willis et al., 2010).

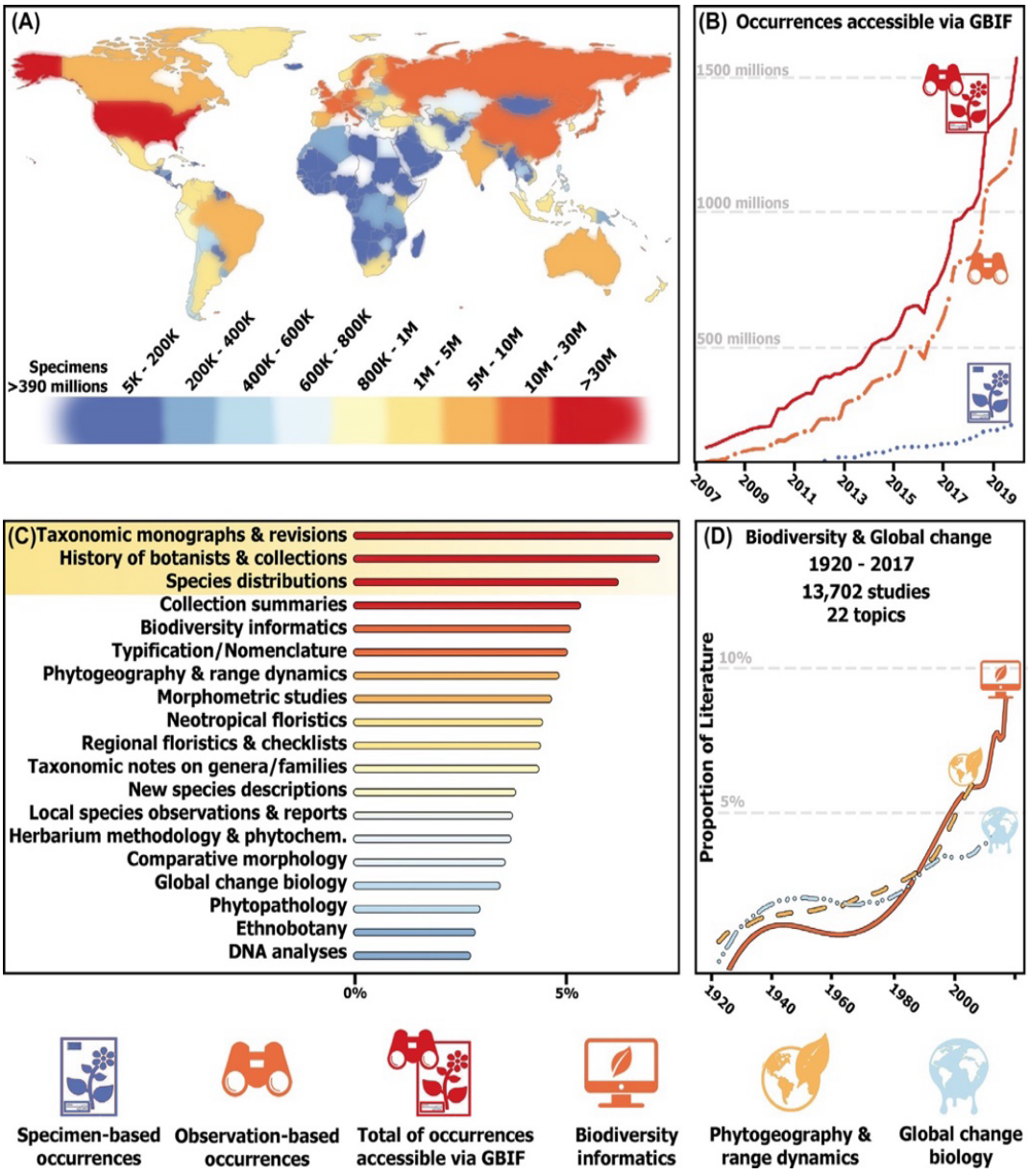


Figure 4. The global distribution and use of herbaria. (A) The global distribution of country-wide herbarium specimens. (B) The contrasting growth of digitized specimen versus observational data mobilized online by the Global Biodiversity Information Facility (GBIF). (C) Literature summary of traditional uses of herbarium specimens (shaded in yellow) versus the (D) expanding non-traditional uses of collections. Icon legend for (B) and (D) shown at bottom. Figure from David, (2023).

Herbarium Digitization and Specimen Accessibility

The widespread utilization of herbaria in research has been greatly facilitated by the digitization of collections that can then be consulted at distance through the web. This process can be divided into different steps: specimen imaging (digitization *per se*), and informatization of the associated data (e.g. species name, collection date, collection place [georeferencing], collector name, ...). Completion of these phases enables the creation of curated databases of herbarium specimens that are easily accessible and directly amenable to scientific research and computational analyses. However, despite numerous technological and methodological advances in recent years (Gutiérrez-Larruscain et al., 2018; Hedrick et al., 2020; Tegelberg et al., 2014; Yost et al., 2018), digitization remains a costly and time-consuming process. Currently, digitally accessible collections are still not common (**Fig. 1 B**). It is estimated that less than 30% of collections have at least location and collection date information available online, and less than 10% have digital images (**Fig. 2 A**, Park et al., 2023). Digitized specimen data are typically submitted to the Global Biodiversity Information Facility (GBIF, https://www.gbif.org/) and/or to regional databases, which often share data with GBIF (e.g., Consortium of California Herbaria, Australasian Virtual Herbarium, eReColNat, Virtual Herbaria - JACQ; **Fig. 2 E**). However, digitized specimens are not always easily accessible and represent only the tip of the iceberg compared to physical collections, thus inadequately describing the vast biodiversity preserved in herbarium collections. Researchers still often need to personally visit herbaria to search for specimens, requesting access and permission to sample from the collection's curators. For various types of analyses, such as DNA studies, small portions of the herbarium plants must be collected. This destructive process can make access to the collections more difficult and is often denied for rare specimens or types (specimens or groups of specimens formally associated with the scientific names of organisms). Furthermore, the part of the plant being studied influences the decision; portions essential for morphological determination of the species, such as leaves, flowers, and fruits, tend to be less accessible compared to the root system (if present), which is less interesting and informative. Nowadays, technological advancement has made it possible to increasingly reduce the quantity of material to be sampled and non-destructive DNA extraction methods have also been described (Shepherd, 2017).

From herbarium specimens to historical DNA

The emergence of high-throughput sequencing technologies has revolutionized the study of herbaria, unlocking new ways to exploring genome-level diversity of historical herbarium samples. Plant specimens in herbarium are known to contain traces of DNA. Herbarium-derived DNA can be defined as historical or ancient DNA (aDNA) by its biochemical characteristics; it is highly fragmented, with distinct patterns of DNA misincorporations. Identifying and evaluating these alterations is a crucial step in verifying the absence of contemporary contaminations within the historical DNA of herbaria. To analyze herbarium specimens and ancient samples, specific laboratories and protocols for aDNA purification, amplification, and sequence analysis are necessary (Brewer et al., 2019; Marinček et al., 2022; Semple et al., 2004; Sugita et al., 2020; Taylor & Swann, 1994). The fragment lengths obtained from herbarium samples are typically extremely short, ranging from 50 to 200 base pairs. This characteristic complicates data assembly and processing, necessitating the utilization of specialized bioinformatic tools or parameters distinct from those employed for modern DNA. Longer fragments exceeding 500 base pairs are often elusive or challenging to obtain from herbarium specimens (Papalini et al., 2023). Herbaria serve as invaluable resources for scientific research, including classification, macroevolutionary inference, and tracing the evolutionary history of wild plants. The use of herbarium aDNA represents a useful resource in large-scale phylogenomic. For instance, for the Kew Angiosperm Tree of life project, more than 28% of the taxa examined were sampled from herbarium. This strategy facilitated the examination of taxa that would have otherwise remained inaccessible within todays' research project timeframes (Baker et al., 2022). Moreover, herbarium collections provide spatiotemporal data crucial for understanding the genetic mechanisms underpinning plant responses to environmental changes (Heberling et al., 2019; Lang et al., 2019; Meineke et al., 2018). For example, they offer insights into the impacts of modern agricultural practices and the introduction of non-native species, that potentially evolve into invasive species (Van Kleunen et al., 2015). Traditionally, herbarium serve as proxies for identifying first-plant colonizers, shedding light on critical aspects of plant ecology and evolution (Bieker et al., 2022; Exposito-Alonso et al., 2018). Crop plant herbarium specimens play also a crucial role in agricultural research, offering valuable resources for studying crop evolution, domestication, movements, adaptations, and changes in genome content (Gutaker et al., 2019;

Kreiner et al., 2022; Mascher et al., 2016; van de Wouw et al., 2010; Vos et al., 2022).

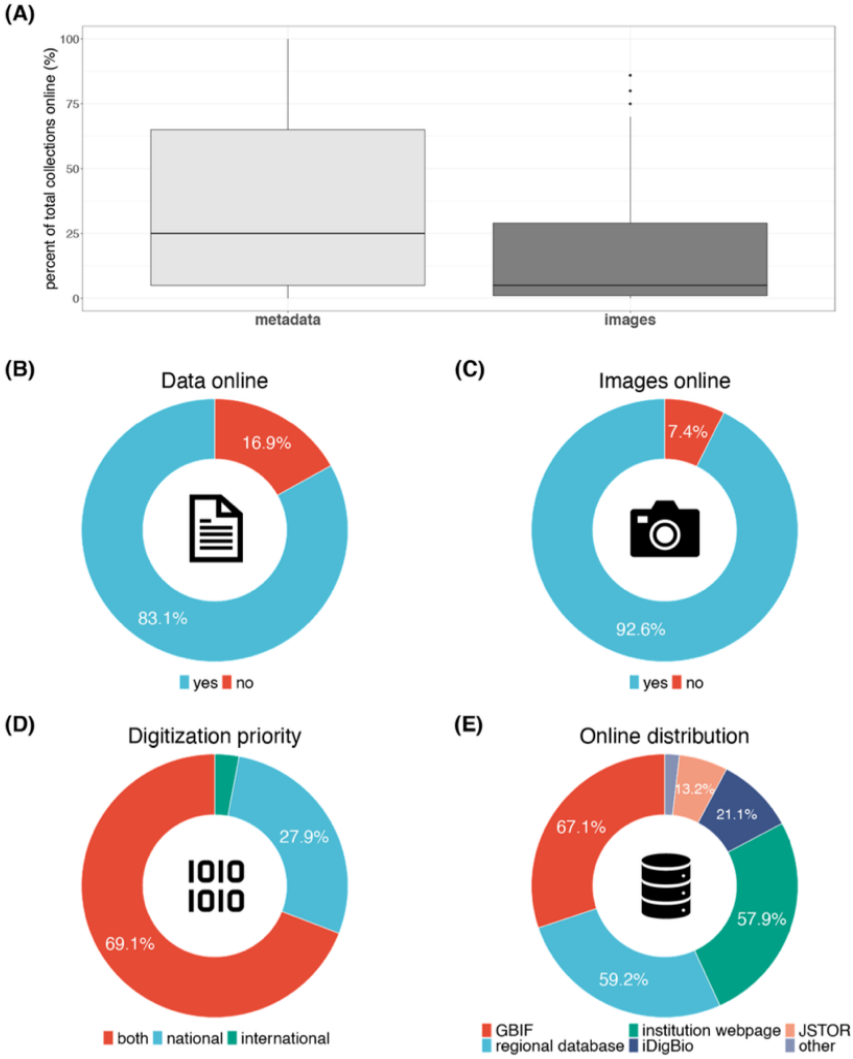


Figure 2. Trends in the digitization of herbarium specimens. Boxplot (A) summarizes the percentage of physical collections in surveyed herbaria that have at least location and date information (metadata) or digital image online. Pie charts (B) and (C) show the percentage of herbaria that have at least some data and images of their collections shared online, respectively. The digitization priority of herbaria in terms of where specimens were collected is illustrated in pie chart (D), while pie chart (E) shows how the surveyed herbaria share and distribute digital specimen data. Figure from Park et a. (2023).

Herbaria and plant-associated microorganisms

Traces of microorganisms can be identified on herbarium samples. In the case of pathogens (fungal, bacterial or even viral taxa), typical disease-related symptoms can be identified on plant organs (usually on the leaves and stem of the plant). For example, as shown in **Fig. 3 A-B**, rusts fungal pathogens create identifiable symptoms such as those observed on leaves of *Sorbus aucuparia* (due to the fungus *Gymnosporangium cornutum*) and *Buxus sempervirens*. Similarly, other more common symptoms of plant pathogens such as yellowing, necrosis, staining and browning can also be identified on herbarium plants. The sampling of small portions of diseased organs also makes it possible through the study of ancient DNA to molecularly identify the pathogen that affected the plant when it was alive. Using this approach it has been possible to trace the historical spread and evolution of several plant pathogens over the past centuries as in the case of *Phytophthora infestans*, the causal agent of potato late blight (Yoshida et al., 2013). or *Xanthomonas citri* pv. *citri*, the causal agent of Asiatic citrus canker (Campos et al., 2021). As far as viruses are concerned, it was also possible to identify their presence through the study of DNA or even small viral RNAs preserved in the herbarium plant tissues (Gagnevin et al., 2021). Thus, using small RNA sequences, Rieux and colleagues (2021) were able to reconstruct the near-complete sequence of Cassava geminiviruses from 50- and 90-year-old herbarium specimens.

In addition to pathogens, traces of plant-associated symbiotic microorganisms can also be found in herbaria. For example, in the roots of many herbarium plant genera belonging to the Fabaceae, it is possible to identify nodule-forming rhizobia (**Fig. 3 A-C**) and, through microscopic observations, the presence of endophytic and symbiotic fungi (**Fig. 3 D-D3**, Bianciotto et al., 2022). In this case as well, molecular analysis is a powerful tool for investigating in greater depth which organisms have colonized the plant in the past. In a recent study, Heberling et al. (2019) showed that it was possible to amplify and sequence DNA of arbuscular mycorrhizal fungi from herbarium root DNA and highlighted significant differences in AMF community composition between different plant species.

Finally, within the herbarium samples it is sometimes possible to find the presence of dried rhizospheric soils, which surround their roots. They may hypothetically preserve the original rhizospheric microbiomes (**Fig. 3 E**, Bianciotto et al., 2022). In the case of herbaria, a very limited number of studies

have targeted their entire complex, ‘asymptomatic’, microbiomes using DNA extracted from either the above-ground (leaves, stems) and underground (roots) organs of herbaria (Bieker et al., 2020, Schubert et al., 2014).

Herbaria were mostly created in the 19th and first half of the 20th century, the period that preceded the current acceleration of global changes whose impact on the soil microbiome is largely unknown. The analysis of microbial aDNA extracted from the soil and roots of herbarium plants should thus inform us on how and to what extent global changes have affected the plant-associated microbiomes not only from a taxonomic, but also a functional point of view.

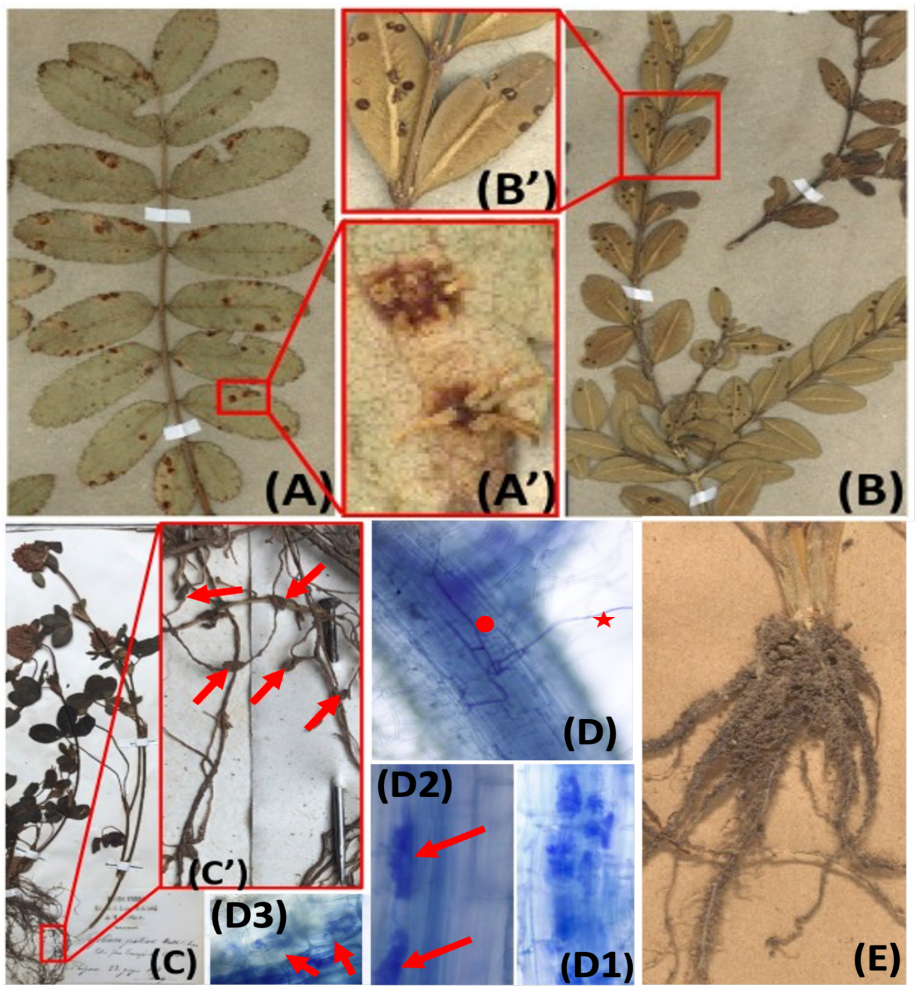


Figure 3. (A–B) symptoms induced by fungal rust pathogens on leaves of (A, A') *Sorbus aucuparia* (fungus: *Gymnosporangium cornutum*) and (B, B') *Buxus sempervirens* (fungus: *Puccinia buxi*); (C–C') nodules induced by N_2-fixing rhizobia on the roots of *Trifolium pratense* collected in 1879; (D–D3) Microscopic observation of cotton blue-stained arbuscular mycorrhizal (AM) fungi in the roots of *S. cereale* plants collected in 1905; arrows, collapsed arbuscules in cortical cells; star, external hypha of an AM fungus reaching and growing on the surface of root epidermal cells (circle); (E) microorganisms-containing dried rhizospheric soil sheathing the roots of a *Secale cereale* plant collected in 1905. Photo credits; (A–B and D), ©Herbier LY, FR-CERESE, UCBLyon1; (C, E) ©TO Herbarium, department of Life Sciences and Systems Biology, University of Turin; (D), Valeria Bianciotto, Gianluca Grasso. Full pictures of plants illustrated in D and E–G are visible on the Recolnat portal (https://explore.recolnat.org/) under accessions LY0662690 (D) and LY0662689 (E–G). Image modified from Bianciotto et al, 2022.

References

Baker, W. J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L. R., Brewer, G., Carruthers, T., Clarkson, J. J., Cook, J., Cowan, R. S., Dodsworth, S., Epitawalage, N., Françoso, E., Gallego, B., Johnson, M. G., Kim, J. T., Leempoel, K., … Forest, F. (2022). A Comprehensive Phylogenomic Platform for Exploring the Angiosperm Tree of Life. *Systematic Biology*, *71*(2), 301–319. https://doi.org/10.1093/SYSBIO/SYAB035

Bianciotto, V., Selosse, M. A., Martos, F., & Marmeisse, R. (2022). Herbaria preserve plant microbiota responses to environmental changes. *Trends in Plant Science*, *27*(2), 120–123. https://doi.org/10.1016/J.TPLANTS.2021.11.012

Bieker, V. C., Battlay, P., Petersen, B., Sun, X., Wilson, J., Brealey, J. C., Bretagnolle, F., Nurkowski, K., Lee, C., Barreiro, F. S., Owens, G. L., Lee, J. Y., Kellner, F. L., van Boheeman, L., Gopalakrishnan, S., Gaudeul, M., Mueller-Schaerer, H., Lommen, S., Karrer, G., … Martin, M. D. (2022). Uncovering the genomic basis of an extraordinary plant invasion. *Science Advances*, *8*(34), 19. https://doi.org/10.1126/

Bieker, V. C., Barreiro, F. S., Rasmussen, J. A., Brunier, M., Wales, N., & Martin, M. D. (2020). Metagenomic analysis of historical herbarium specimens reveals a postmortem microbial community. *Molecular Ecology Resources*, *20*(5), 1206–1219. https://doi.org/10.1111/1755-0998.13174

Bieker, V. C., & Martin, M. D. (2018). Implications and future prospects for evolutionary analyses of DNA in historical herbarium collections. *Botany Letters*, *165*(3–4), 409–418. https://doi.org/10.1080/23818107.2018.1458651

Brewer, G. E., Clarkson, J. J., Maurin, O., Zuntini, A. R., Barber, V., Bellot, S., Biggs, N., Cowan, R. S., Davies, N. M. J., Dodsworth, S., Edwards, S. L., Eiserhardt, W. L., Epitawalage, N., Frisby, S., Grall, A., Kersey, P. J., Pokorny, L., Leitch, I. J., Forest, F., & Baker, W. J. (2019). Factors Affecting Targeted Sequencing of 353 Nuclear Genes From Herbarium Specimens Spanning the Diversity of Angiosperms. *Frontiers in Plant Science*, *10*, 477275. https://doi.org/10.3389/FPLS.2019.01102/BIBTEX

Burbano, H. A., & Gutaker, R. M. (2023). Ancient DNA genomics and the renaissance of herbaria. *Science*, *382*(6666), 59–63. https://doi.org/10.1126/SCIENCE.ADI1180/ASSET/DA5C6B1D-632B-448E-B439-BC13363F0094/ASSETS/IMAGES/LARGE/SCIENCE.ADI1180-F2.JPG

Campos, P. E., Crego, C. G., Boyer, K., Gaudeul, M., Baider, C., Richard, D., Pruvost, O., Roumagnac, P., Szurek, B., Becker, N., Gagnevin, L., & Rieux, A. (2021). First historical genome of a crop bacterial pathogen from herbarium specimen: Insights into citrus canker emergence. *PLOS Pathogens*, *17*(7), e1009714. https://doi.org/10.1371/JOURNAL.PPAT.1009714

Chialva, M., & Bonfante, P. (2018). Plant microbiota: From model plants to Mediterranean crops. *Phytopathologia Mediterranea*, *57*(1). https://doi.org/10.14601/Phytopathol_Mediterr-22052

Daru, B. H., Davies, T. J., Willis, C. G., Meineke, E. K., Ronk, A., Zobel, M., Pärtel, M., Antonelli, A., & Davis, C. C. (2021). Widespread homogenization of plant communities in the Anthropocene. *Nature Communications 2021 12:1*, *12*(1), 1–10. https://doi.org/10.1038/s41467-021-27186-8

Davis, C. C. (2023). The herbarium of the future. *Trends in Ecology and Evolution*, *38*(5), 412–423. https://doi.org/10.1016/j.tree.2022.11.015

Ding, W. N., Ree, R. H., Spicer, R. A., & Xing, Y. W. (2020). Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science*, *369*(6503), 578–581. https://doi.org/10.1126/SCIENCE.ABB4484

Exposito-Alonso, M., Becker, C., Schuenemann, V. J., Reiter, E., Setzer, C., Slovak, R., Brachi, B., Hagmann, J., Grimm, D. G., Chen, J., Busch, W., Bergelson, J., Ness, R. W., Krause, J.,

Burbano, H. A., & Weigel, D. (2018). The rate and potential relevance of new mutations in a colonizing plant lineage. *PLOS Genetics*, *14*(2), e1007155. https://doi.org/10.1371/JOURNAL.PGEN.1007155

Gagnevin, L., Rieux, A., Lett, J., Roumagnac, P., Szurek, B., Campos, P., Baider, C., Gaudeul, M., & Becker, N. (2021). Herbaria, a Window into the Evolutionary History of Crop Pathogens. *Natural History Collections in the Science of the 21st Century*, 195–218. https://doi.org/10.1002/9781119882237.ch13

Gutaker, R. M., Weiß, C. L., Ellis, D., Anglin, N. L., Knapp, S., Luis Fernández-Alonso, J., Prat, S., & Burbano, H. A. (2019). The origins and adaptation of European potatoes reconstructed from historical genomes. *Nature Ecology & Evolution 2019 3:7*, *3*(7), 1093–1101. https://doi.org/10.1038/s41559-019-0921-3

Gutiérrez-Larruscain, D., Santos-Vicente, M., Anderberg, A. A., Rico, E., & Martínez-Ortega, M. M. (2018). Large–scale digitization of herbarium specimens: Development and usage of an automated, high–throughput conveyor system. *TAXON*, *67*(1), 165–178. https://doi.org/10.12705/671.10

Heberling, J. M., Prather, L. A., & Tonsor, S. J. (2019). *The Changing Uses of Herbarium Data in an Era of Global Change: An Overview Using Automated Content Analysis*. *69*(10). https://doi.org/10.1093/biosci/biz094

Hedrick, B. P., Heberling, J. M., Meineke, E. K., Turner, K. G., Grassa, C. J., Park, D. S., Kennedy, J., Clarke, J. A., Cook, J. A., Blackburn, D. C., Edwards, S. V, & Davis, C. C. (2020). Digitization and the Future of Natural History Collections. *BioScience*, *70*(3), 243–251. https://doi.org/10.1093/biosci/biz163

Jaenicke-Després, V., Buckler, E. S., Smith, B. D., Gilbert, M. T. P., Cooper, A., Doebley, J., & Pääbo, S. (2003). Early Allelic Selection in Maize as Revealed by Ancient DNA. *Science*, *302*(5648), 1206–1208. https://doi.org/10.1126/SCIENCE.1089056

Kistler, L., Bieker, V. C., Martin, M. D., Pedersen, M. W., Ramos Madrigal, J., & Wales, N. (2020). Ancient Plant Genomics in Archaeology, Herbaria, and the Environment. *Annual Review of Plant Biology*, *71*(Volume 71, 2020), 605–629. https://doi.org/10.1146/ANNUREV-ARPLANT-081519-035837/CITE/REFWORKS

Kreiner, J. M., Latorre, S. M., Burbano, H. A., Stinchcombe, J. R., Otto, S. P., Weigel, D., & Wright, S. I. (2022). Rapid weed adaptation and range expansion in response to agriculture over the past two centuries. *Science*, *378*(6624), 1079–1085. https://doi.org/10.1126/SCIENCE.ABO7293/SUPPL_FILE/SCIENCE.ABO7293_DATA_S1_AND_S2.ZIP

Lang, P. L. M., Willems, F. M., Scheepens, J. F., Burbano, H. A., & Bossdorf, O. (2019). Using herbaria to study global environmental change. *New Phytologist*, *221*(1), 110–122. https://doi.org/10.1111/NPH.15401

Lavoie, C. (2013). Biological collections in an ever changing world: Herbaria as tools for biogeographical and environmental studies. *Perspectives in Plant Ecology, Evolution and Systematics*, *15*(1), 68–76. https://doi.org/10.1016/J.PPEES.2012.10.002

Macdougall, A. S., Loo, J. A., Clayden, S. R., Goltz, J. G., & Hinds, H. R. (1998). Defning conservation priorities for plant taxa in southeastern New Brunswick, Canada using herbarium records. *Biological Conservation*, *86*, 325–338.

Meredith, A. L. 1996. Roles of natural history collections. Annals of the Missouri Botanical Garden 83: 536–545.

Marinček, P., Wagner, N. D., & Tomasello, S. (2022). Ancient DNA extraction methods for herbarium specimens: When is it worth the effort? *Applications in Plant Sciences*, *10*(3), e11477. https://doi.org/10.1002/APS3.11477

Marsico, T. D., Krimmel, E. R., Carter, J. R., Gillespie, E. L., Lowe, P. D., McCauley, R., Morris, A. B., Nelson, G., Smith, M., Soteropoulos, D. L., & Monfils, A. K. (2020). Small herbaria

contribute unique biogeographic records to county, locality, and temporal scales. *American Journal of Botany*, *107*(11), 1577–1587. https://doi.org/10.1002/AJB2.1563

Mascher, M., Schuenemann, V. J., Davidovich, U., Marom, N., Himmelbach, A., Hübner, S., Korol, A., David, M., Reiter, E., Riehl, S., Schreiber, M., Vohr, S. H., Green, R. E., Dawson, I. K., Russell, J., Kilian, B., Muehlbauer, G. J., Waugh, R., Fahima, T., … Stein, N. (2016). Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nature Genetics 2016 48:9*, *48*(9), 1089–1093. https://doi.org/10.1038/ng.3611

Meineke, E. K., Davis, C. C., & Davies, T. J. (2018). The unrealized potential of herbaria for global change biology. *Ecological Monographs*, *88*(4), 505–525. https://doi.org/10.1002/ECM.1307

Mishler, B. D., Guralnick, R., Soltis, P. S., Smith, S. A., Soltis, D. E., Barve, N., Allen, J. M., & Laffan, S. W. (2020). Spatial phylogenetics of the North American flora. *Journal of Systematics and Evolution*, *58*(4), 393–405. https://doi.org/10.1111/JSE.12590

Nualart, N., Neus Ibáñez, &, Soriano, I., & López-Pujol, J. (2017). Assessing the Relevance of Herbarium Collections as Tools for Conservation Biology. *Bot. Rev.*, *83*, 303–325. https://doi.org/10.1007/s12229-017-9188-z

Papalini, S., Di Vittori, V., Pieri, A., Allegrezza, M., Frascarelli, G., Nanni, L., Bitocchi, E., Bellucci, E., Gioia, T., Pereira, L. G., Susek, K., Tenaillon, M., Neumann, K., & Papa, R. (2023). Challenges and Opportunities behind the Use of Herbaria in Paleogenomics Studies. *Plants 2023, Vol. 12, Page 3452*, *12*(19), 3452. https://doi.org/10.3390/PLANTS12193452

Park, D. S., Feng, X., Akiyama, S., Ardiyani, M., Avendaño, N., Barina, Z., Bärtschi, B., Belgrano, M., Betancur, J., Bijmoer, R., Bogaerts, A., Cano, A., Danihelka, J., Garg, A., Giblin, D. E., Gogoi, R., Guggisberg, A., Hyvärinen, M., James, S. A., … Davis, C. C. (2023). The colonial legacy of herbaria. *Nature Human Behaviour 2023 7:7*, *7*(7), 1059–1068. https://doi.org/10.1038/s41562-023-01616-7

Primack, D., Imbres, C., Primack, R. B., Miller-Rushing, A. J., & Del Tredici, P. (2004). Herbarium specimens demonstrate earlier flowering times in response to warming in Boston. *American Journal of Botany*, *91*(8), 1260–1264. https://doi.org/10.3732/AJB.91.8.1260

Rieux, A., Campos, P., Duvermy, A. *et al.* Contribution of historical herbarium small RNAs to the reconstruction of a cassava mosaic geminivirus evolutionary history. *Sci Rep* 11, 21280 (2021). https://doi.org/10.1038/s41598-021-00518-w

Semple, J. C., Mcminn-Sauder, | Harper, Stover, | Melissa, Lemmon, A. R., Lemmon, E. M., James, |, Beck, B., & Beck, J. B. (2004). Goldenrod herbariomics: Hybrid-sequence capture reveals the phylogeny of Solidago. *Am J Bot*, *110*. https://doi.org/10.1002/ajb2.16164

Simkin, R. D., Seto, K. C., McDonald, R. I., & Jetz, W. (2022). Biodiversity impacts and conservation implications of urban land expansion projected to 2050. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(12), e2117297119. https://doi.org/10.1073/PNAS.2117297119/SUPPL_FILE/PNAS.2117297119.SD02.XLSX

Shepherd LD (2017) A non-destructive DNA sampling technique for herbarium specimens. PLoS ONE 12(8): e0183555. https://doi.org/10.1371/journal.pone.018355

Schubert, M., Ermini, L., Sarkissian, C. et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat Protoc 9, 1056–1082 (2014). https://doi.org/10.1038/nprot.2014.063

Sugita, N., Ebihara, A., Hosoya, T., Jinbo, U., Kaneko, S., Kurosawa, T., Nakae, M., & Yukawa, T. (2020). Non-destructive DNA extraction from herbarium specimens: a method particularly suitable for plants with small and fragile leaves. *Journal of Plant Research*, *133*(1), 133–141. https://doi.org/10.1007/S10265-019-01152-4/FIGURES/2

Swarts, K., Gutaker, R. M., Benz, B., Blake, M., Bukowski, R., Holland, J., Kruse-Peeples, M., Lepak, N., Prim, L., Romay, M. C., Ross-Ibarra, J., Sanchez-Gonzalez, J. D. J., Schmidt, C., Schuenemann, V. J., Krause, J., Matson, R. G., Weigel, D., Buckler, E. S., & Burbano, H. A.

(2017). Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science*, *357*(6350), 512–515. https://doi.org/10.1126/SCIENCE.AAM9425

Taylor, J. W., & Swann, E. C. (1994). DNA from Herbarium Specimens. *Ancient DNA*, 166–181. https://doi.org/10.1007/978-1-4612-4318-2_11

Tegelberg, R., Mononen, T., & Saarenmaa, H. (2014). High-Performance digitization of natural history collections: Automated imaging lines for herbarium and insect specimens. *Taxon*, *63*(6), 1307–1313. https://doi.org/10.12705/636.13

Thiers, B. M. (2022). The World's Herbaria 2022: A Summary Report Based on Data from Index Herbariorum. *Index Herbariorum*. http://sweetgum.nybg.org/science/ih/

van de Wouw, M., van Hintum, T., Kik, C., van Treuren, R., & Visser, B. (2010). Genetic diversity trends in twentieth century crop cultivars: A meta analysis. *Theoretical and Applied Genetics*, *120*(6), 1241–1252. https://doi.org/10.1007/S00122-009-1252-6/FIGURES/3

Van Kleunen, M., Dawson, W., Essl, F., Pergl, J., Winter, M., Weber, E., Kreft, H., Weigelt, P., Kartesz, J., Nishino, M., Antonova, L. A., Barcelona, J. F., Cabezas, F. J., Cárdenas, D., Cárdenas-Toro, J., Castaño, N., Chacón, E., Chatelain, C., Ebel, A. L., … Pyšek, P. (2015). Global exchange and accumulation of non-native plants. *Nature*, *525*(7567), 100–103. https://doi.org/10.1038/NATURE14910

Vos, R., Andel, T. van, Gravendeel, B., Kakakiou, V., Michels, E., Schijlen, E., & Stefanaki, A. (2022). A Tomato Genome From The Italian Renaissance Provides Insights Into Columbian and Pre-Columbian Exchange Links And Domestication. *BioRxiv*, 2022.06.07.495125. https://doi.org/10.1101/2022.06.07.495125

White, D. M., Huang, J.-P., Jara-Muñoz, O. A., Madriñán, S., Ree, R. H., & Mason-Gamer, R. J. (2021). The Origins of Coca: Museum Genomics Reveals Multiple Independent Domestications from Progenitor Erythroxylum gracilipes. *Syst. Biol*, *70*(1), 1–13. https://doi.org/10.1093/sysbio/syaa074

Willis, C. G., Ellwood, E. R., Primack, R. B., Davis, C. C., Pearson, K. D., Gallinat, A. S., Yost, J. M., Nelson, G., Mazer, S. J., Rossington, N. L., Sparks, T. H., & Soltis, P. S. (2017). Old Plants, New Tricks: Phenological Research Using Herbarium Specimens. *Trends in Ecology and Evolution*, *32*(7), 531–546. https://doi.org/10.1016/j.tree.2017.03.015

Willis, C. G., Ruhfel, B. R., Primack, R. B., Miller-Rushing, A. J., Losos, J. B., & Davis, C. C. (2010). Favorable Climate Change Response Explains Non-Native Species' Success in Thoreau's Woods. *PLOS ONE*, *5*(1), e8878. https://doi.org/10.1371/JOURNAL.PONE.0008878

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., Thines, M., Weigel, D., & Burbano, H. A. (2013). The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. *ELife*, *2*(2), 731. https://doi.org/10.7554/ELIFE.00731

Yost, J. M., Sweeney, P. W., Gilbert, E., Nelson, G., Guralnick, R., Gallinat, A. S., Ellwood, E. R., Rossington, N., Willis, C. G., Blum, S. D., Walls, R. L., Haston, E. M., Denslow, M. W., Zohner, C. M., Morris, A. B., Stucky, B. J., Carter, J. R., Baxter, D. G., Bolmgren, K., … Mazer, S. J. (2018). Digitization protocol for scoring reproductive phenology from herbarium specimens of seed plants. *Applications in Plant Sciences*, *6*(2), 1022. https://doi.org/10.1002/APS3.1022

Chapter 3. Optimization of protocols and methods for the analysis of microbial and viral communities in herbaria

The study of microbial communities using herbarium specimens requires specific methodologies and protocols to handle small amounts of ancient material, to extract highly degraded fragmented DNA while minimizing external and modern contamination. The workflow typically followed for these studies involves several key steps (**Fig. 1**).

The first step involves the identification of specimens within herbarium collections. It necessitates an assessment of the preservation conditions of the specimens and their relevance to the study. Searching for the right specimens within herbaria, is facilitated by the use of digital databases like e-Recolnat for the herbaria in France (https://www.recolnat.org/fr/) that allow consulting herbaria at distance through the internet. Once selected, the specimens are sampled to obtain the necessary quantities of material for DNA extraction.

DNA extraction is a crucial step and need to be performed in specialized laboratories known as "white laboratories." These laboratories are designed to minimize contamination with modern DNA. Key features include positive air pressure, daily UV/bleach treatments to sterilize surfaces and air from biological contaminants, restricted access limited to authorized and adequately trained personnel, and mandatory personal protective equipment, such as full-body overalls, masks, and double gloves for all operators.

After extraction, the DNA samples are converted into libraries ready for sequencing. This phase requires advanced molecular biology techniques to obtain enough DNA from highly degraded fragments.

The sequencing data is analyzed using specific bioinformatic tools. These tools are optimized for very short sequences typical of degraded DNA and for validation of historical authenticity to distinguish between ancient and modern DNA. Bioinformatic analysis allows for the study of the structure and composition of microbial communities present in herbarium specimens. Using

specific tools and databases, it is possible to identify microbial taxa and trace their evolution over time.

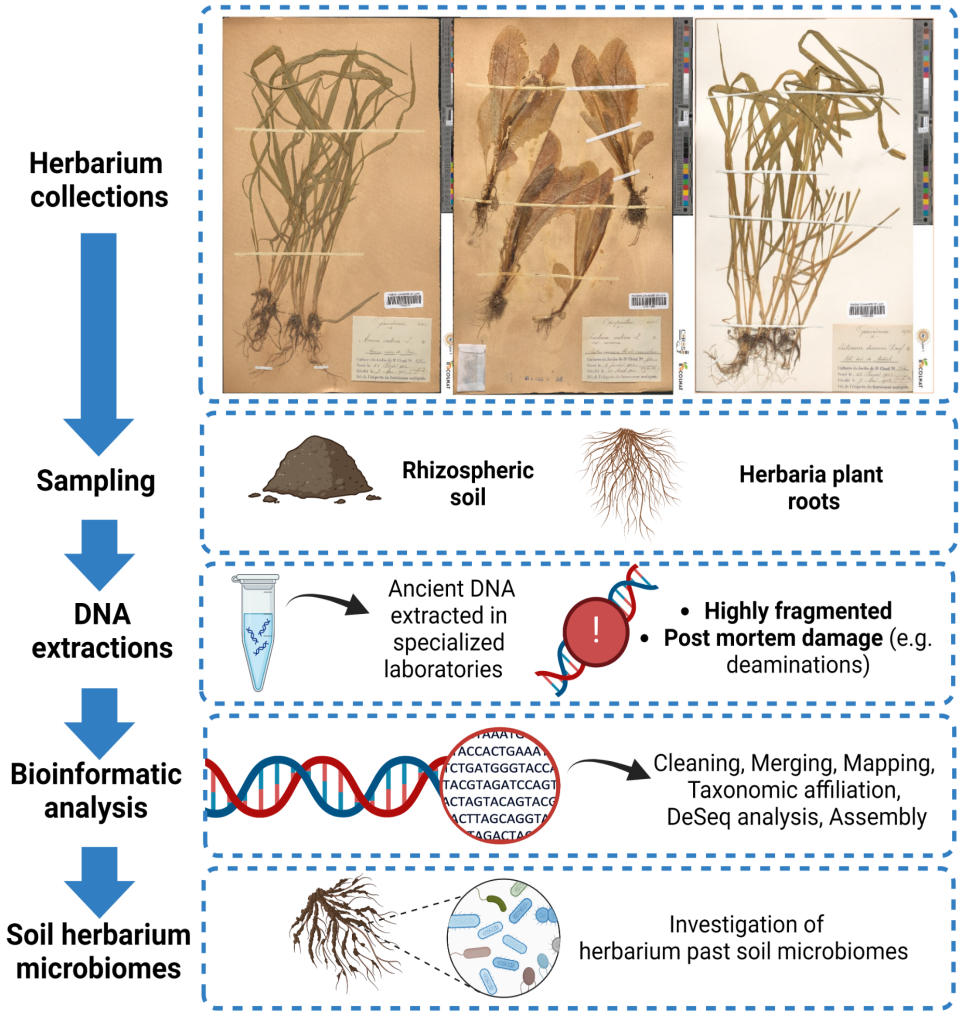


Figure 1. General workflow of the study of herbarium soil and root microbiomes.

This chapter that details these protocols corresponds to a recently (2023) published chapter in the book Viral Metagenomics from the book series Methods in Molecular Biology (https://doi.org/10.1007/978-1-0716-3515-5_15). The chapter presents the different methods and protocols used in the analysis of herbarium samples of soil and roots and has made it possible to discriminate the presence of viruses by obtaining their whole genome sequence from herbarium samples. These protocols and methods are those that are implemented in Chapter 4 to extract soils from a large variety of herbarium soils collected over the entire French territory. They are also implemented in Chapter 5 to analyze in detail the soil and root microbiomes associated to specimens of four cultivated plants.

Chapter 15

Identification of DNA Viruses in Ancient DNA from Herbarium Samples

Gianluca Grasso, Silvia Rotunno, Régis Debruyne, Lucie Bittner, Laura Miozzi, Roland Marmeisse, and Valeria Bianciotto

Abstract

Herbaria encompass millions of plant specimens, mostly collected in the nineteenth and twentieth centuries that can represent a key resource for investigating the history and evolution of phytopathogens. In the last years, the application of high-throughput sequencing technologies for the analysis of ancient nucleic acids has revolutionized the study of ancient pathogens including viruses, allowing the reconstruction of historical genomic viral sequences, improving phylogenetic based molecular dating, and providing essential insight into plant virus ecology. In this chapter, we describe a protocol to reconstruct ancient plant and soil viral sequences starting from highly fragmented ancient DNA extracted from herbarium plants and their associated rhizospheric soil. Following Illumina high-throughput sequencing, sequence data are de novo assembled, and DNA viral sequences are selected, according to their similarity with known viruses.

Key words Ancient DNA, DNA viruses, Herbaria, Museomics

Introduction

Plants host numerous viruses belonging to different families, causing, for several of them, important threats to crop plants. Since the causal association between specific disease symptoms and viruses is rather recent in the history of phytopathology, little is known regarding the origin, spread and prevalence of these pathogens in a recent past.

Natural history collections, and more specifically herbaria encompass millions of plant specimens, mostly collected in the 19th and 20th centuries, that can be searched for the presence and prevalence of specific pathogens and pests. Straightforward visual observation of herbarium plants have been carried out in the case of pathogens producing unambiguous symptoms as in the case of anther smut fungi infecting *Silene* flowers (Antonovics et al., 2003) or of the Horse-chestnut leaf miner (Lees et al., 2011). Alternatively, plant pathogens can be identified among degraded (ancient) aDNA molecules extracted from infected collection specimens. This approach has been reported for fungal (e.g. [Bradshaw et al., 2020) or Oomycete (Yoshida et al., 2013) eukaryotic pathogens and has led to the reconstruction of an entire historical genomic sequence of the Citrus canker bacterial agent (Campos et al., 2021).

A similar approach has been reported to investigate DNA virus evolution and prevalence in past human populations. For example, the systematic search for Hepatitis B virus sequences in ancient human DNA sequences has highlighted the temporal and spatial dispersal of different virus genotypes during the last ca 10,000 years (Kocher et al., 2021). Alternatively, the systematic search for any known pathogen sequences in the DNA extracted from 5th-8th century human remains from a common settlement in Germany identified several cases of co-infection by hepatitis B, smallpox viruses and Parvovirus B, highlighting the poor health status of this local human population (Bonczarowska et al., 2021).

In the case of plants, small RNA extracted from 90-year-old Cassava herbarium specimens allowed identification of a Cassava Mosaic Geminivirus sequence that was used to estimate its evolutionary rate (Rieux et al., 2021), showing how analyses of ancient viral genomic sequence data obtained from historical samples can substantially improve phylogenetic based molecular dating studies. Historical data on viral sequences can also provide essential insight into plant virus ecology, as demonstrated by the analysis of Barley yellow dwarf luteovirus sequences obtained from RNA extracted from herbarium specimens dating from

the end of 19th century to the first half of the 20th century (Malmstrom et al., 2007).

In this chapter, we describe a protocol for the extraction of DNA from herbarium plants and their associated rhizospheric soil. Following Illumina high throughput sequencing of the highly fragmented ancient DNA and its *de novo* assembly, associated DNA viral sequences are identified allowing to reconstruct the corresponding ancient plant and soil viromes.

Materials

Herbarium plant and soil sampling

1. 2 ml sterile centrifuge tubes
2. tweezers
3. sterile razor blades

DNA extraction

All experiments are performed in dedicated spaces within a cleanroom with positive pressure to prevent contamination from the outside environment, wearing laboratory coveralls and with specific cleaning procedures of equipment and spaces (with 2.6% bleach and/or 20 minutes UV-crosslinking at 256 nm).

1. High Pure Viral Large Volume extraction kit (Roche Diagnostics).
2. DNEasy Powersoil Pro kit (QIAGEN).
3. DNA low-binding 1.5 and 2 mL microtubes.
4. Sediment Lysis Buffer (SLB_{conc}): 20 mM Tris-HCl pH 9.0, 10 mM Calcium chloride, 100 mM DTT, 0.5% w/v SDS, 6.25% w/v Polyvinylpyrrolidone.
5. 20 mg/mL Proteinase k solution.
6. Binding Buffer (BB): 5 M Guanidinium HCl, 120 mM Sodium acetate pH 5.2, 0.05% v/v Tween-20, 40% v/v isopropanol (*see* **Note 1**).
7. Elution buffer (EBT): EB buffer (from QIAGEN DNEasy Powersoil Pro kit) supplemented with 0.05% v/v Tween-20.
8. Molecular Biology Grade ethanol.
9. Nuclease-Free water.
10. Qubit dsDNA High Sensitivity kit (Invitrogen)
11. Mixer Mill (with microtube accessories).
12. Hybridization oven or other heating device.
13. Rotary mixer/shaker.
14. Large volume refrigerated centrifuge (up to at least 3,000xg).
15. Microcentrifuge (up to at least 15,000xg)
16. Fluorometer (Invitrogen Qubit or equivalent)
17. 15 mL Falcon tubes

Library preparation and sequencing

1. Oligonucleotides IS1, IS2 and IS3 to make-up truncated P5 and P7 adapters (Meyer & Kircher, 2010)

2. Custom (7 nucleotides) indexed Illumina P5 and P7 Primers (Kircher et al., 2012)
3. NEBNext End-Repair module (New England Biolabs)
4. NEBNext Quick Ligation Module (New England Biolabs)
5. Bst polymerase large fragment (New England Biolabs)
6. Minelute PCR purification kit (QIAGEN)
7. PB_{acidic}: PB buffer (from minelute kit) supplemented with 60 mM Sodium Acetate pH 5.3
8. SsoAdvanced Universal SYBR supermix (Bio-Rad)
9. Qubit dsDNA High Sensitivity kit (Invitrogen)
10. NucleoMag cleanup and size selection beads kit (Macherey-Nagel)
11. Magnet for Solid Reversible Phase Immobilization (SPRI) purification using 1.5 mL microtubes (Invitrogen Dynamag or equivalent)
12. Microcentrifuge (up to at least 15,000xg)
13. Incubation block with heated lid (Eppendorf thermomixer or equivalent)
14. Real-Time PCR thermocycler
15. Fluorometer (Invitrogen Qubit or equivalent)
16. Device for capillary electrophoresis of DNA fragments (Labchip HT Perkin-Elmer with dedicated consumables/reagents or equivalent)
17. Illumina sequencer and dedicated consumables and chemistry

Bioinformatics analyses

1. fastQC(v.0.12.0) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).
2. fastp (v. 0.23.1) (https://github.com/OpenGene/fastp).
3. BWA (v. 0.7.17) (https://bio-bwa.sourceforge.net/).
4. samtools (v. 1.14) (http://www.htslib.org/).
5. SPAdes (v. 3.15.1) (https://github.com/ablab/spades).
6. QUAST (v. 5.2.0) (https://quast.sourceforge.net/).
7. CAP3 (VersionDate: 02/10/15) (https://doua.prabi.fr/software/cap3).
8. DIAMOND (v. 2.0.15) (https://github.com/bbuchfink/diamond).
9. MEGAN6 Community Edition (v. 6.24.23) (https://software-ab.cs.uni-tuebingen.de/download/megan6/welcome.html).

Methods

Herbarium plant and soil sampling

The protocol was successfully implemented on DNA extracted from roots and associated rhizospheric soil (**Fig. 2**) of different plant species stored in herbaria for 115-120 years. For each plant specimen, about 20 *ca* 0.5 cm-long root fragments are collected using sterile fine tweezers and scalpel blades. Soil particles aggregated around the roots (between 10-100 mg per plant specimen (Fig. 3), are detached from the roots and collected using tweezers. Root and soil samples are transferred into sterile tubes and stored at room temperature before DNA extraction.

Figure 2. Root system of herbarium specimen of Triticum aestivum collected in 1856 before (A) and after (B) the sampling of fragments of root and rhizospheric soil.

DNA extraction

This protocol is a modified version of Murchie et al., 2021's 'Cold Spin Extraction' method. (*see* **Note 2**)

1. Preheat the SLB_{conc} buffer at 50°C.
2. For each sample, weigh up to 250 mg of soil, or up to 50 mg of root in a 2 mL microtube. Prepare one extraction blank (no soil or root sample) for every series of extractions (i.e. for a total of 16 extractions, consider 15 samples and one extraction blank).
3. Transfer each sample in a single PowerBead tube provided in the DNEasy Powersoil Pro kit (already containing garnet beads and 750 µL of 181 mM NaPO4 and 121 mM of guanidinium isothiocyanate).
4. Rinse each sample tube with 500 µL of SLB_{conc} solution to collect the leftover of soil/root samples attached to the tube walls. Transfer the suspension to the corresponding PowerBead tube (making up a final volume of 1250 µL of digestion solution).
5. Homogenize the samples by a 5 min agitation of the PowerBead tubes in a mixer mill at a 25 beats/s mixing frequency. If necessary, repeat this step once or twice to obtain a homogeneous suspension.
6. Add 16 µL of 20 mg/mL Proteinase K solution (for a final concentration of approx. 0.25 mg/ml in the digestion solution).
7. Set the PowerBead tubes in a rotating shaker (speed 16-18 rotations/min) for an overnight digestion (20-24 h) in a hybridization oven at 35°C in the dark. Ensure that the digestion solution, sample, and PowerBeads are moving at each oscillation.
8. Remove PowerBead tubes from the oven and centrifuge 5 min at 10,000xg (the maximum speed recommended for PowerBead tubes).
9. Transfer supernatants to a DNA low-binding 2 mL tube and freeze them at -20°C. Recover as much digest solution as possible at this step. Tiny portions of the pellet can be pipetted without consequences (they will be eliminated on step 13). The extraction protocol can be stopped after that step (usually until the next day).
10. Preheat the EBT buffer at 30°C.
11. Thaw the digested supernatants, and centrifuge them briefly. Pipet each of them into a 15 mL Falcon tube filled with 13 mL of Binding Buffer.
12. Spin the Falcon tubes at 3,000 x g for a minimum of 3 hours (up to overnight) in a refrigerated centrifuge at 4°C.

13. Decant the supernatant after centrifugation, not disturbing the dark pellet at the bottom of the tube, and add it to a high-volume silica column (High Pure Extender Assembly; Roche Diagnostics).
14. Centrifuge the high-volume silica columns at 1,000 x g for 2 min. In case the entire volume has not passed through, renew the centrifugation step.
15. Detach the silica column from the assembly and put it in a 2 mL collection tube.
16. Add 500 µL of the Inhibitor Removal Buffer of the High Pure Viral Nucleic Acid Large Volume (Roche Diagnostics) to the column and centrifuge at 3,000 x g for 1 min at room temperature.
17. After centrifugation, transfer the column to a new collection tube and add 450 µL of the Wash Buffer (High Pure Viral Nucleic Acid Large Volume; Roche Diagnostics) to the column. Centrifuge at 6,500 x g for 1 minute at room temperature.
18. Repeat step 17 for a second wash.
19. Transfer the column to a new collection tube and centrifuge to dry the silica columns at 15,000 x g for 1 min.
20. Elute the DNA off the silica column with 25 µL EBT. Centrifuge at 15,000 x g for 1 min.
21. Repeat step 20 for a total elution volume of 50 µL.
22. Total DNA estimate is performed via fluorometric quantitation of 1 µL of each extract.
23. Store the extracted DNA at -20°C until processed into libraries and at -80°C for long-term storage.

Library preparation and sequencing

This protocol is a modified version of the dsDNA library preparation method by [11], implementing the double indexing strategy of [12]. All specimen extracts and blank extractions are to be processed in the same fashion.

Truncated adapter preparation

1. Incubate the 200 µM truncated P5 and P7 mixes 10 min at 95°C and let them cool down to 12°C slowly at -0.1°C /s.
2. Mix equal amounts of truncated P5 and P7 to obtain a 100 µM P5+P7 adapter mix. Vortex mix and centrifuge briefly.

3. Aliquot the adapter mix and freeze. Use each aliquot only once and then discard.

Universal library preparation

1. Prepare an End-Repair reaction premix containing : 5 µL of 10X NEBNext End-Repair buffer, 2.5 µL of NEBNext End-Repair Enzyme mix, and 27.5 µL of Nuclease-Free water per library.
2. Use 35 µL of the premix for each individual library. Add 15 µL of DNA extract (it is not necessary to use a specific quantity of DNA for library construction) for a final reaction volume of 50 µL.
3. Incubate 15 minutes at 25° C. Transfer at least 5 minutes at 4°C.
4. Proceed with a minelute purification using 6 volumes (300 µL) of PB_{acidic} for one volume of repaired DNA. Elute the DNA off the silica column twice with 15 µL of EBT for a final volume of approximately 29 µL.
5. Dilute a P5+P7 adapter mix aliquot to 20 µM. Make up a ligation premix containing: 10 µL of 5X NEBNext T4 quick ligation buffer, 7.5 µL of Nuclease-Free water, and 2.5 µL of the diluted adapter mix.
6. Add 20 µL of that premix to each library. Vortex mix and centrifuge briefly.
7. Add 1.5 µL of NEBNext T4 ligase to each individual library for a reaction volume of 50 µL (final concentration of 12 U of ligase / µL and 1 uM of P5+P7 adapter mix).
8. Incubate 90 minutes at 22°C.
9. Proceed with a minelute purification using 5 volumes (250 µL) of PB_{acidic} for binding DNA. Elute each library twice with 15 µL of EBT for a final volume of approximately 29 µL.
10. Prepare a fill-in reaction premix with the following reagents: 4 µL of Nuclease-free water 4 µL of 10X Thermopol buffer, 1 µL of 10 mM dNTPs, and 2 µL of Bst Polymerase large fragment (16 U per reaction).
11. Dispense 11 µL of the fill-in premix into each library tube (for a total reaction volume of 40 µL containing 1X Thermopol buffer and 250 uM of each dNTP). Vortex and quick spin.
12. Incubate 20 minutes at 37°C and then transfer at 80°C for another 20 minutes incubation.
    Store libraries at -20°C.

Library indexing and characterization

1. Amplify 10 µL of each library using a unique pair of custom P5 and P7 indexing primer in a PCR reaction using 1X SsoAdvanced Supermix and 500 nM of each primer. Due to various levels of inhibitions between samples, the total reaction volume can be adjusted from 40 (by default) up to 100 µL.
2. Perform the PCR amplification of the libraries in Real-Time with the following conditions: 2 minutes of hotstart denaturation at 98°C, followed by 20 cycles of 10 s at 98°C, 20 s at 60°C and 20 s at 72°C.
3. Address the level of fluorescence for each library separately at each cycle: remove individual libraries from the thermocycler when they shows a start in the plateau phase of the amplification.
4. Quantify each amplified library via fluorometric quantification.
5. Dilute the library according to the requirements of your capillary electrophoresis equipment. For a Labchip HT characterization, libraries are diluted to a 0.2-2.0 ng/µL range and evaluated with the NGS 3k chip and consumables.
6. Based on both fluorometric quantification and electrophoretic size distribution of the fragments, calculate the molarity of each individual library.

Pooling and Sequencing

1. Make an equimolar pool of the libraries to sequence, based on their estimated molarity.
2. Make a SPRI purification of the pool using 1.25X volume of NucleoMag purification beads.
3. Estimate the molarity of this final pool (via fluorometric and electrophoretic analyses) and prepare the Illumina sequencing accordingly.
4. Perform the shotgun sequencing of the pooled libraries on the relevant Illumina platform/chemistry. In our case, we sequenced the pool of DNA libraries made from either the roots or the soils herbarium samples on a Novaseq 6000 platform in Paired-End sequencing (2*50 bp). We aimed at generating approximately 20 million reads for each library.

Bioinformatics analyses

Preprocessing (quality control, trimming and merging)

1. Check quality of the reads using FastQC (Andrews, 2010) and visualize the HTML reports generated by FastQC using a web browser:
   >fastqc path_to_sample_name.fq.gz
2. Trim the reads to remove primer sequences and short reads and merge paired reads with Fastp (Chen et al., 2018) using the following command. For the trimming use the default commands except for the -l option (length of reads to trim) which has to be set to 25 nucleotides and the --overlap_len_require option (overlap needed between different pair reads, default value 30) which is here lowered to 10 nucleotides. These parameters are here adapted to the short length of the reads, i.e. usually less than 100 bp for ancient/historical DNA obtained from herbarium samples. As -l value was chosen 25 nucleotides to facilitate merging, mapping and assembly, in order to achieve high accuracy, since the trimmed sequences are longer than the default parameter of fastp (15 nucleotides) (*see* **Note 3**).
   >fastp -h -g -l 25 --adapter_fasta list_adapter.fasta --overlap_len_require 10
   -m -I sample_name_R1.fq -I sample_name_R2.fq -o sample_name_R1_trim.fastq.gz -O sample_name_R2_trim.fastq.gz --merged_out sample_name_trim_merge.fastq.gz --unpaired1 sample_name_R1_trim_unpaired.fastq.gz --unpaired2 sample_name_R2_trim_unpaired.fastq.gz

Removal of reads mapping to reference plant genomes

Sequences are then mapped to the corresponding plant reference genome (using available genomes present in the NCBI genomes database, https://www.ncbi.nlm.nih.gov/home/genomes/) to remove the plant nuclear, mitochondrial, and plastid genomes from the metagenomic datasets. The tool BWA aln (Li & Durbin, 2009) , appropriate for processing short sequences, is used for mapping Illumina reads to the plant genomes. Sequences that do not map to plants are then extracted using Samtools (Danecek et al., 2021). The hypothesis is that they correspond to the host plant microbiome (i.e., eubacteria, archaea, fungi, viruses).

1. Index the reference plant genome downloaded from the NCBI genomes database
   >bwa index -a bwtsw reference_genome.fasta
2. Align the merged sequences to the plant reference genome
   >bwa aln reference_genome.fasta sample_name_trim_merge.fastq.gz> sample_name_trim_merge.sai
3. Convert alignment in SAM format
   >bwa samse reference_genome.fasta sample_name_trim_merge.sai sample_name_trim_merge.fastq.gz >sample_name_trim_merge.sam
4. Convert SAM file into BAM file
   >samtools view -b -S sample_name_trim_merge.sam >sample_name_trim_merge.bam
5. Sort the BAM file
   >samtools sort sample_name*_trim_merge.bam -o sample_name_trim_merge_sort.bam*
6. Index the BAM file
   >samtools index sample_name*_trim_merge_sort.bam*
7. Calculate final statistics
   >samtools flagstat sample_name_trim_merge_sort.bam > sample_name_statistic_mapping.txt
8. Select unmapped sequences
   >samtools view -b -f 4 sample_name_trim_merge_sort.bam > sample_name_trim_merge_sort_unmapped.bam
9. Recovery fastq files for the unmapped sequences
   >samtools fastq sample_name_trim_merge_sort_unmapped.bam > sample_name_unmapped.fq

To exclude any contamination from sample manipulation in the herbaria, during sampling and analysis, repeat the same alignment procedure on the unmapped reads using the most recent version of the human genome (GCF_000001405.26) as the reference genome.

<u>*De novo assembly*</u>

1. Perform the *de novo* assembly of the reads using the software SPAdes with default parameters (*see* **Note 4**):

> spades.py -s sample_name_unmapped.fq.gz -o spades_results_sample_name/

2. Evaluate the assembly (i.e., number of assembled scaffolds, N50, N90, scaffold length distribution) with the tool QUAST, using default parameters (*see* **Note 5**):

> quast.py -k spades_results_sample_name/contigs.fasta spades_results_sample_name/scaffolds.fasta --single clean_reads.fq.gz -o quast_results_sample_name/

3. Collapse redundant scaffolds with CAP3 using default parameters:

> cap3 spades_results_sample_name/scaffolds.fasta -x sample_name

4. Concatenate in one file the obtained contigs and singlets sequences:

> cat spades_results_sample_name/scaffolds.fasta.sample_name.contigs spades_results_sample_name/scaffolds.fasta.sample_name.singlets > sample_name_cap3.fastac

Identification of viral sequences

To identify viral sequences, contigs are aligned against a protein reference database using the DIAMOND tool. A non-redundant protein database needs to be downloaded and formatted according to the following steps.

1. Download the non-redundant (nr) protein database from the NCBI website:

> wget ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz

2. Download the file that maps NCBI protein accession numbers to taxon ids form the NCBI website:

> wget ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/accession2taxid/prot.accession2taxid.FULL.gz

3. Download the nodes.dmp and names.dmp files from the NCBI taxonomy website:

> wget ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/taxdmp.zip

4. Create a DIAMOND-formatted database file (see **Note 6**):

> diamond makedb --in nr.gz -d nr.dmnd --taxonmap prot.accession2taxid.FULL.gz --taxonnodes nodes.dmp --taxonnames names.dmp

5. Run DIAMOND in blastx-mode and save the output in a tabular file (*see* **Note 7**):

> diamond blastx -f 6 –sensitive –quiet -d nr.dmnd -q sample_name_cap3.fasta -o sample_name.csv

6. Run DIAMOND in blastx-mode and save the output into DIAMOND alignment archive (DAA) supported by MEGAN:

> diamond blastx -d nr.dmnd -q sample_name_cap3.fasta -a sample_name.daa

Visualize DIAMOND results

The taxonomic distribution of the contigs annotated with DIAMOND can be then visualized with the tool MEGAN6.

1. Download the file mapping NCBI-nr accessions "megan-map-Feb2022.db.zip" to taxonomic and functional classes (NCBI, GTDB, EC, eggNOG, InterPro2GO, SEED) from *https://software-ab.cs.uni-tuebingen.de/download/megan6/welcome.html* and unzip it.
2. Open MEGAN6, click on the "File" menu, and select "Import From BLAST". In the tab "Files", at the point n.1, import your DAA file, and in the tab "Taxonomy", click on "Load MeganMapDB mapping file" and load the file "megan-map-Feb2022.db". Click on "Apply" to visualize your data (**Fig. 3**).
3. In order to extract viral contig sequences, select the node of interest, open the drop down menu by clicking the right mouse button and select the "Extract reads" option.

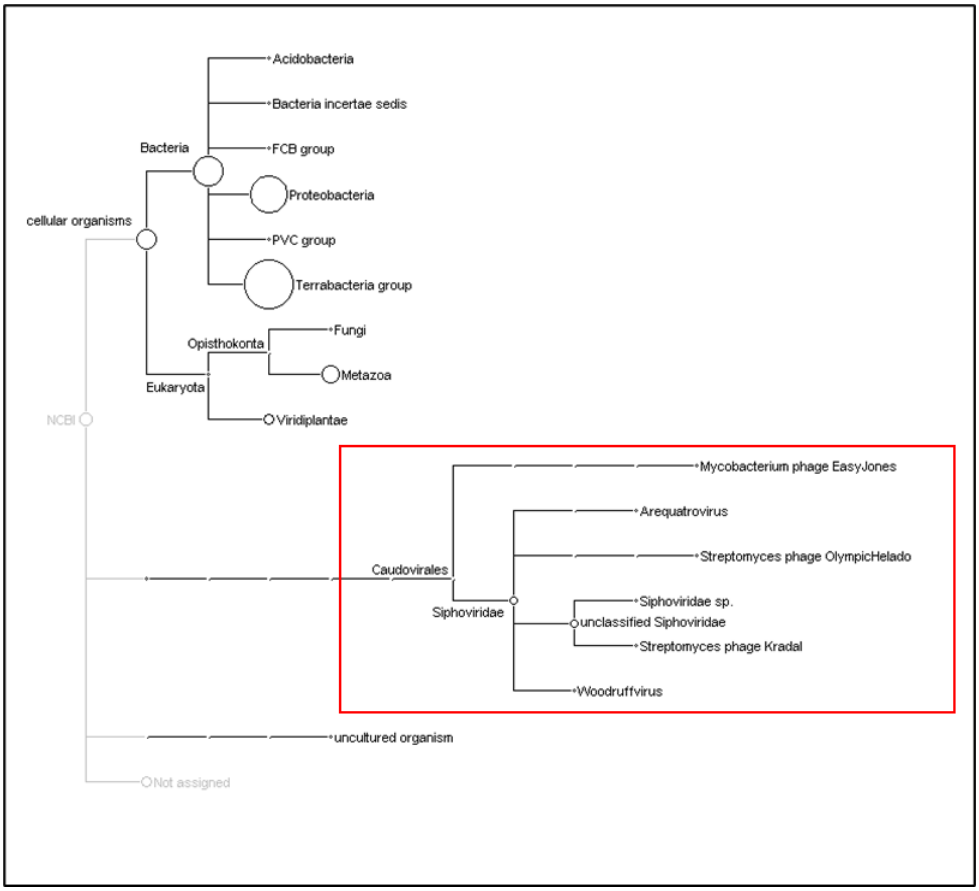


Figure 3. Visualization of DIAMOND results using MEGAN6. Red square indicate viral taxonomic nodes.

Notes

1.When preparing the BB buffer, note that the reaction is endothermic and that Guanidinium will only go into solution once the nuclease-free water is added to the mix.

2.DNA extraction was performed using the former PowerBead tubes from the DNA Powersoil kit (QIAGEN). It can readily be applied using the current version of the DNEasy Powersoil Pro kit (QIAGEN) by modifying step 3 in section 3.2. The tubes in the new kit contain only the beads,; it is thus necessary to add, before step 4, 750 µL of the CD1 solution (from the same kit).

3. Use -h option for obtaining result report on HTML, -g for removing polyG, -m for merging the reads, -adapter_fasta for indicating the fasta list of adapters (that contained adapters and their reverse complement). As indicated in the main text, the *-overlap_len_require* option was also introduced to reduce the overlap length during merging (from 30, default, to 10 nucleotides). Finally, with -q option, it is also possible to eliminate all the reads below a certain quality threshold. In our case, the raw data (in the FastQC reports) showed high quality, and the -q option was not used.

4. Use *-t* option for setting the number of threads and *-m* option to set memory limit in Gb, according to the characteristics of your IT infrastructure.

5. Use *-t* option for setting the number of threads and *-m* option to set memory limit in Gb, according to the characteristics of your IT infrastructure. Use *–silent* option if you do not want printed on screen detailed information about each step; the information will be stored anyway in the log file.

6. Use *-p* option for setting the number of threads according to the characteristics of your IT infrastructure. By default, DIAMOND uses all available threads.

7. Use *-p* option for setting the number of threads according to the characteristics of your IT infrastructure. By default, diamond uses all available threads. Use *-e* option to set the e-value threshold to report an alignment. Use *-f* option to format the output file. Use *–sensitive* option to enable the sensitive mode designed for full sensitivity for hits of >40% identity. Use *-k* option to set the maximum number of target sequences per query to report alignments for. Use *–quiet* option to disable all terminal output.

Acknowledgements

Work on ancient plant-associated microbiota was supported by the Muséum National d'Histoire Naturelle grant ATM 2021 (HoloHerbier). GG was supported by a PhD grant from the University of Turin and a mobility grant from the French-Italian University (program da Vinci 2022). We would like to thank the curators and staff of the herbaria of the Muséum National d'Histoire Naturelle in Paris, of the Lyon botanical garden and of the Universities of Montpellier, Lyon 1 and Clermont-Auvergne to give us access to herbarium specimens for implementation of the protocols described in this chapter. We thank the "Plateau de Paléogénomique et Génétique Moléculaire" (P2GM, MNHN, Paris), for granting access to their resources and facility.

References

Antonovics J, Hood ME, Thrall PH, Abrams JY, Duthie GM. Herbarium studies on the distribution of anther-smut fungus (Microbotryum violaceum) and Silene species (Caryophyllaceae) in the eastern United States. Am J Bot. 2003 Oct;90(10):1522-31. doi: 10.3732/ajb.90.10.1522. PMID: 21659105.

Lees, D.C., Lack, H.W., Rougerie, R., Hernandez-Lopez, A., Raus, T., Avtzis, N.D., Augustin, S. and Lopez-Vaamonde, C. (2011), Tracking origins of invasive herbReferencesivores through herbaria and archival DNA: the case of the horse-chestnut leaf miner. Frontiers in Ecology and the Environment, 9: 322-328.https://doi.org/10.1890/100098

Bradshaw M, Tobin PC. Sequencing Herbarium Specimens of a Common Detrimental Plant Disease (Powdery Mildew). Phytopathology. 2020 Jul;110(7):1248-1254. doi: 10.1094/PHYTO-04-20-0139-PER. Epub 2020 Jun 3. PMID: 32407253.

Yoshida K, Schuenemann VJ, Cano LM, Pais M, Mishra B, Sharma R, Lanz C, Martin FN, Kamoun S, Krause J, Thines M, Weigel D, Burbano HA. The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. Elife. 2013 May 28;2:e00731. doi: 10.7554/eLife.00731. Erratum in: Elife. 2013;2:e01108. PMID: 23741619; PMCID: PMC3667578

Campos PE, Groot Crego C, Boyer K, Gaudeul M, Baider C, Richard D, Pruvost O, Roumagnac P, Szurek B, Becker N, Gagnevin L, Rieux A. First historical genome of a crop bacterial pathogen from herbarium specimen: Insights into citrus canker emergence. PLoS Pathog. 2021 Jul 29;17(7):e1009714. doi: 10.1371/journal.ppat.1009714. PMID: 34324594; PMCID: PMC8320980.

Kocher, A., Papac, L., Barquera, R., Key, F. M., Spyrou, M. A., Hübler, R., Rohrlach2, A. B., Aron, F., Stahl, R., Wissgott, A., Van Bömme, F., Pfefferkorn, M., Alissa, M., Villalba-Mouco, V., Neumann, G. U., Rivollat, M., Van De Loosdrecht, M. S., Majander, K., Tukhbatova, R. I., … Kocher, A. (2021). Ten millennia of hepatitis B virus evolution. *Science*, *374*(6564), 132. https://doi.org/https://www.science.org/doi/10.1126/science.abi5658

Bonczarowska JH, Susat J, Mühlemann B, Jasch-Boley I, Brather S, Höke B, Brather-Walter S, Schoenenberg V, Scheschkewitz J, Graenert G, Krausse D, Francken M, Jones TC, Wahl J, Nebel A, Krause-Kyora B. Pathogen genomics study of an early medieval community in Germany reveals extensive co-infections. Genome Biol. 2022 Dec 13;23(1):250. doi: 10.1186/s13059-022-02806-8. PMID: 36510283; PMCID: PMC9746117.

Rieux A, Campos P, Duvermy A, Scussel S, Martin D, Gaudeul M, Lefeuvre P, Becker N, Lett JM. Contribution of historical herbarium small RNAs to the reconstruction of a cassava mosaic geminivirus evolutionary history. Sci Rep. 2021 Oct 28;11(1):21280. doi: 10.1038/s41598-021-00518-w. PMID: 34711837; PMCID: PMC8553777.

Malmstrom, C.M., Shu, R., Linton, E.W., Newton, L.A. and Cook, M.A. (2007), Barley yellow dwarf viruses (BYDVs) preserved in herbarium specimens illuminate historical disease ecology of invasive and native grasses. Journal of Ecology, 95: 1153-1166. https://doi.org/10.1111/j.1365-2745.2007.01307.x

Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010 Jun;2010(6):pdb.prot5448. doi: 10.1101/pdb.prot5448. PMID: 20516186.

Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 2012 Jan;40(1):e3. doi: 10.1093/nar/gkr771. Epub 2011 Oct 21. PMID: 22021376; PMCID: PMC3245947.

Murchie T.J., M. Kuch, A.T. Duggan, M.L. Ledger, K. Roche, J. Klunk, E. Karpinski, D. Hackenberger, T. Sadoway, R. 2021. Optimizing extraction and targeted capture of ancient

environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. Quat. Res. 99:305-328

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Chen S, Yanqing Z, Chen Y., Gu J.; fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics, Volume 34, Issue 17, 1 September 2018, Pages i884–i890, https://doi.org/10.1093/bioinformatics/bty560

Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25, 1754-1760. [PMID: 19451168]

Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., O Pollard M., Whitwham A., Keane T., McCarthy S.A., Davies R.M., Li H.. Twelve years of SAMtools and BCFtools , GigaScience, Volume 10, Issue 2, February 2021, giab008, https://doi.org/10.1093/gigascience/giab008

Chapter 4. Herbaria as repositories of ancient soil samples

Gianluca Grasso[1,2,3,†], Régis Debruyne[4,], The Herbonaute consortium#, Eva Perez Pimpare [5], Martino Adamo[1], Naoise Nunan[6], Mercedes Mendez-Millan[7], Valeria Bianciotto[3,#], Roland Marmeisse[2,3,#]

1. Department of Life Science and Systems Biology (DBIOS), Università degli Studi di Torino, 10125 Torino, Italy.

2. Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université Des Antilles, 75005 Paris, France.

3. Institute for Sustainable Plant Protection (IPSP), National Research Council (CNR), SS, Viale Mattioli 25, 10125, Torino, Italy.

4. Muséum National d'Histoire Naturelle, Archéozoologie, Archéobotanique: Sociétés, Pratiques et Environnements (AASPE: UMR 7209 CNRS-MNHN), CP 56, 43 rue Buffon, F-75005 Paris, France.

5, Infrastructure Récolnat, Direction Générale Déléguée aux Collections MNHN, Paris

6, UMR 7618, iEES, UMR7618 CNRS-Sorbonne Université, Paris

7 IRD, Plateforme Alysés, Bondy

Introduction

Soil microbes play a crucial role in essential soil functions such as nutrient cycling, pathogen control, symbiotic interactions, and bioremediation, while also supporting plant life through nutrient availability, protection, defense, and tolerance to biotic and abiotic stresses (Banerjee & A Heijden, 2023; Wei et al., 2019; Zhang et al., 2021)

However, in the absence of historical archives, the only available data on the impact of current environmental changes on soil microbial diversity come mainly from short-term impact studies that compare soils subjected to different management practices or along environmental gradients (Berg & Cernava, 2022; Nic Lughadha et al., 2020; Rull, 2021; Zhou et al., 2020). These studies, which rely on high-throughput sequencing techniques to evaluate microbial diversity, have been conducted in the last 15 years in the 21st century (Zhou et al., 2020) and thus represent a rather limited period of time.

Recently, herbaria have been identified as potential sources of ancient soil samples inadvertently left associated to the root systems of dried plants (Bianciotto et al., 2022). From different of these soils it was possible to extract historical DNA whose sequencing revealed the presence of microbial sequences and even allowed assembling the full genome of an unknown plant virus (Grasso et al., 2024; Chapter 3). To be exploited to retrace the recent trajectories of soil microbial communities it is however essential to demonstrate that herbarium soil samples are not anecdotic but are rather numerous, associated to many different plant taxa, representative of large geographic areas encompassing different climates and soils and collected over significant periods of time. It is also essential to demonstrate that historical DNA extraction and conversion to sequencing libraries can be performed on a wide range of these soils whatever their origin, composition and collection time.

In the present study we evaluated these requirements by looking at digitized herbarium plant specimens preserved in 32 institutional herbaria in France. These herbaria were fully or partially digitized to produce a curated database of about 2333 high-resolution pictures freely consultable through the web (https://www.recolnat.org/fr/). The study focused on a set of 14 different commonly cultivated crop plants collected in mainland France.

From a sample of 30 soil herbarium specimens of *Triticum* and *Hordeum* we further performed soil DNA extraction and sequencing library construction to

evaluate if and how collection date, soil characteristics and putative local environmental conditions affected historical soil DNA extractability and preservation.

Materials and Methods

Annotation of a herbarium database and soil sampling

More than 3500 digitized herbarium specimens (3585) from the Recolnat database attributed to 14 cultivated plant taxa collected in France (European territory) were selected. These database was annotated through a citizen science project (http://lesherbonautes.mnhn.fr/). This manual annotation project (*'Aux racines de l'agriculture en France'*; http://lesherbonautes.mnhn.fr/missions/22858123) consisted in entering in a database hand-written information of the specimen's label as well as in indicating the presence of roots and of soil associated to the roots. For the current project, the two most relevant information indicated, or not, on the label were the specimen collection date and place. Collection place referred to the "commune" administrative level whose mean area is of 14.9 km^2 in continental France; thus corresponding to an average linear geographic distance of less than 5 km between two neighboring communes. An additional information of interest regarded, often informal, indications that suggested whether the plants were collected in agricultural fields.

From the indications present in this database, we selected 30 plants of either *Triticum aestivum* or *Hordeum vulgare* for soil sampling (**Supplementary Table S1**). Selection criteria were as followed: presence of soil on the root systems, known and diverse collections dates and places and indication suggesting that the plants were collected in cultivated fields. At least 100 mg (dry weight) of soil were collected on each plant; 50 mg for DNA extraction and 50mg for chemical analyses.

Chemical analyses of herbarium soils

Concentrations in total organic carbon (Corg) and nitrogen (N), along with ^{13}C and ^{15}N contents, were determined using a Thermo Fisher Scientific Elemental Analyzer Flash 2000 HT connected to an Isotope Ratio Mass Spectrometer Delta V Advantage (EA-IRMS). When necessary, for the determination of Corg and ^{13}C content, carbonates were first removed by treating soils with 2% HCl. For N and ^{15}N measurements, bulk soil samples were always used directly. Isotope ratios are expressed in the conventional delta (δ in ‰) notation, with $\delta^{13}C$ values

referenced to the Pee Dee Belemnite (PDB) carbonate standard and $\delta^{15}N$ values referenced to atmospheric nitrogen value.

Total major (Na, Mg, Al, Ca, P, Mn, Fe) and trace (Zn, Cd, Pb) element contents were determined by ICP-MS at the Institut de Physique du Globe (IPSP), Paris, after acid digestion following the methodology used by Valdés et al., 2014. The procedure included placing 25 mg samples in perfluoroalkoxy (PFA) Savillex™ vessels for treatment at 150°C for 48 hours with a combination of nitric acid (HNO_3) and hydrofluoric acid (HF). A mixture of HF and perchloric acid ($HClO_4$) was then added and digestion proceeded at 150°C for 24 hours. After two treatments with HNO_3 at 150°C to ensure complete evaporation of the acids, the residue was diluted to 35 mL with HNO_3. Analytical precision and accuracy were ensured through routine replicate analyses, the use of target materials, and the MESS-3 certified reference material. Validation data showed an accuracy with a relative error not exceeding 5%.

X-ray fluorescence (XRF) spectroscopy was used on native soil samples to determine SiO_2 content that cannot be assayed by ICP-MS on HF-digested samples. A portable X-ray fluorescence (XRF) spectrometer was used as presented by Goudard et al., (2024) for the measurement of mineral elements content, including rare earth, directly on herbarium plant samples.

DNA extraction and sequencing library construction

Total DNA was extracted following the method of Murchie et al., 2021 optimised for ancient sediment samples, with minor modifications ((Grasso et al., 2024), described in chapter 3). Control extraction and blank samples were systematically performed to mitigate the risk of in-laboratory contamination. Herbarium DNA were characterised by fluorometric quantification (Qubit II, Invitrogen) and capillary electrophoresis (Labchip GX Touch, Perkin-Elmer) after their conversion in Illumina sequencing libraries using a double-indexing PCR (Kircher et al., 2012).

Data analysis

Data analysis was performed using R v.4.2.2. The database obtained from the *Les Herbonautes* platform was analyzed using the *tidyverse* v.2.0.0 (Wickham et al., 2019), *ggplot2* v.3.5.1 (Wickham H, 2016)., and *ggrepel* v.0.9.4 packages (Slowikowski K, 2024).

The spatial coverage of samples obtained from the entire Herbonautes database and for the 30 selected *Triticum* and *Hordeum* plants sampled for soil was analyzed using the R-package *geodata* v.0.5-9 (Hijmans et al., 2024), *terra* v.1.7.55 (Robert & Hijmans, 2024) and *tidyterra* v.4.2.3 (Hernangómez, 2023). Various local soil parameters such as pH, total nitrogen, SOC (Soil Organic Carbon), CEC (Cation Exchange Capacity), and bulk density were also extracted from the SoilGrid database (https://soilgrids.org/, Poggio et al., 2021) based on the herbarium plant collection site positions using the average municipal values of the parameters created using *QGIS v 3.36.0-Maidenhead*. Herbarium sample values were then confronted with all average values of the whole of mainland France to evaluate their representativeness across the entire geographic area under consideration.

A regression model (Generalized Linear Mixed Model, GLMM) was implemented to explore the relationships between quantities of extracted DNA and soil characteristics (Carbon organic %, Nitrogen %, δ13C, δ15N, Na, Mg, Al, P, K, Ca, Mn, Zn, Cd and SiO_2, contents) as well as collection site variables (bioclimatic and soil variables). As in the case of soil variables that were extracted from the Soil Grid database, bioclimatic ones were extracted from the WorldClim version 2.1 database (https://www.worldclim.org/, Fick & Hijmans, 2017). Models were built using the same methods described in (Adamo et al., 2021). Iron (Fe), carbon organic percentage, bulk soil density, annual temperature and pH were all correlated with other variables of the dataset (Pearson's $r > 0.7$, **Supplementary Fig.S1**) and were thus removed from the analysis. GLMM constructed with *lme4* (Bates et al., 2015) were fitted to these data using a Poisson distribution and a log link function. All variables were scaled and the plant genus variable was used as random factor. GLMM were built using all the non-collinear variables and the non-associated factors (herbarium collections, collection sites, collection years and plant genus) selected after data exploration (the equation is in R notation):

m1 = lme4::lmer(DNA-ng ~ Na + Mg + Al + P + K + Ca + Mn + Zn + Cd + year + X.Nitro + SiO2 + soilN + elev + soc + annual_prec + cat_ex + δ15N + δ13C + region + herb + (1|genus), data = data).

Once the initial model had been fitted, a model selection was performed by backward elimination. Model reduction was carried out based on Aikaike information criterion values in order to simplify it and avoid overfitting (Johnson

& Omland, 2004). The final model was valued by checking overdispersion and standard residuals plots (**Supplementary Fig.S2**, Zuur & Ieno, 2016)

The data was also analyzed through a Principal Component analysis (PCA) using the packages *ggcorrplot* v.0.1.4.1 (Kassambara, 2023), *FactoMineR* v.2.9 (Lê et al., 2008) and *factoextra* v.1.0.7 (Kassambara & Mundt, 2020).

Results and discussion

Characteristics and temporal and spatial distribution of herbarium samples

The database built by *Les Herbonautes* network includes 3585 herbarium accessions of 14 species of crop plants in the genera *Avena, Cannabis, Cicer, Fagopyrum, Hordeum, Lens, Linum, Lupinus, Panicum, Phacelia, Secale, Sorghum, Triticum* and *Zea*.

From these 3585 specimens, 2333 had a precise collection location (commune level, excluding sample collected in botanical gardens and from Paris). These later specimens cover a time span from 1700 to 2014, but only 3 were collected in the 18th century and 14 in the 21st one; 7.6% had no associated collection date (**Fig 1.C**).

Geographically, the plants are distributed evenly across mainland France (**Fig 1. A**). 30.4% of the specimens had roots and soil associated, 29.3% had only roots, while the remaining 40.3% had only aboveground organs (**Fig 1.B**).

The percentages of root-soil presence in the specimens were calculated for each plant genus (**Supplementary Fig. S3**). Plants belonging to the genera *Avena, Hordeum, Polygonum, Secale*, and *Triticum* had a higher percentage of specimens with soil (always greater than 45%) compared to plants in the genera *Cannabis, Cicer, Sorghum* and *Zea* where most specimens lacked both soil and roots. In these cases, the absence of roots could certainly be explained by the large size and thickness of the mature plants that make them difficult to place them entire in a herbarium folder. Other plant genera (*Lens, Linum, Lupinus*, and *Phacelia*) had predominantly (more than 44%) specimens with roots without attached soil. The presence of soil on the specimens could have been influenced by the morphology and architecture of the plants' root systems. In the Poaceae (such as *Avena, Hordeum, Secale* and *Triticum*), presence of highly ramified fibrous adventitious crown roots probably facilitates root system extirpation that retains soil particles difficult to wash away during sample preparation.

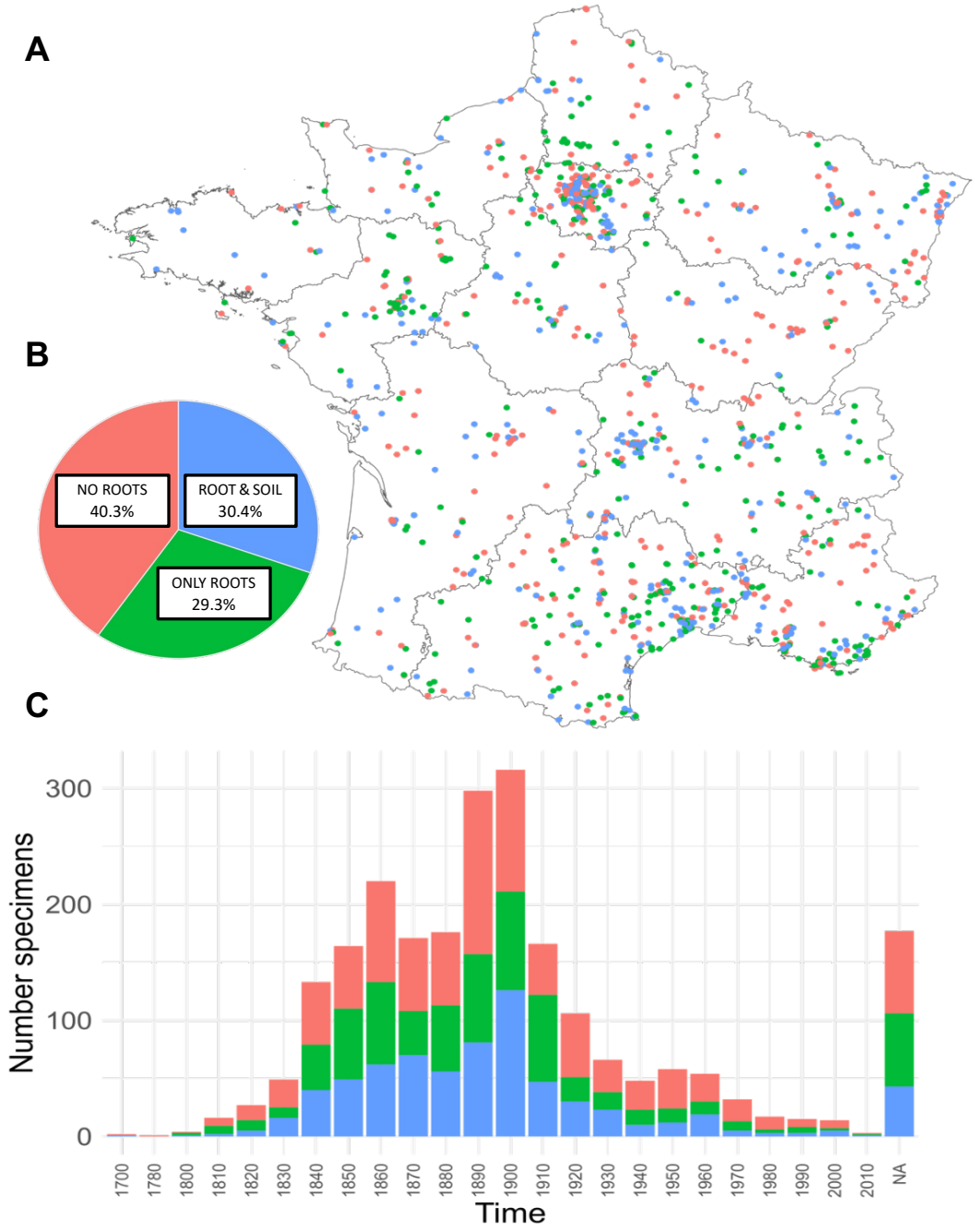


Figure 1. Main characteristics of the annotated crop plant dataset. **A**. Map showing the geographical distribution of the plant specimens across mainland France. **B**. Pie chart of the percentages of specimens with roots and soil (red), only roots (green), and only aboveground organs (blue). **C**. Bar plot representing the temporal distribution of the specimens from the 18th century to 2010; NA, plants without known collection date

Representativeness of herbarium specimens in France

Herbarium cultivated plants were collected all over the entire French mainland territory with however some collection hot spots around the city of Paris and along the Mediterranean coast in South France (**Fig. 1A**). Regarding collection dates, most (70.9%) of the plants were collected over a period of less than a century between the 1840s and the 1920s, a period that precedes today's global environmental changes (**Fig. 1C**).

In the case of plants with roots and associated soil that could be sampled to study past soil microbial diversity, we extracted from soil and climatic databases several putative geographic (elevation), climatic (annual mean temperature, annual precipitation) and soil (bulk density, cation exchange capacity, organic carbon density, pH, soil organic carbon, total nitrogen) characteristics related to their collection sites. These predicted values were then confronted to the distribution of the same parameters across the entire French territory (**Fig. 2**). For each of these nine parameters, we observed a clear superposition in the distribution of the values of the collected samples and of the

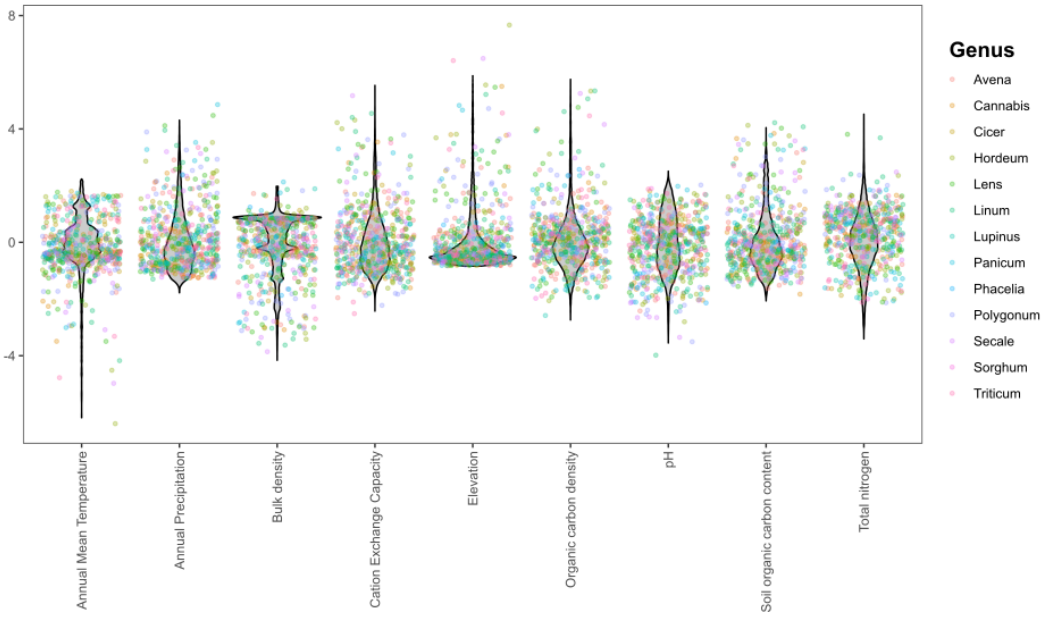


Figure 2. Representativeness of herbarium specimens in continental France. The plot shows in the background (in grey) the violin plots of all French municipal averages of the values taken into consideration (annual mean temperature, annual precipitation. Cation exchange capacity, elevation, organic carbon density, pH, soil organic carbon content, soil total nitrogen); in the foreground are represented the different values of the variables (estimated from the GPS position of the collecting city) of the individual herbarium specimens present in the database and colored by genus.

values observed across the entire French territory. It thus appears that soils associated to cultivated plants stored in herbaria are likely to be representative of the diversity of agricultural soils and of agricultural areas of France that are distributed across different climatic regions (oceanic, Mediterranean, continental and mountain).

DNA extraction of soil herbarium specimens

To promote herbarium soils as sources of DNA for the survey of ancient soil microbial communities we needed to demonstrate that historical DNA could indeed be extracted from a variety of these soils. For this purpose, we selected a set of 30 samples associated to *T. aestivum* and *H. vulgare* plants collected over a period of 150 years (1821-1977) and representative of the pedo- and bioclimatic regions of France (Supplementary **Figure S3 & Supplementary Table S2**). Diversity of these soils was confirmed by the direct measurement of their elemental composition. Diversity could be illustrated by the concentration ranges of Si (from 8.42 to 51.96 g/100g) and Ca (from 0.26 to 19.24 g/100g) that suggest that the studied soils have developed on very different parent rock materials **Supplementary Table S3**).

DNA extraction was successfully achieved using a single protocol for all 30 herbarium soil samples. All DNA samples were further transformed in Illumina libraries, demonstrating that all DNA extracts were free of inhibitors that would have prevented enzymatic (proteinase K, DNA polymerase) manipulation.

The extracted DNA samples exhibited very similar fragmentation profiles. The log-normal distribution of the DNA fragments peaked between 61 and 86 bp across the various soil samples (**Fig. 3.A**) and no correlation between peak position and age of the samples was observed (**Fig. 3.B**). This could suggest that most DNA fragmentation occurred at initial stages of soil storage, possibly during sample drying and then slowed down or even almost ceased. This observation is at odd with results reported by Weiß et al. (2016) who observed a weak correlation between herbarium plant (*Solanum* spp. and *Arabidopsis*) DNA fragmentation and plant collection date. In this latter study size of the extracted plant DNA fragments are however in the same range as those observed for soil DNA.

As opposed to DNA fragmentation, amounts of extracted DNA greatly differed between soils, ranging from 9.3 $ng.g^{-1}$ of soil to 589.5 $ng.g^{-1}$ (**Supplementary Table S1).** Statistical analyses were performed to test whether DNA extraction yield could be related to collection date or to one or several of the soil and collection site characteristics. A PCA analysis suggested that predicted pH, bulk soil density (bdod) values, and measured Mg and SiO_2 contents, but not sample age, are among the key factors that contribute most to

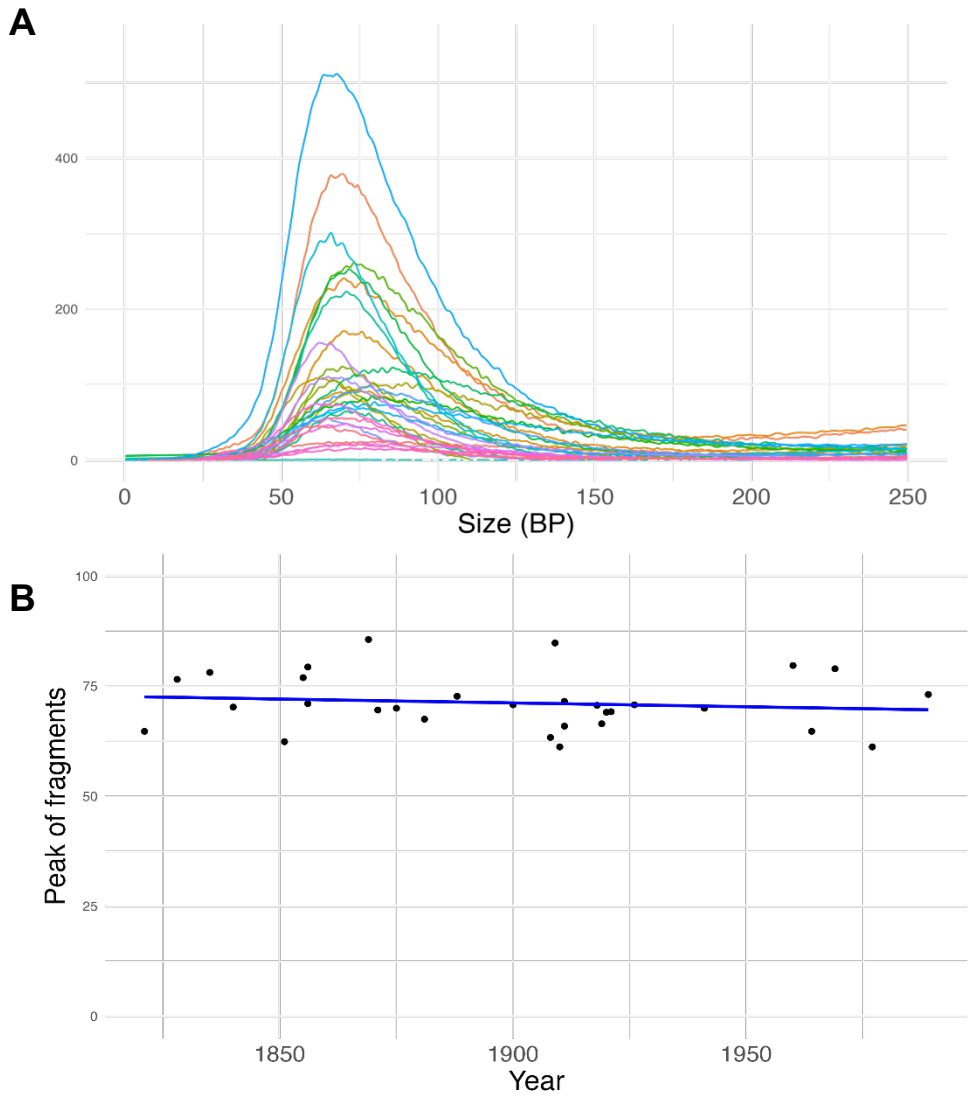


Figure 3. A. Degradation profile of herbarium soil extracted DNA. **B.** Plot of the peaks of the extracted DNA curves and the year of herbarium sample collection. The correlation is nearly zero (-0.12) and shows no influence of the collection date on the degradation of the extracted DNA.

the variability between samples (**Fig S6**). Thus, high pH and bdod (highly correlated to each other) values seem to negatively affect soil DNA extraction yield. These results are somehow in linewith those reported by (Dequiedt et al., 2011) who showed that at the scale of the entire French territory, several soil characteristics, but not climatic ones correlate with the amount of extractable microbial DNA.

GLMM were implemented to more precisely define the factors that affected DNA recovery (**Fig. 4**). Three factors emerged, total soil N and Na contents appeared significantly ($P < 0.01$ and $P < 0.05$, respectively) and positively correlated to DNA yield, while SiO_2 content was negatively correlated ($P<0.01$).

The quantity of extracted DNA from an ancient soil possibly reflects both the amount of DNA (microbial biomass) present in the fresh soil when it was sampled and the post-harvest time-dependent DNA degradation process. While it is difficult to speculate about the roles of SiO_2 and Na concentrations to the aDNA extraction yields, a high nitrogen content is likely to reflect soil fertility and is certainly positively correlated to microbial biomass.

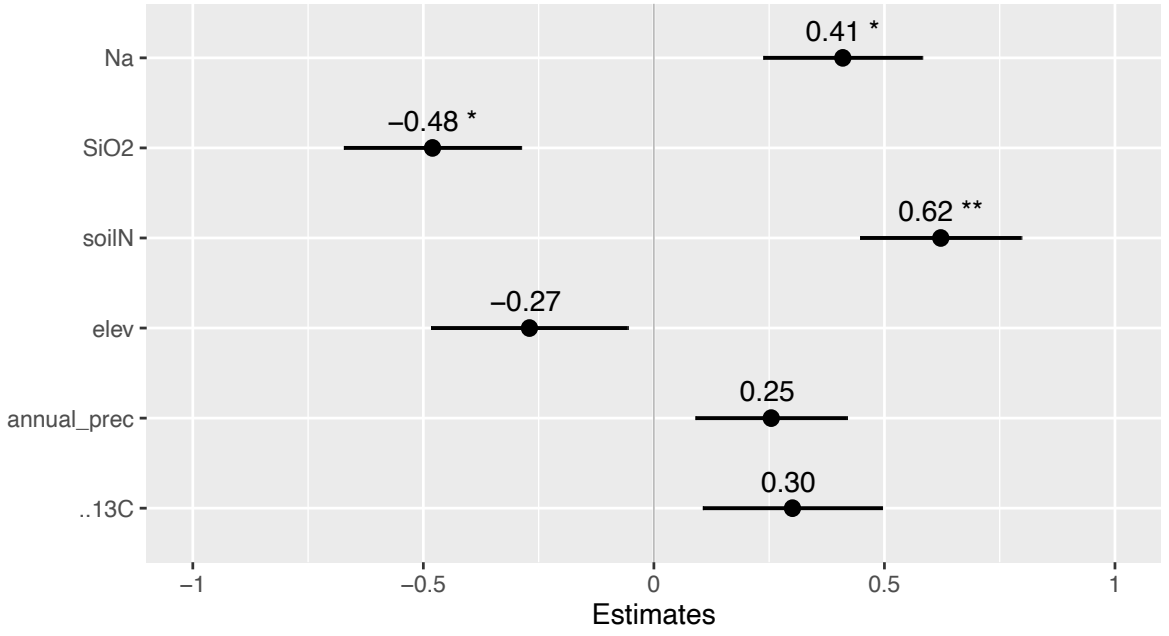


Figure 4. Regression model results highlighting the most significant variables in relation to the amount of extractable DNA from the analyzed samples (above). Significant variables in the model are marked with * (P value < 0.05) and ** (P value < 0.01).

Conclusion

Herbaria are undergoing a renaissance as sources of genomic data for exploring plant evolution, ecology, and diversity. Ancient DNA preserved in herbarium specimens offers unprecedented insights into historical plant and soil communities. This chapter demonstrates how certain crop plants (Poaceae, Polygonaceae, and Leguminosae) are particularly suitable for studying the soil and/or roots of plants preserved in herbaria. The presence of roots and soil is far from rare in herbarium plants, and numerous samples can provide this material to investigate large geographic areas during a long (150-200 years) period. Additionally, herbarium soil appears to preserve DNA, potentially containing traces of the original microbiome (as will be described in the following chapter). These samples should allow us to investigate how soil and plant microbial communities responded to environmental and climate changes since the 1800s.

Acknowledgements

Work on herbarium soil-associated microbiota was supported by the Emergence program of Sorbonne Université (HerbaSoil). GG was supported by a PhD grant from the University of Turin and a mobility grant from the French-Italian University (program da Vinci 2022). We thank “Les Herbonautes” platform for allowing us to investigate soil presence in French herbarium collections. We thank all the herbaria where soil samples were collected (MNHN, Université Lyon I, Universitè Montpellier II, Herbiers Universitaires of Clermont-Ferrand, Jardin botanique de Lyon). We thank the Platform P2GM from the MNHN, for granting access to their ancient DNA facilities.

References

Adamo, M., Chialva, M., Calevo, J., Bertoni, F., Dixon, K., & Mammola, S. (2021). Plant scientists' research attention is skewed towards colourful, conspicuous and broadly distributed flowers. *Nature Plants 2021 7:5*, *7*(5), 574–578. https://doi.org/10.1038/s41477-021-00912-2

Banerjee, S., & A Heijden, M. G. (2023). Soil microbiomes and one health. *Nature Reviews Microbiology*, *21*, 6–20. https://doi.org/10.1038/s41579-022-00779-w

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/JSS.V067.I01

Berg, G., & Cernava, T. (2022). The plant microbiota signature of the Anthropocene as a challenge for microbiome research. *Microbiome*, *10*(1), 1–12. https://doi.org/10.1186/S40168-021-01224-5/FIGURES/1

Bianciotto, V., Selosse, M. A., Martos, F., & Marmeisse, R. (2022). Herbaria preserve plant microbiota responses to environmental changes. *Trends in Plant Science*, *27*(2), 120–123. https://doi.org/10.1016/J.TPLANTS.2021.11.012

Chialva, M., & Bonfante, P. (2018). Plant microbiota: From model plants to Mediterranean crops. In *Phytopathologia Mediterranea* (Vol. 57, Issue 1, pp. 123–145). Firenze University Press. https://doi.org/10.14601/Phytopathol_Mediterr-22052

Dequiedt, S., Saby, N. P. A., Lelievre, M., Jolivet, C., Thioulouse, J., Toutain, B., Arrouays, D., Bispo, A., Lemanceau, P., & Ranjard, L. (2011). Biogeographical patterns of soil molecular microbial biomass as influenced by soil characteristics and management. *Global Ecology and Biogeography*, *20*(4), 641–652. https://doi.org/10.1111/J.1466-8238.2010.00628.X

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/JOC.5086

Goudard, L., Blaudez, D., Sirguey, C., Purwadi, I., Invernon, V., Rouhan, G., & van der Ent, A. (2024). Prospecting for rare earth element (hyper)accumulators in the Paris Herbarium using X-ray fluorescence spectroscopy reveals new distributional and taxon discoveries. *Annals of Botany*, *133*(4), 573–584. https://doi.org/10.1093/AOB/MCAE011

Grant, J. R., Enns, E., Marinier, E., Mandal, A., Herman, E. K., Chen, C. Y., Graham, M., Van Domselaar, G., & Stothard, P. (2023). Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research*, *51*(W1), W484–W492. https://doi.org/10.1093/NAR/GKAD326

Grasso, G., Rotunno, S., Debruyne, R., Bittner, L., Miozzi, L., Marmeisse, R., & Bianciotto, V. (2024). Identification of DNA Viruses in Ancient DNA from Herbarium Samples. *Methods in Molecular Biology*, *2732*, 221–234. https://doi.org/10.1007/978-1-0716-3515-5_15/COVER

Hernangómez, D. (2023). Using the tidyverse with terra objects: the tidyterra package. *Journal of Open Source Software*, *8*(91), 5751. https://doi.org/10.21105/JOSS.05751

Hijmans, R. J., Barbosa, M., Ghosh, A., Mandel, A., & Maintainer,]. (2024). *Title Download Geographic Data*. https://github.com/rspatial/geodata/issues/

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology and Evolution*, *19*(2), 101–108. https://doi.org/10.1016/j.tree.2003.10.013

Kassambara, A. (2023). *Visualization of a Correlation Matrix using "ggplot2" [R package ggcorrplot version 0.1.4.1]*. https://doi.org/10.32614/CRAN.PACKAGE.GGCORRPLOT

Kassambara, A., & Mundt, F. (2020). Kassambara, A. and Mundt, F. (2020) Factoextra Extract and Visualize the Results of Multivariate Data Analyses. *R Package Version 1.0.7. - References - Scientific Research Publishing*. https://scirp.org/reference/referencespapers?referenceid=3067217

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1). https://doi.org/10.1093/NAR/GKR771

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, *25*(1), 1–18. https://doi.org/10.18637/JSS.V025.I01

Murchie, T. J., Kuch, M., Duggan, A. T., Ledger, M. L., Roche, K., Klunk, J., Karpinski, E., Hackenberger, D., Sadoway, T., Macphee, R., Froese, D., Poinar, H., & Corresponding, *. (2021). Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. *Quaternary Research*, *99*, 305–328. https://doi.org/10.1017/qua.2020.59

Nic Lughadha, E., Bachman, S. P., Leão, T. C. C., Forest, F., Halley, J. M., Moat, J., Acedo, C., Bacon, K. L., Brewer, R. F. A., Gâteblé, G., Gonçalves, S. C., Govaerts, R., Hollingsworth, P. M., Krisai-Greilhuber, I., de Lirio, E. J., Moore, P. G. P., Negrão, R., Onana, J. M., Rajaovelona, L. R., … Walker, B. E. (2020). Extinction risk and threats to plants and fungi. *Plants, People, Planet*, *2*(5), 389–408. https://doi.org/10.1002/PPP3.10146

Pearman, J. K., Biessy, L., Howarth, J. D., Vandergoes, M. J., Rees, A., & Wood, S. A. (2022). Deciphering the molecular signal from past and alive bacterial communities in aquatic sedimentary archives. *Molecular Ecology Resources*, *22*(3), 877–890. https://doi.org/10.1111/1755-0998.13515

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL*, *7*(1), 217–240. https://doi.org/10.5194/SOIL-7-217-2021

Robert, M., & Hijmans, J. (2024). *Package "terra" Type Package Title Spatial Data Analysis*.

Rull, V. (2021). Biodiversity crisis or sixth mass extinction? *EMBO Reports*, *23*(1). https://doi.org/10.15252/EMBR.202154193

Siano, R., Lassudrie, M., Cuzin, P., Briant, N., Loizeau, V., Schmidt, S., Ehrhold, A., Mertens, K. N., Lambert, C., Quintric, L., Noël, C., Latimier, M., Quéré, J., Durand, P., & Penaud, A. (2021). Sediment archives reveal irreversible shifts in plankton communities after World War II and agricultural pollution. *Current Biology*, *31*(12), 2682-2689.e7. https://doi.org/10.1016/J.CUB.2021.03.079

Valdés, J., Guiñez, M., Castillo, A., & Vega, S. E. (2014). Cu, Pb, and Zn content in sediments and benthic organisms from San Jorge Bay (northern Chile): Accumulation and biotransference in subtidal coastal systems. *Ciencias Marinas*, *40*(1), 45–58. https://doi.org/10.7773/CM.V40I1.2318

Wei, Z., Gu, Y., Friman, V. P., Kowalchuk, G. A., Xu, Y., Shen, Q., & Jousset, A. (2019). Initial soil microbiome composition and functioning predetermine future plant health. *Science Advances*, *5*(9), 759–784. https://doi.org/10.1126/SCIADV.AAW0759/SUPPL_FILE/AAW0759_SM.PDF

Weiß, C. L., Schuenemann, V. J., Devos, J., Shirsekar, G., Reiter, E., Gould, B. A., Stinchcombe, J. R., Krause, J., & Burbano, H. A. (2016). Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *R. Soc. Open Sci*, *3:160239*. https://doi.org/10.1098/rsos.160239

Wickham, H., Averick, M., Bryan, J., Chang, W., D', L., Mcgowan, A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/JOSS.01686

Zhang, J., Cook, J., Nearing, J. T., Zhang, J., Raudonis, R., Glick, B. R., Langille, M. G. I., & Cheng, Z. (2021). Harnessing the plant microbiome to promote the growth of agricultural crops. *Microbiological Research*, *245*, 126690. https://doi.org/10.1016/J.MICRES.2020.126690

Zhou, Z., Wang, C., & Luo, Y. (2020). Meta-analysis of the impacts of global change factors on soil microbial diversity and functionality. *Nature Communications 2020 11:1*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-16881-7

Zuur, A. F., & Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, *7*(6), 636–645. https://doi.org/10.1111/2041-210X.12577

Supplementary

Fig S1. Multi-collinearity analysis among environmental variables. Above the diagonal: Pearson's r correlation coefficient. Below the diagonal: scatterplots and histograms of data distribution. Pearson's correlation values above the threshold of collinearity used in this study (> 0.7).

Variable;: Al (aluminum), Fe (Iron), X.carbon (percent of organic carbon measured with EA-IRMS), X.nitro (nitrogen percentage measured with EA-IRMS), bdob (soil bulk density from SoilGrid), soilN (total nitrogen soil from SoilGrid), elev (elevation), soc (soil organic content from SoilGrid), annual_prec (annual mean precipitation), annualT: (annual mean temperature), pH.

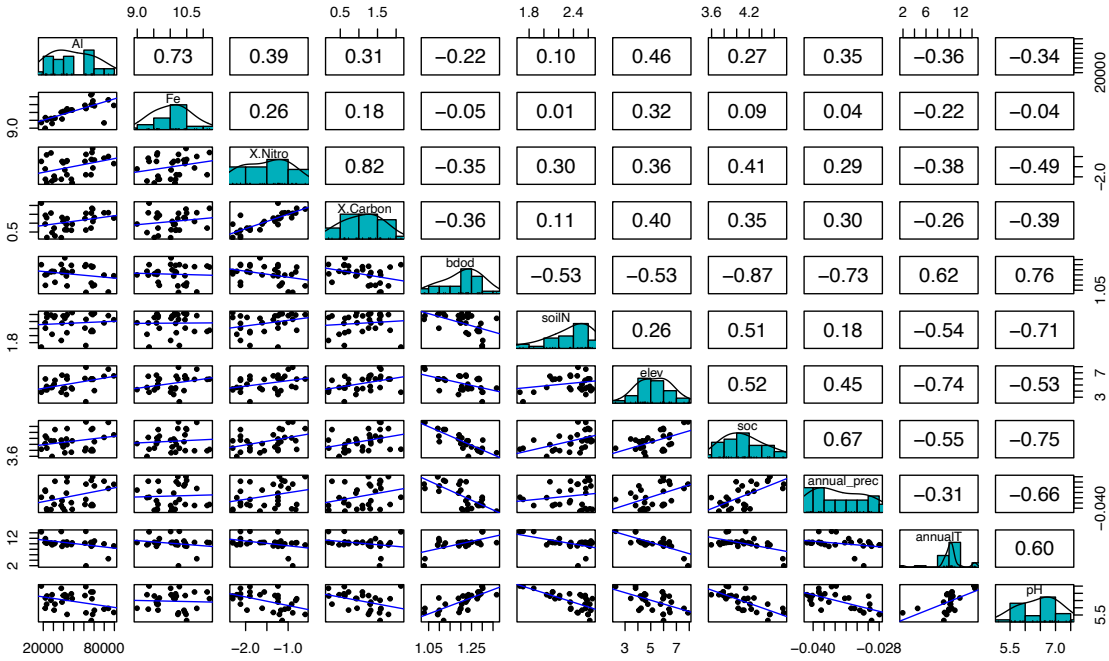


Figure S2. Standard validation plots for the selected GLMM model. Plots were automatically generated using the function *check_model* in the R packages performance on the model create with lme4 package.

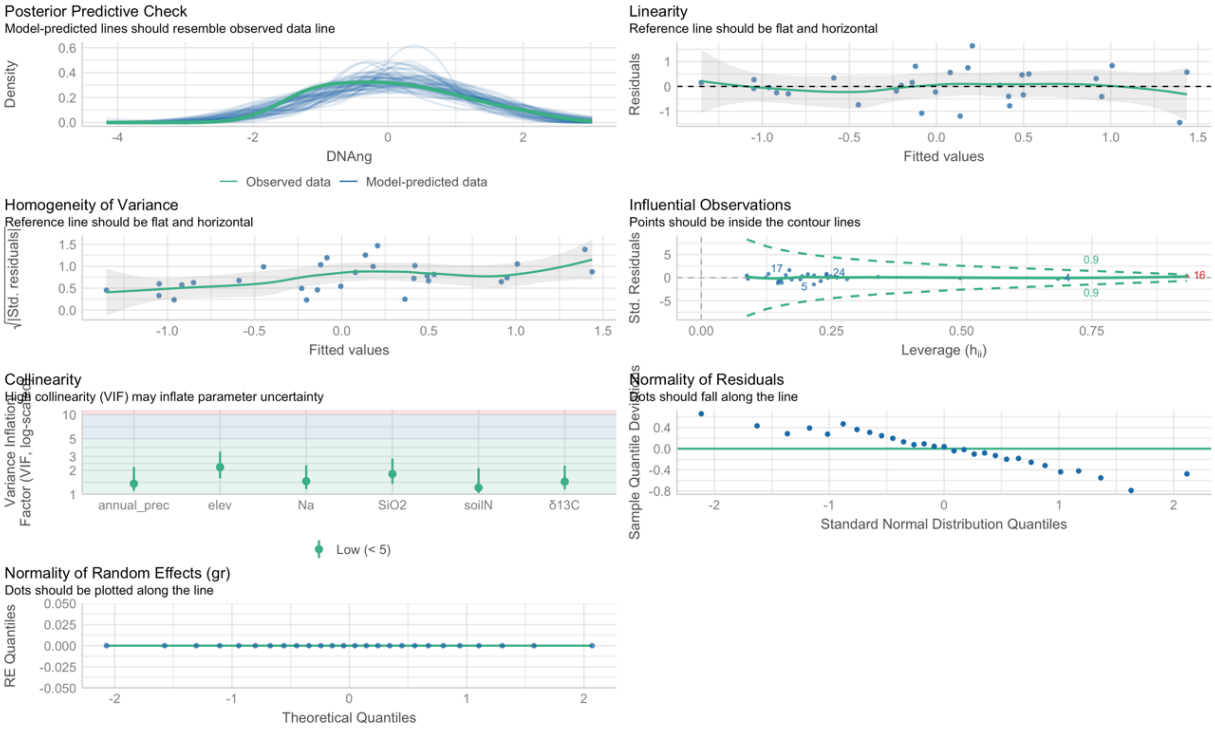


Figure S3. Pie charts of the percentages of herbarium specimens with roots and soil (red), with only roots (green), and without both soil and roots (stems and leaves without roots and soil, blue). Plant species are grouped by genera (only genus *Triticum* encompassed different species).

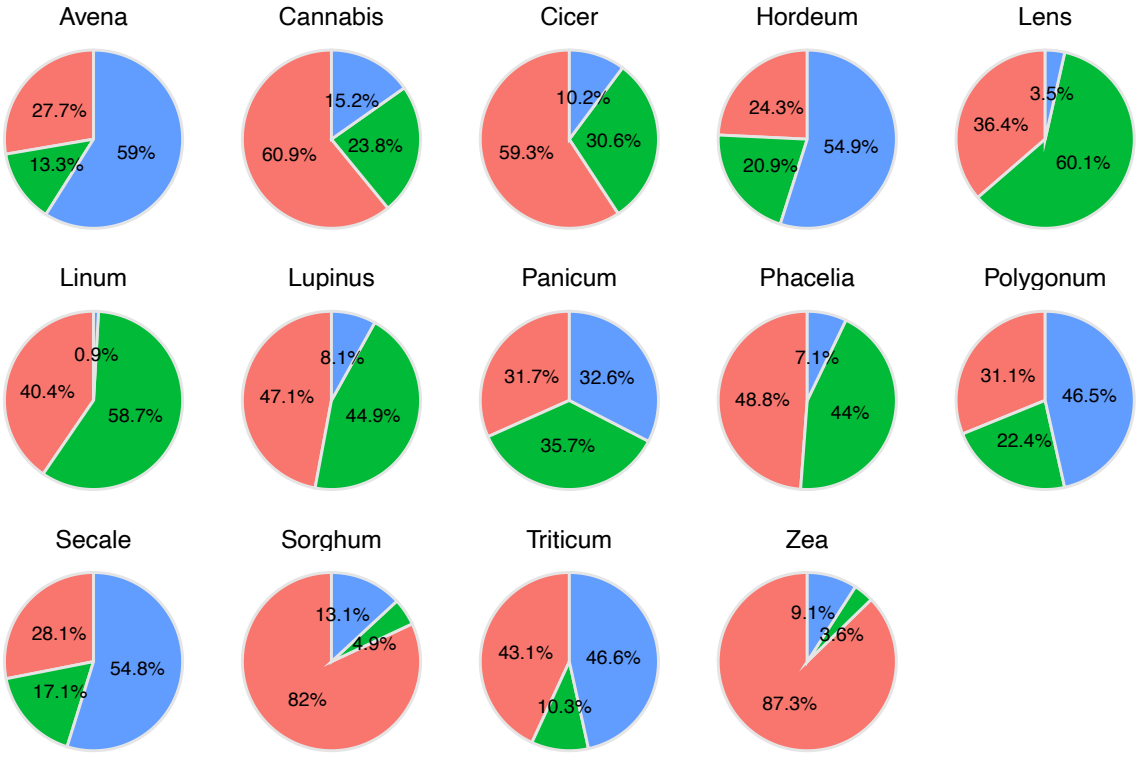


Figure S4 Sampling locations and dates of the 30 *Triticum aestivum* and *Hordeum vulgare* associated soil herbarium sample from which DNA was extracted. A. Map showing the geographical origin of the 30 soil samples in France. **B**.. Bar plot giving the sampling dates of the 30 soil samples from the 1820s to the 1970s.

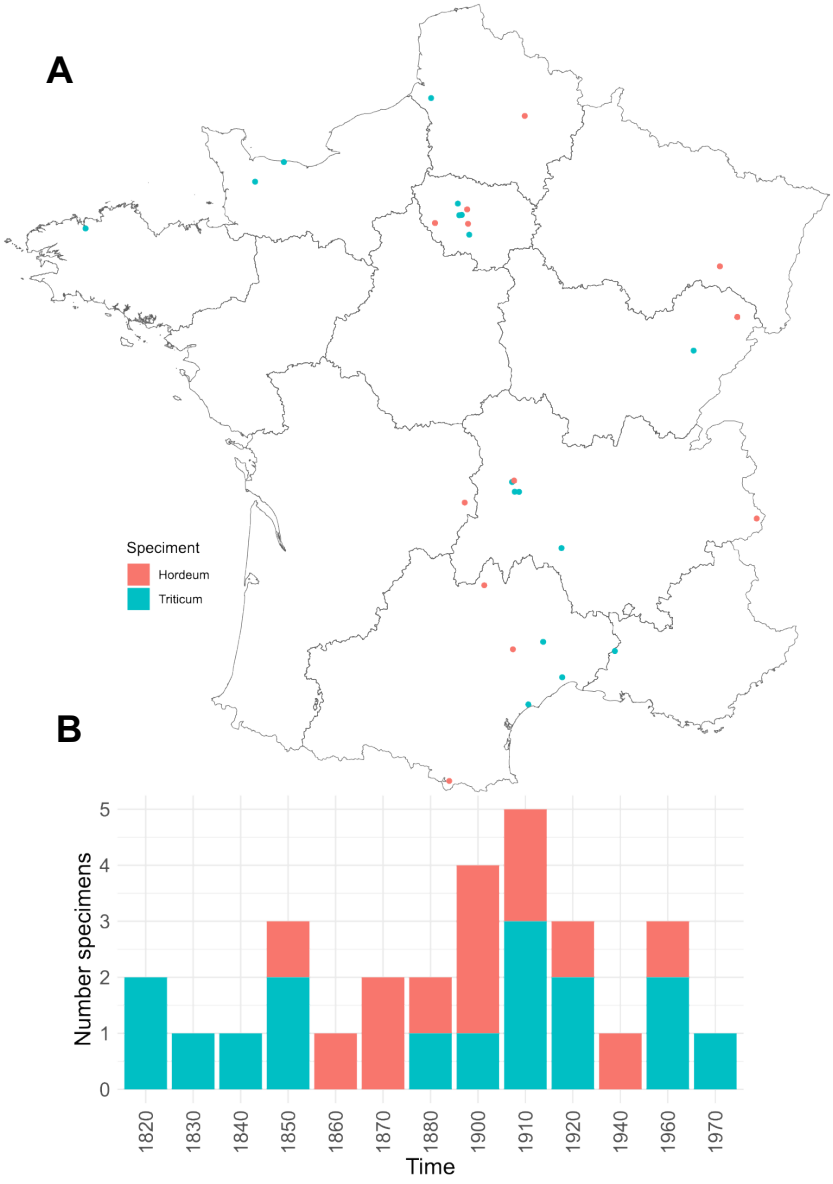


Figure S5. Characteristics of the selected 30 *Triticum aestivum* and *Hordeum vulgare*-associated soil herbarium samples from which DNA was extracted. The plot shows in the background (in grey) the violin plots of all French municipal averages of the values taken into consideration (annual mean temperature, annual precipitation. Cation exchange capacity, elevation, organic carbon density, pH, soil organic carbon content, soil total nitrogen); in the foreground are given the actual values of the variables (dots, estimated from the GPS position of the collection sites) for the 30 herbarium specimens (red, *Hordeum*, blue, *Triticum*) selected for DNA extraction.

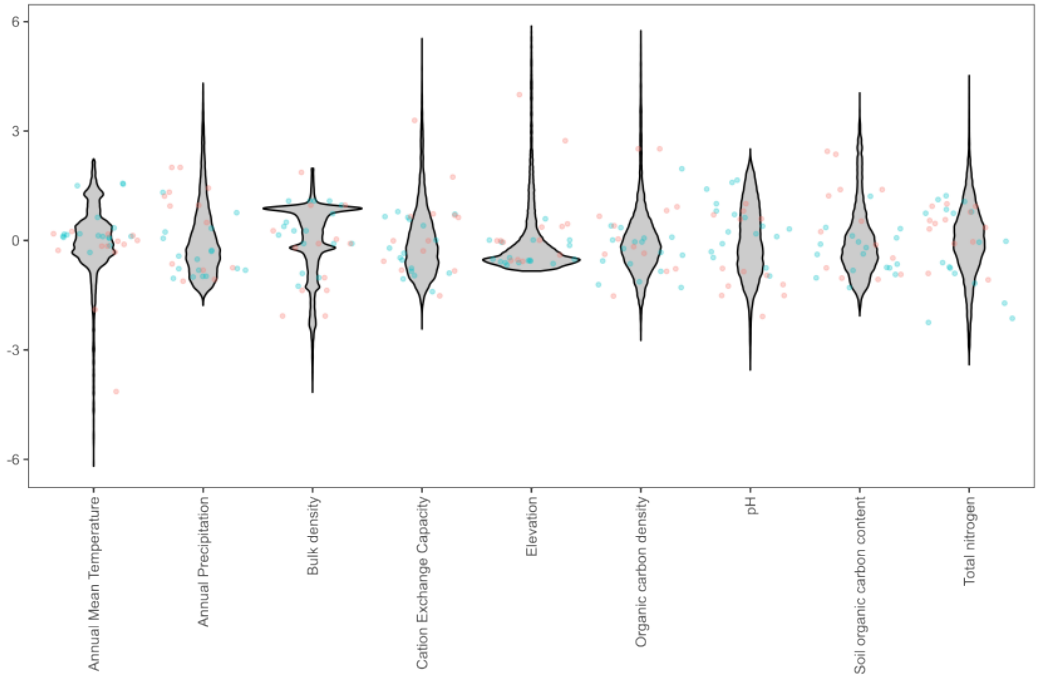


Figure S6. PCA analysis of herbarium soils parameters and their relative contribution to the total amount of extractable DNA (ng/50mg of soil) from soil samples. The analysis considered the parameters measured experimentally (Organic Carbon % (X.Carbon), Nitrogen % (X.Nitrogen), $\delta^{13}C$, $\delta^{15}N$, Na, Mg, Al, P, K, Ca, Mn, Zn, Cd, and SiO_2 contents) as well as those extracted from the SoilGrid and WorldClim database using the GPS coordinates of herbarium collection sites (pH, total nitrogen, Soil Organic Carbon (soc), Cation Exchange Capacity (cec), bulk density (bdod), annual temperature (annualT) and annual precipitation (annual_prec)). In addition to these parameters, the sample collection date (year) was also included.

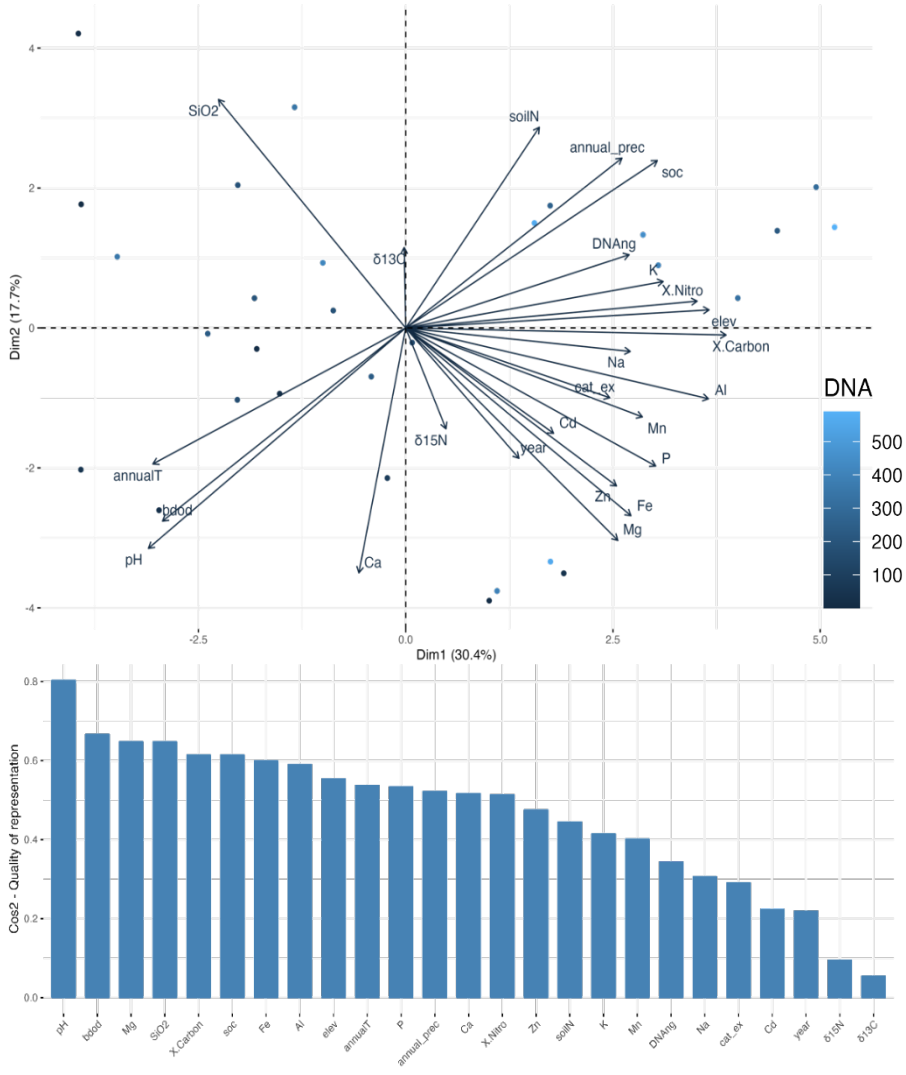


Table S1. Characteristics of the 30 plant specimens selected for soil DNA extraction. The Recolnat accession No. allow consulting the characteristics of the sample on the Recolnat web site (https://explore.recolnat.org/search/botanique/type=index).

sample name	Species	Collection date	Recolnat accession number	Total volume of DNA (ng/50mg)	French commune
1	Hordeum vulgare	1909	LY0665455	229.0	Saint-Quentin
2	Hordeum vulgare	1908	LY0666127	393.7	Clermont-Ferrand
3	Hordeum vulgare	1856	P03626872	159.9	Le Kremlin-Bicêtre
4	Hordeum vulgare	1919	SLA004450	14.0	Saint-Jean-et-Saint-Paul
5	Hordeum vulgare	1871	P03658192	214.7	Chagey
6	Hordeum vulgare	1898	P03629916-15	322.5	Epinal
7	Hordeum vulgare	1969	P01969577	350.4	Grigny
8	Hordeum vulgare	1911	P02637839	299.2	Bonneval-sur-Arc
9	Hordeum vulgare	1869	P03658191	253.9	Chagey
10	Hordeum vulgare	1921	P02638268	589.5	Llo
11	Hordeum vulgare	1875	MPU1385062	183.8	Rambouillet
12	Hordeum vulgare	1941	MPU742924	448.2	Florentin-la-Capelle
13	Hordeum vulgare	1888	MPU267705	397.8	Ussel
A	Triticum aestivum	1900	LY0664463	10.5	Suresnes
B	Triticum aestivum	1926	MPU091810	20.7	Portiragnes
C	Triticum aestivum	1911	CLF140821	40.3	Montpellier
D	Triticum aestivum	1856	MPU146493	85.1	Aulas
E	Triticum aestivum	1960	CLF349322	558.8	Tallende
F	Triticum aestivum	1881	CLF349329	33.4	Corent
G	Triticum aestivum	1835	P03239048	411.9	Arromanche-les-Bains
H	Triticum aestivum	1964	CLF140818	14.4	Clermont-Ferrand
I	Triticum aestivum	1821	P06768667	17.5	Saint-Lô
L	Triticum aestivum	1851	P02375291	9.3	Béhen
M	Triticum aestivum	1855	P03364631	43.8	Barbentane
N	Triticum aestivum	1918	P03364524	559.3	Morlaix
O	Triticum aestivum	1910	CLF140823	181.0	Verrières-le-Buisson
P	Triticum aestivum	1828	LYJB052767	340.9	Besançon
Q	Triticum aestivum	1840	P03257193	196.0	Ballancourt-sur-Essonne
R	Triticum aestivum	1920	CLF140808	228.6	Igny
S	Triticum aestivum	1977	MPU1193998	150.0	Saint-Laurent-d'Aigouze

Table S2. Bioclimatic values associated to the 30 plant herbarium specimens from which soils were collected for DNA extraction. These parameters were used to construct GLMM models to investigate which one of them affected extracted DNA. Values were extracted from the SoilGrid and Worldclim databases using the GPS coordinates of the collection sites using the geodata package in R.

sample_name	Genus	Region	bulk_density (kg/dm³)	nitrogen(g/kg)	elevation (m)	soc (g/kg)	annual_precipitation(mm)	annual_mean_temperature (°)	cation_exchange (cmol(c)/kg)	pH
1	Hordeum	Nord	1.29	5.51	92.10	41.94	698.93	9.71	17.70	6.97
2	Hordeum	Sud-est	1.29	4.49	371.35	44.05	660.67	10.51	22.87	7.09
3	Hordeum	Ile_de_France	1.37	6.20	89.59	41.17	593.60	10.40	23.06	6.90
4	Hordeum	Sud-ovest	1.22	4.34	627.63	54.17	1120.36	10.67	23.08	6.80
5	Hordeum	Centro	1.09	6.74	394.53	92.52	1282.53	9.57	20.78	5.54
6	Hordeum	Nord	1.12	6.87	403.14	89.78	1160.80	9.10	22.79	5.12
7	Hordeum	Ile_de_France	1.21	6.33	66.52	48.12	603.98	10.55	19.48	6.82
8	Hordeum	Sud-est	1.03	6.06	2716.89	115.65	1055.98	-1.95	32.17	5.65
9	Hordeum	Centro	1.09	6.74	394.53	92.52	1282.53	9.57	20.78	5.54
10	Hordeum	Sud-ovest	1.03	6.59	1985.30	113.95	950.89	4.54	26.70	5.96
11	Hordeum	Ile_de_France	1.23	6.33	168.59	61.25	675.38	10.02	15.22	5.84
12	Hordeum	Sud-ovest	1.20	5.57	620.16	68.90	1135.80	9.96	17.63	5.90
13	Hordeum	Sud-ovest	1.19	6.01	661.93	75.07	1051.48	9.23	18.56	5.68
A	Triticum	Ile_de_France	1.23	7.05	78.07	61.45	613.72	10.34	19.44	6.63
B	Triticum	Sud-ovest	1.30	2.95	9.59	47.87	746.69	14.55	18.68	7.47
C	Triticum	Sud-ovest	1.30	2.81	45.55	46.38	781.44	14.36	17.41	7.51
D	Triticum	Sud-ovest	1.10	4.88	399.97	71.96	891.51	11.88	20.64	6.68
E	Triticum	Sud-est	1.28	4.90	430.34	44.22	674.04	10.24	23.01	6.85
F	Triticum	Sud-est	1.28	4.63	389.35	46.09	674.44	10.40	22.88	6.63
G	Triticum	Nord	1.19	6.59	31.29	65.33	854.71	10.40	21.82	6.65
H	Triticum	Sud-est	1.28	4.49	371.35	44.05	660.67	10.51	22.87	7.09
I	Triticum	Sud-est	1.20	5.59	69.64	56.01	916.05	10.47	16.85	5.97
L	Triticum	Nord	1.24	4.15	100.00	36.43	775.05	9.61	17.60	6.91
M	Triticum	Sud-est	1.30	3.47	45.66	62.97	660.48	14.48	18.89	7.35
N	Triticum	Nord	1.10	6.95	59.09	59.67	1010.74	11.01	15.62	5.83
O	Triticum	Ile_de_France	1.20	6.87	118.33	67.58	619.28	10.22	17.87	6.25
P	Triticum	Centro	1.13	6.55	320.19	89.49	1134.19	10.27	23.35	6.25
Q	Triticum	Ile_de_France	1.25	6.55	83.22	50.04	623.17	10.55	19.83	6.85
R	Triticum	Ile_de_France	1.23	6.68	98.04	61.62	611.43	10.33	17.16	6.52
S	Triticum	Sud-ovest	1.22	5.56	742.69	55.68	727.07	9.07	22.61	6.39

Table S3 Chemical and physical characteristics associated to the soil collected on 30 selected plant herbarium specimens for DNA extraction. These parameters were used to construct GLMM models to investigate which one of them affected extracted DNA. Values were obtained on soil samples (50 mg) using ICP-MS and X-ray Fluorescence methods.

sample_name	DNA_volume(ng/50mg soil)	Na (ppm)	Mg (ppm)	Al (ppm)	P (ppm)	K (ppm)	Ca (ppm)	Mn (ppm)	Fe (ppm)
1	229.0	9753.2	4015	47698	574	15572	17451	501	25932
2	393.7	26209.2	23085	67023	2627	18319	63859	1090	67464
3	159.9	8021.9	4077	41246	1126	15999	17805	516	22657
4	14.0	1150.8	9420	43322	914	19100	175940	266	26505
5	214.7	15493.0	7099	68644	2793	33687	3696	1458	35561
6	322.5	6358.1	34814	67364	900	32098	25914	1268	36215
7	350.4	3776.1	1560	23582	429	10159	11380	445	13620
8	299.2	19280.8	12476	62106	1122	21522	8695	806	28307
9	253.9	4036.2	4322	43489	1523	15848	49734	1195	27752
10	589.5	17529.6	13176	84101	856	26011	10379	655	49246
11	183.8	7479.7	2149	30762	648	11780	3626	301	15207
12	448.2	27597.7	3210	79873	1112	41202	4651	465	11342
13	397.8	9560.6	12545	89449	1262	36717	4100	864	34476
A	10.5	2151.7	10398	48505	464	24400	27152	167	25536
B	20.7	4994.8	6336	61898	561	21873	38261	530	29073
C	40.3	3641.0	3192	18109	980	7244	192388	291	11625
D	85.1	3066.2	102355	22155	1025	8460	162698	864	15846
E	558.8	9028.3	11715	61096	1134	19381	101474	915	34745
F	33.4	12925.4	21098	67969	1943	9957	45504	1233	72438
G	411.9	4133.7	5306	42807	532	17561	37442	486	24479
H	14.4	17719.3	15216	68908	2204	19140	58747	1042	45987
I	17.5	4900.6	1050	21965	463	9317	2583	131	8055
L	9.3	8837.5	1503	28711	388	14125	3064	1066	10905
M	43.8	4379.0	7617	37796	577	11679	162635	639	22316
N	559.3	8154.0	5027	70353	1163	21537	4458	1100	33481
O	181.0	1968.1	6416	23323	2363	10906	54269	505	14763
P	340.9	6238.0	1916	31568	470	15793	13363	443	15313
Q	196.0	4501.7	3869	39711	720	16903	22965	489	23408
R	228.6	5344.5	7763	27633	1530	10732	94569	632	14938
S	150.0	10082.3	4904	38532	916	17001	88542	404	14522

sample_name	Zn (ppm)	Cd (ppm)	Pb (ppm)	weight herbarium soil (g)	SiO2 (g/100g of soil)	%nitrogen	δ15N (‰vsAir)	%carbon	δ13C(‰vsPDB)
1	59	0.22	16	0.056	37.59	0.122	3.794	1.822	-26.436
2	101	0.24	44	0.0521	23.93	0.151	9.623	2.058	-25.613
3	63	0.20	30	0.052	27.16	0.190	7.344	1.872	-26.264
4	49	0.10	14	0.0523	19.78	0.154	3.743	1.865	-26.366
5	104	0.22	86	0.0531	28.75	0.551	6.698	6.020	-26.620
6	105	0.26	46	0.0508	22.37	0.325	6.767	NA	NA
7	50	0.19	33	0.0512	32.81	0.101	3.142	1.207	-26.548
8	99	0.24	17	0.0507	33.05	0.406	6.046	4.741	-25.390
9	70	0.17	38	0.0529	32.40	0.304	5.746	3.525	-26.873
10	110	0.20	20	0.053	16.03	0.372	0.511	4.906	-26.582
11	144	0.18	19	0.0517	36.57	0.142	2.678	2.059	-26.690
12	87	0.50	37	0.0523	29.86	0.337	4.390	4.883	-26.502
13	118	0.19	57	0.0433	29.68	0.265	6.644	3.614	-26.774
A	188	0.14	24	0.0542	33.06	0.095	6.421	NA	NA
B	101	0.31	44	0.054	30.97	0.154	4.698	1.836	-25.748
C	51	0.14	81	0.052	NA	0.292	8.931	NA	NA
D	54	0.26	88	0.05	8.42	0.118	4.838	NA	NA
E	292	0.36	32	0.0522	16.48	0.287	5.763	3.530	-25.901
F	111	0.29	21	0.0539	18.69	0.446	7.657	6.302	-25.120
G	51	0.20	42	0.0506	33.34	0.285	8.505	2.507	-25.887
H	168	0.25	146	0.055	24.72	0.210	8.889	2.993	-25.262
I	20	0.07	16	0.0532	51.96	0.203	6.889	2.594	-26.819
L	29	0.21	25	0.0532	47.79	0.109	3.271	1.603	-25.595
M	63	0.13	16	0.0541	25.18	0.117	5.305	1.200	-25.860
N	104	0.13	72	0.0515	35.49	0.311	7.255	3.725	-26.997
O	91	0.15	1197	0.0535	37.43	0.400	8.368	3.691	-25.859
P	44	0.21	18	0.0505	41.90	0.124	5.831	1.323	-26.275
Q	59	0.18	22	0.0494	37.38	0.270	6.372	3.010	-25.819
R	93	0.22	119	0.0525	27.20	0.444	7.769	4.812	-25.893
S	48	0.39	17	0.0506	32.58	0.293	2.866	3.748	-26.783

Chapter 5. Ancient microbiomes as mirrored by DNA extracted from century-old herbarium plants and associated soil

Gianluca Grasso[1,2,3], Régis Debruyne[4,*], Martino Adamo[1], Olivier Rué [5], Franck Lejzerowicz[6], Lucie Bittner[2,7,#], Valeria Bianciotto[3,#], Roland Marmeisse[2,3,#]

1. Department of Life Science and Systems Biology (DBIOS), Università degli Studi di Torino, 10125 Torino, Italy.

2. Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université Des Antilles, 75005 Paris, France.

3. Institute for Sustainable Plant Protection (IPSP), National Research Council (CNR), SS, Viale Mattioli 25, 10125, Torino, Italy.

4. Muséum National d'Histoire Naturelle, Archéozoologie, Archéobotanique: Sociétés, Pratiques et Environnements (AASPE: UMR 7209 CNRS-MNHN), CP 56, 43 rue Buffon, F-75005 Paris, France.

5. Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France.

6. Section for Aquatic Biology and Toxicology, University of Oslo, Oslo, Norway.

7. Institut Universitaire de France, Paris, France.

Introduction

The contribution of the soil microbiome to soil and plant health is widely recognized (Banerjee & van der Heijden, 2022; Bardgett & Van Der Putten, 2014). A current challenge is to evaluate how ongoing global changes are affecting the diversity of these microbial communities, with potential consequences on their contribution to ecosystem processes. This challenge is essentially addressed in short-term studies that compare microbial communities along environmental gradients or in manipulated mesocosms or field experiments. From these studies, emerge a number of common signatures (e.g., decrease in host-specificity, increase in abundance of rare microbes, pathogens, and hypermutators), but also of differences in the response of soil microbiomes to environmental constraints (e.g., shifts in pH, temperature, and humidity values; Berg & Cernava, 2022; Zhou et al., 2020). These short-term studies, which rely on high-throughput sequencing techniques to evaluate the microbial diversity, have almost all been carried out in the last 15 years in the 21st century (Zhou et al. 2020). These studies therefore monitor short-term responses to environmental changes of soil microbial communities that have already experienced several decades of global changes. This is particularly relevant in the case of arable soils, which in different parts of the world have been managed since the late 1940s-1960s by intensive farming practices. These practices broke completely with the former ones, which did not depend on the use of chemical fertilisers and pesticides to grow high-yielding crop varieties.

In this context, we can ask what have already been the long-term consequences on soil microbiomes of intensive farming practices. Indeed, several ecological processes occur at a pace incompatible with short-term monitoring. This is the case of biological invasions and diversity impoverishment, two facets of the global biodiversity crisis. This crisis has been put forth and quantified by studies targeting animals and plants (Nic Lughadha et al., 2020; Pyšek et al., 2020; Rull, 2021). In the case of microorganisms, invasions - but rarely extinctions - are documented essentially for symptom-producing plant or animal pathogens, for a few mushroom-producing fungal soil saprotrophs as well as for bacterial or fungal root symbionts (Desprez-Loustau et al., 2007; Thakur et al., 2019; Wang et al., 2023). No data are available for the great majority of "elusive" soil microorganisms. Agricultural practices such as the release of microbial inoculants (Liu et al., 2022) or mass movements of entire microbiomes through long-distance trade and travel (Zhu et al., 2017) suggest however that widespread

invasion of soils ecosystems by alien microbial species and genetic elements is likely.

In order to evaluate long-term changes in the composition of soil microbial communities, it is necessary to access ancient soil samples and to implement appropriate protocols to reveal the composition of their original microbiomes. Regarding the later prerequisite, molecular paleomicrobiology, based on the analysis of degraded ancient DNA (aDNA) extracted from historical samples, is certainly the most relevant approach (Grasso et al., 2024a). Paleomicrobiology allows for a diachronic approach of the evolution of microbial diversity. It allows documenting taxonomic and functional diversity of microorganisms in ancient, up to tens of thousands years old items preserved in favourable environments such as caves or permafrost. It has been implemented for samples as diverse as archaeological artefacts such as dental calculus, desiccated or mineralized faeces, thus giving access to ancient human/animal oral and intestinal microbiomes (Brealey et al., 2021; Fellows Yates et al., 2021; Maixner et al., 2021; Quagliariello et al., 2022). These data have been used to infer the diet and local environmental conditions to which were exposed the corresponding hosts. Regarding "non-host" associated environmental samples, the analysis of aDNA preserved in stratified marine or freshwater sediments recapitulates the temporal dynamics of aquatic microbial communities and how they responded to past environmental changes (Barouillet et al., 2022; Keck et al., 2020; Siano et al., 2021).

Regarding archived soil samples, although large collections do exist in different countries, the corresponding samples have, for most of them, been excavated in the last 50 years. Older referenced samples are rare and limited to very few collection sites, as exemplified by the Rothamsted long-term agronomic experiments that started in the 1840s and for which soil samples were continuously collected (Poulton, 2011). We recently identified herbaria as potential repositories of (rhizospheric) soils that were left attached to the roots of many herbaceous plants preserved in these collections (Bianciotto et al., 2022). These collections that may encompass more than 400 millions of plant samples globally (https://sweetgum.nybg.org/science/wp-content/uploads/2023/11/The_Worlds_Herbaria_2022.pdf) have been mostly established in the 19th and first half of the 20th century, a period that predates the so-called "great acceleration" in global changes (Steffen et al., 2015). For the many plant species, herbaria can contain hundreds of different samples collected

over the entire species' distribution areas and several decades. These collections can thus be used to trace plant adaptation to global changes through time, from anatomical, physiological and genomic point of views (DeLeo et al., 2020; Denney & Anderson, 2020; Lang et al., 2022, Burbano & Gutaker, 2023).

Initial studies have targeted plants in herbaria to study their original microbiomes using metabarcoding (Daru et al., 2018; Heberling & Burke, 2019) or shotgun sequencing (Bieker et al., 2020; Weiß et al., 2016) of plant-extracted DNA. The objective of the present study is to extend the approach to herbarium plant-associated desiccated rhizospheric soils in order to access the past diversity of the soil microbiome and ultimately infer its time-dependent response to environmental changes. This experimental framework needs first to be validated from both experimental and analytical points of view. It is necessary to evaluate the ancient nature of the extracted DNA and to demonstrate that the deduced biological data are in line in terms of biodiversity and community assembly rules with the data commonly reported for extant soil microbiomes. For this validation, we chose four cultivated and taxonomically diverse plants; *Lactuca sativa* (lettuce) and three cereals, oat, rye and durum wheat. Most of the studied plants were grown in a common garden in the suburbs of Paris (France) and collected in the first years of the 20th century; that is about 120 years before the start of the study.

Materials and methods

Herbarium specimens

All plant specimens (**Table S1, Fig. 1.A**), four or five per plant species, came from the Bonaparte's collection detained by the herbarium of the University of Lyon (herbarium code, LY), France and can be visualised on the RECOLNAT website (https://explore.recolnat.org). Specimens belonged to the species *Avena sativa*, *Lactuca sativa*, *Triticum durum* and *Secale cereale*. Most plants (16 out of 19) were originally cultivated in the same garden, in "Saint-Cloud" in the suburbs of Paris. The remaining three specimens were collected in three other sites in France (**Table S1**). Several plants grown in Saint-Cloud are thought to be genetically distinct, as they belonged to a specific variety mentioned on the herbarium's labels. Plants were collected over a short period of time, between the years 1903 and 1907.

For each specimen, about 20 fine root fragments - ca. 0.5 cm-long - were collected using sterile tweezers and scalpel blades. Soil particles aggregated around the roots (between 10 and 100 mg per specimen) were detached from the roots and collected separately. For two of the *L. sativa* plants, no rhizospheric soil was present around the roots. For three specimens (belonging to *L. sativa, S. cereale* and *T. durum*), leaf fragments were also sampled. The collected material was transferred into sterile tubes and stored at room temperature until DNA extraction.

DNA extraction manipulation and sequencing

DNA extraction and its manipulation were performed in the dedicated facility of the platform "Paléogénomique et génétique moléculaire" at the "Musée de l'Homme" in Paris (https://www.ecoanthropologie.fr/fr/paleogenomique-et-genetique-moleculaire-6206). The platform is equipped with clean rooms under positive air pressure and UV-sterilised daily to prevent contamination from the outside environment. Strict separation is maintained between spaces dedicated to sample preparation, DNA extraction, control, amplification, and sequencing library preparation. Laboratory coveralls were worn for all experiments and specific cleaning procedures of equipment and spaces were implemented (2.6% bleach and/or 20 min UV (256 nm)-cross-linking).

Total DNA was extracted following the method of Murchie et al. (2021) optimised for ancient sediment samples, with minor modifications (Grasso et al.

2024b). Three control extraction and library preparation blank samples were systematically performed to mitigate the risk of in-laboratory contamination. End-repaired DNA extracts were used to prepare Illumina libraries using a double-indexing PCR with a variable number of 9-14 cycles (Kircher et al., 2012) per individual library (**Table S4**). All library blanks were submitted to 14 cycles of amplification but failed to produce any detectable product but low proportions of adapter dimers. Amplified libraries were characterised by fluorometric quantification (Qubit II, Invitrogen) and capillary electrophoresis (Labchip GX Touch, Perkin-Elmer) and pooled in equimolar amounts. The multiplexed libraries were sequenced on a Illumina NovaSeq 6000 instrument (SP flowcell, 800 M paired-end 2x50 bp reads) at the iGenSeq sequencing facility ("Institut du Cerveau", https://igenseq.institutducerveau-icm.org/).

Sequence filtering, quality control and read mapping

Reads were subjected to quality control, adapters trimming and merging using FastQC v.0.12.0 (Andrews, 2010) and fastp v.0.23.1 (Chen et al., 2018). For the trimming, default commands were used except for two options to account for the short length of the reads: *--length_required* option was set to 25 nucleotides, and *–overlap_len_require* was lowered from 30 to 10 nucleotides. Reads were then mapped using BWA v. 0.7.17 (Li & Durbin, 2009) to the corresponding plant reference genomes (**Table S2**) to filter out the plant sequences. The pool of unmapped reads that includes the host plant microbiome (*e.g.* eubacteria, archaea, fungi, micro-eukaryotes, viruses) was extracted using Samtools v. 1.14 (Danecek et al., 2021). The unmapped reads were further mapped to the h28 human reference genome (accession No. GCF_000001405.26) to exclude any human contamination.

Ancient DNA validation and comparison to modern microbiomes

The rate of cytosine deamination at the extremities of the reads was assessed using MapDamage2 (Jónsson et al., 2013) based on the frequencies of transitions and transversions in alignments to reference genomes. MapDamage was separately run on the plant reads (**Table S2**) and on reads affiliated to the 10 most abundant bacterial genera identified in the microbiome dataset (cf. details on the classification in the next section) using reference genomes from NCBI (April 2023; **Table S3**). For both plants and bacteria, the distribution of metagenomic

fragment lengths was compared for the different groups using ANOVA testing in R 4.2.2 (R Core Team, 2021).

Microbiome reads were compared with modern metagenomes from the Earth Microbiome Project (EMP, Thompson et al., 2017) using the Qiita platform (Gonzalez et al., 2018). Qiita offers access to a supercomputer to upload, search, and analyse metagenomic samples jointly using Qiime2 (v.2023.2, Bolyen et al., 2019). Herbarium metagenomes were uploaded to Qiita, merged with the EMP samples, aligned using Bowtie2 (Langmead et al., 2009) to the WoLr2 database ("Web of Life" release 2), and classified using Woltka (v.0.1.4, Zhu et al., 2022) to obtain feature tables across the same WoL reference genomes. These tables were used to measure alpha and beta diversity into Qiita, using Qiime2. To explore differences in composition and diversity among various microbiomes, the core_metrics function of Qiime2 v.2023.2.0 (Bolyen et al., 2019) was employed.

Analysis of microbial community composition and diversity

The microbiome reads were taxonomically annotated using Kaiju v1.9.0 (Menzel et al., 2016), which confronts translated sequences to a database of bacterial, archeal and protists (including fungi) proteomes (nr_euk data bank release 2021-03, https://github.com/bioinformatics-centre/kaiju). Abundance tables of phylum and genus level were obtained and further explored in R. The Shannon alpha diversity metric was calculated for each sample using the *fossil* package v.0.4.0 (Vavrek MJ, 2011) and the vegan package v.2.6-4 (Oksanen et al., 2022). Statistical differences were assessed using the Kruskal-Wallis test. For beta diversity, a Bray-Curtis dissimilarity matrix was calculated using *vegan* based on feature tables created using *fossil* and Hellinger transformed using *labdsv* 2.1-0 (Roberts, 2023). This matrix was explored with Principal Coordinates Analysis (PCoA, Anderson & Willis, 2003) using *ape* v.5.7-1 (Paradis et al., 2004). Permutational analysis of variance (PERMANOVA) were conducted using the 'adonis' function of *vegan*, scrutinizing community structure across taxa (genus and phylum levels), collection sites, and sample type (soil *versus* root). Diversity measures for Bacteria-Archaea and Eukarya were repeated both at phylum and genus levels. DESeq2 v.1.26.0 (Love et al., 2014) was run to detect differentially-abundant genera across sample types. In addition to these analyses, the sequences were also analysed with Simka, a *de novo* tool for comparative metagenomics that represents each data set as a k-mer spectrum based on read similarity. It is

thus possible to analyse a priori, without any taxonomic affiliation, the influence the different studied factors (site, matrix, and plant host) on the total sequence datasets of each sample (Benoit et al., 2020).

Results

DNA extraction, sequencing and sequence analysis

DNA was successfully extracted from all 39 herbarium leaf, root, and soil samples representing four plant species collected in four geographic sites in France *ca* 120 years ago. Total amounts of extracted DNA ranged from 64 to 398 ng for root/leaf samples and from 787 to 1203 ng for soil samples. Electrophoretic separation of amplified library DNA fragments produced, in all cases, a peak lower than 200 bp, corresponding to a majority of DNA inserts shorter than 70 bp, thus justifying a NovaSeq 6000 run of only 50 cycles. After demultiplexing, quality control, paired-end reads merging and elimination of reads shorter than 25 bp, we obtained an average of 21.8 ± 7.7 (min-max: 8.5-40.3) millions of reads per sample (**Figure S1.A**). On average, 67.5 % (50.5-85.4) of the merged fragments were shorter than 50 bp. The mean DNA fragment length for the different sample types were 40.7, 42.1, and 48.0 bp for the leaf, root, and rhizosphere samples, respectively (**Figure S1.B**). The average read duplication rate was only on average 3 % per library. Average percentage of reads mapping onto plant genomes varied between 87.5 % (86.8-88.6), 62 % (24.09-85.9), and 23.54 % (3.8-49.8) for leaves, roots, and soil samples respectively (**Table S4**). On average 0.4 % of reads mapped to the human genome with only three samples that had percentages above 0.7 % (**Table S4**).

Size distributions were evaluated separately for the sequences that mapped onto a plant genome, and for the microbiome reads that mapped onto genomes of the 10 most represented bacterial genera *Streptomyces, Nocardioides, Pseudomonas, Flavobacterium, Bradyrhizobium, Lentzea, Microbacterium, Sphingomonas, Massilia, Devosia.* These 10 genera belong to three different phyla (Proteobacteria, Actinobacteria and Bacteroidetes). Sample average length of merged reads (*i.e.* the size of extracted DNA fragments) never exceeded 50 bp and statistically supported (P<0.001, ANOVA) differences between sample types (**Figure 1.C**) were observed. On average, plant DNA fragments were ca. 10 nucleotides longer than microbiome DNA fragments. The length of the microbiome fragments also differed between bacterial genera. Fragments attributed to *Flavobacterium* were *ca.* 15 nucleotides longer than fragments attributed to the nine other genera. Of the ten microbial taxa, *Flavobacterium* had the lowest genome GC-content (around 35 % versus 61-72 % for the other 9 bacterial genera, **Fig. 1.C**). The sample type (root, rhizospheric soil or leaf) from which DNA was extracted had a significant effect on DNA

fragment size. Except for sequences that mapped to *Streptomyces*, *Massilia* and *Lentzea*, DNA extracted from soil was less fragmented than plant DNA or than microbial DNA extracted from plant tissues (roots or leaves, $P<0.05$, ANOVA **Figure 1.B**).

Besides high fragmentation, another typical feature of aDNA is cytosine deamination. This damage preferentially occurs at DNA fragment overhangs and results in higher apparent C-to-T and G-to-A substitutions at the 5' and 3' ends of the molecules, respectively. Substitutions were quantified for the fragments mapping to host plant genomes or to the selected bacterial genera. In all samples and for all taxa, the rates of C-to-T and G-to-A substitutions were higher than background at the 5' and 3' terminal positions, respectively and displayed typical curves for ancient DNA libraries when progressing inward (illustrated in **Fig. 1** for *Triticum durum* and two representative bacterial genera). For plant sequences, substitution rates at the first position of the DNA fragments were in the range of 0.014-0.022; while it was slightly higher (0.022-0.032) in the case of bacterial sequences.

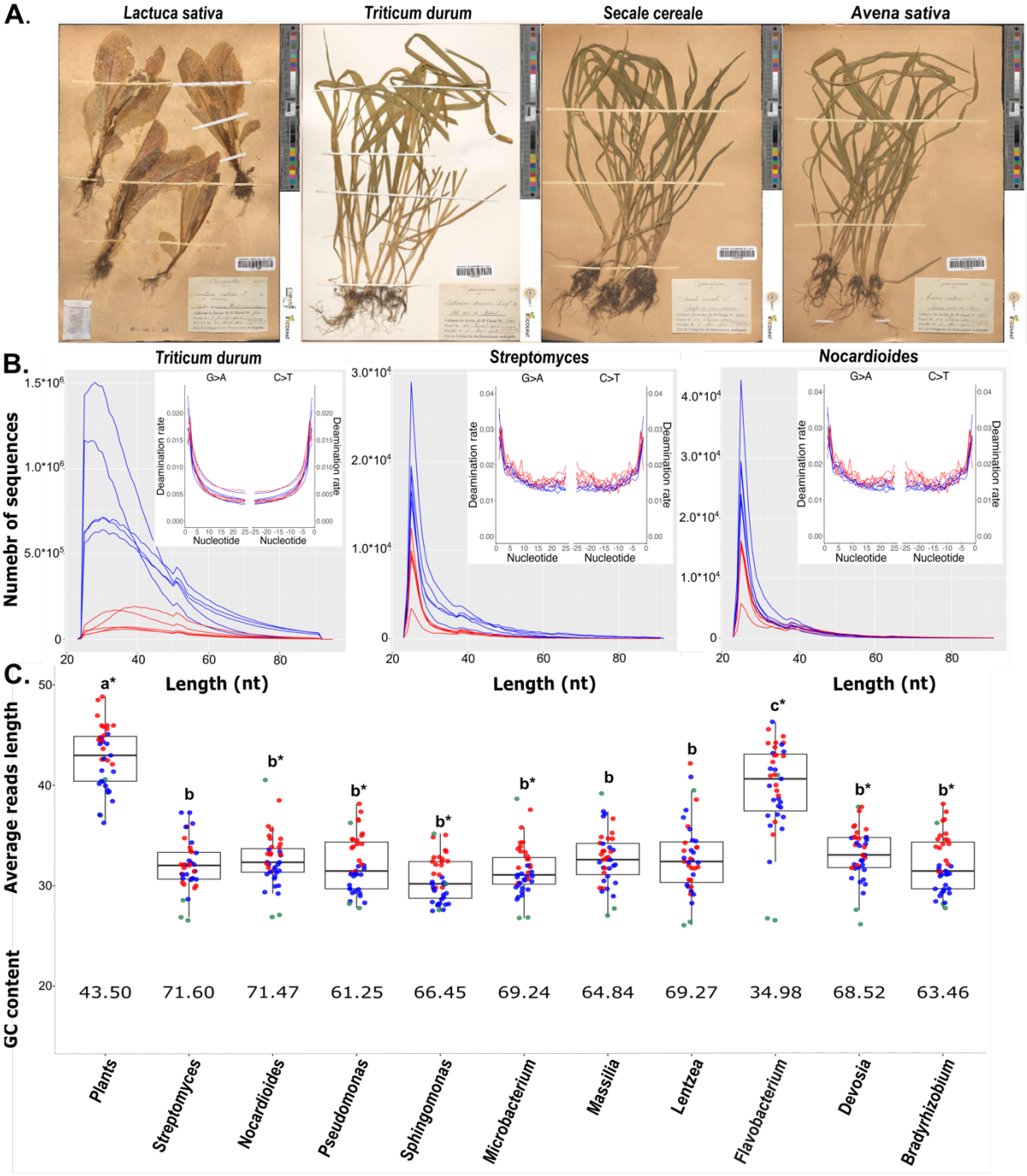

A.
Lactuca sativa
Triticum durum
Secale cereale
Avena sativa
B.
Triticum durum
Streptomyces
Nocardioides
Numebr of sequences
G>A
C>T
Deamination rate
Nucleotide
C.
Length (nt)
Average reads length
GC content
a*
b
b*
c*
43.50
71.60
71.47
61.25
66.45
69.24
64.84
69.27
34.98
68.52
63.46
Plants
Streptomyces
Nocardioides
Pseudomonas
Sphingomonas
Microbacterium
Massilia
Lentzea
Flavobacterium
Devosia
Bradyrhizobium

Figure 2. Analysis of DNA extracted from herbarium samples highlights typical signatures of ancient DNA and suggests plant/soil and species-dependent effects on patterns of degradation. A. an illustration of the herbarium plants with associated soil used for the experiments; numbers in parenthesis refer to their "RecolNat" accession numbers (https://explore.recolnat.org/search/botanique/type=index). B. Length distribution and deamination profiles of DNA fragments (merged reads) extracted from *Triticum durum* roots (blue curves) and associated soil (in red) affiliated to *T. durum*, *Streptomyces* and *Nocardioides*. Deamination profiles are illustrated for the first 25 nucleotides at the 5' end (G to A transitions) and 3' one (C to T transitions). C. Distribution (box plots) of the average lengths of the sequenced DNA fragments extracted from 39 leaf (green), root (blue) and soil-associated (red) herbarium specimens. Each box plot refers to the fragments assigned to either a plant genome (4 species), or one of the 10 most abundant bacterial genera. Different letters below the plots indicate statistically supported ($P<0.001$) differences between taxa; while asterisks indicate, for each of the taxa, significant ($P<0.05$) differences in the average length of the fragments extracted from soil and plant tissues.

Below ground herbaria microbial communities cluster with extant soil ones

Integration of herbarium microbiomes into the 817 metagenomes collected for EMP in highly contrasted environments, including soil and plant organs, was performed to evaluate if conservation-in herbaria altered the taxonomic diversity of their microbial communities, and if they were inadvertently contaminated by other communities. PCoA ordination of all metagenomes labelled by environment and project (EMP or herbarium) showed that soil and root herbarium samples grouped with metagenomes from extant non-saline soils and sediments. Herbarium-associated metagenomes were clearly distinct from metagenomes of animals, aquatic environments (fresh or saline waters, saline sediments), and plant surfaces (aerial surfaces of vascular plants or surfaces of algae, **Fig.2**). Two of the three herbarium leaf microbiomes and only one herbarium root microbiome did not tightly cluster with extant soil metagenomes.

Taxonomic annotation and alpha diversity

Using Kaiju, 85.2% of the microbiome sequences did not receive any taxonomic affiliation. At the phylum level, 95.82 % of the annotated sequences were assigned to Bacteria, 2.74 % to Eukarya, 1.36 % to Archaea and 0.08 % to viruses. A genus-level assignation was obtained for 53.23 % of the bacterial sequences and for 83.16 % of the Eukaryotic ones. The five most represented bacterial phyla were, in the following order, Proteobacteria, Actinobacteria, Bacterioidetes, Planctomycetes and Acidobacteria. Eukarya was mainly

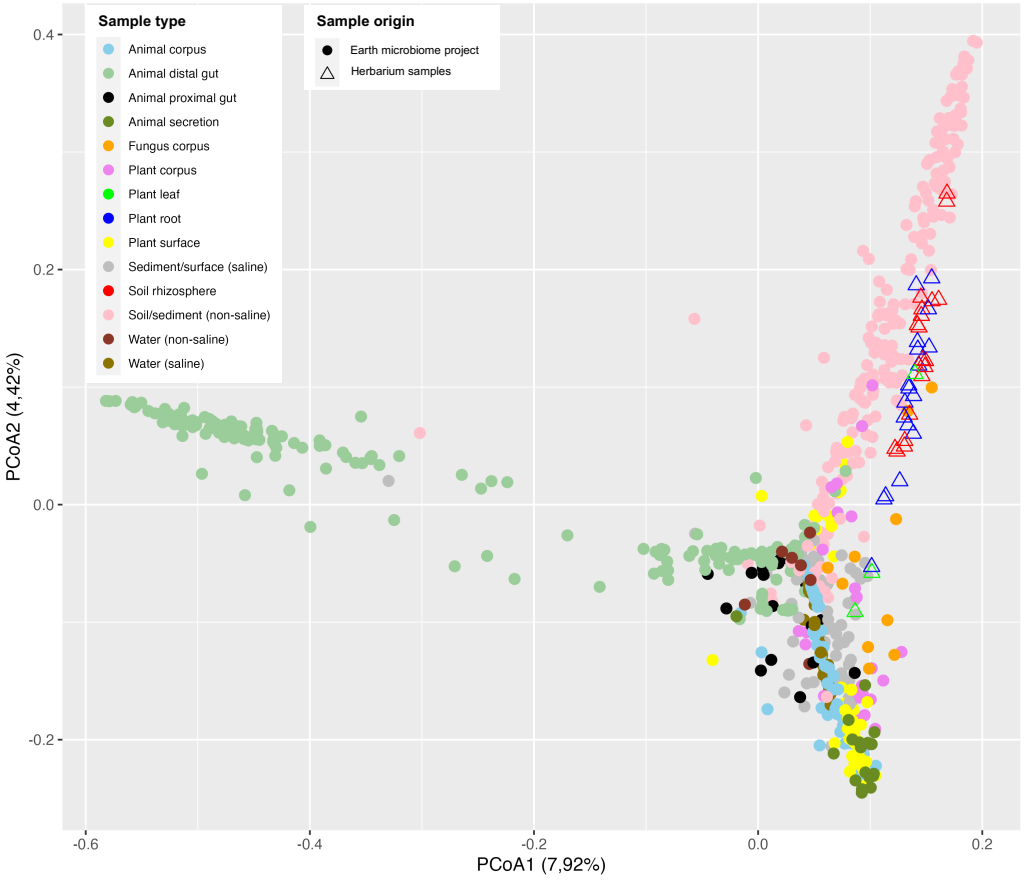


Figure 2. Herbarium associated soil and root microbial communities cluster with extant soil communities. Herbarium soil and root microbial communities (empty triangles) cluster close to extant soil communities (pink dots) in a global analysis of microbial communities sampled in highly contrasted environments worldwide (dots). Principal Coordinates Analysis (PCoA) ordination is based on Bray Curtis indices computed using the Qiita management platform. Note that two of the three herbarium samples located under the 0.0 value of the PCoA2 axis are two of the three communities from leaves (green triangles).

represented by fungi (essentially Ascomycota, Basidiomycota and Mucoromycota that encompassed symbiotic Glomeromycotina). More than one percent of the sequences were also affiliated to other unicellular eukaryotes (*e.g*. Oomycota, Ciliophora, or Apicomplexa, **Fig.3.A**). Since DNA was extracted from minute amounts of material, no Metazoa genome alignment was attempted.

Shannon alpha diversity indices calculated at the genus level were globally higher for soil compared to root samples ($P<0.01$; Kruskal-Wallis). This global difference holds true for each individual herbarium plant sample (
3.B). At the genus level, in the case of bacteria from the collection site of Saint-Cloud, a large proportion of the genera were identified in both soil and roots (53 and 71 % for genera represented by at least 0.3 or 0.05 % of the total number of reads, respectively). Genera identified only in soil outnumbered those identified only in roots (39 *versus* 25 or 104 *versus* 42 for genera represented by at least 0.3 or 0.05 % of the total number of reads, respectively). A similar distribution pattern between roots and soil was also observed for eukaryotic taxa (**Fig. S2**).

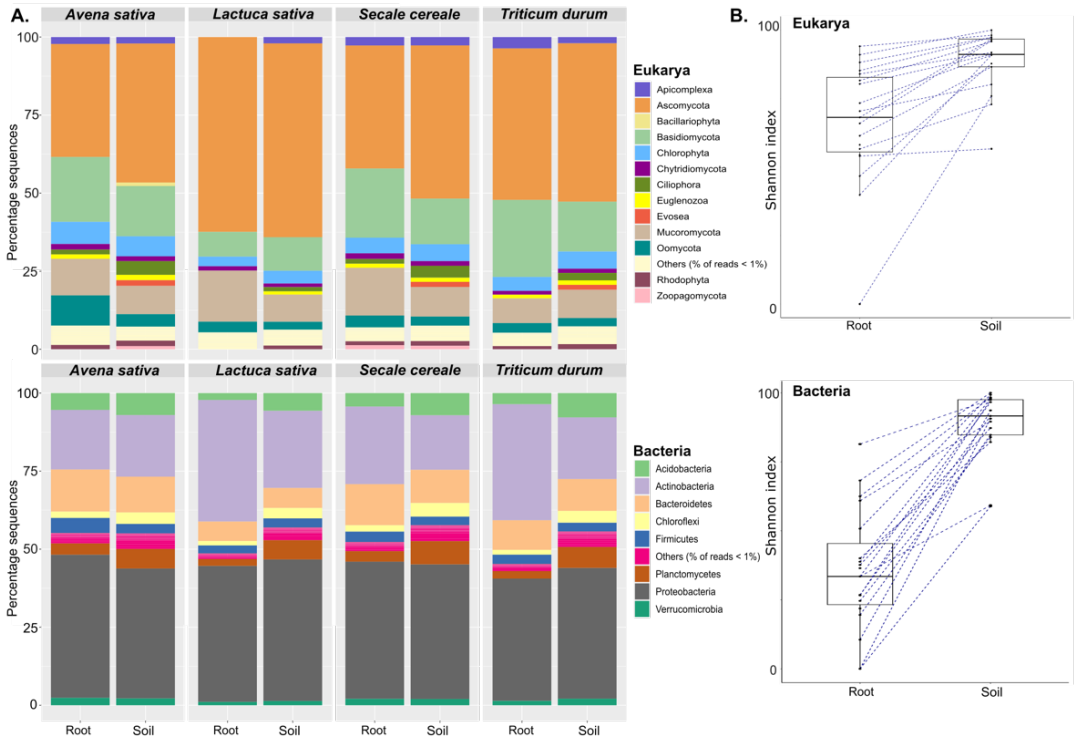


Figure 3. Taxonomic (alpha) diversity of herbarium-associated soil and root microbial communities. A. Taxonomic composition (phylum level) of eukaryotic (top) and bacterial (bottom) microbial communities associated with the roots or the soils of the four studied plant species. Each bar represents the mean value of the different replicates of samples collected in Saint Cloud garden in Paris (excluding samples collected from other collection sites). Phyla represented by less than 1% of the total number of annotated sequences were pooled in a single category (Others). B. Distribution of normalised Shannon biodiversity indices (calculated at the genus level) of root and soil eukaryotic and bacterial communities. Dotted lines connect the values of the root and soil communities of each individual herbarium plant sample. *Lactuca* herbarium plant samples from which no soil samples could be collected were excluded from this analysis.

Beta diversity analyses

The factors that affected the microbial community structure were evaluated at both phylum and genus levels. At the phylum level, sample type (root versus soil) had the strongest effect, over plant host species, collection site, and interaction of these terms ($P < 0.001$, PERMANOVA). Clearly, microbial compositions from different sample types separate on the NMDS ordination ($k=3$, stress < 0.05, **Fig 4.A and Table S5**). The phyla Planctomycetes, Chloroflexi, Acidobacteria, Ciliophora and, to a lesser extent, Verrucomicrobia, that were more abundant in soil contributed most to the differences between soil and root communities (**Fig. 4.A**). Root communities tended to be slightly richer in fungal (Basidiomycota, Mucoromycota) and Actinobacteria sequences than soil ones (**Fig 3.A**).

At the genus level, a similar sample type effect for both Bacteria and Eukarya on community composition was observed ($P<0.001$, PERMANOVA, **Table S5**). At that taxonomic level, a plant species effect was also detected ($P<0.001$, PERMANOVA) in eukaryotic and bacterial root communities (**Fig. S3 and Table S6**). PCoA ordination of the samples also illustrated the separation between the soil and root samples, and to some extent the separation between root samples from different plant species (**Fig. S3**). While the first axis of the PCoA ordination highlighted the differences between root and soil communities, the second axis separated soil and root communities of plants collected in Saint-Cloud from soil and root communities of plants collected in the other three sites. (**Fig 4**). Regarding plant species, for bacteria, both axes contributed to the separation of *Lactuca* and *Secale* root samples, while for Eukarya, both axes contributed to the separation of *Lactuca* samples from all other species' root samples (**Fig. S3**).

In the case of microbial genera, we used DESeq2 to identify taxa whose abundance differed between root and rhizospheric soil samples (**Fig.5**). Clear taxonomic differences between soil or root-enriched bacterial taxa were observed. While 8 of the 19 root-enriched taxa belonged to Actinomycetes (42 %), none belonged to Planctomycetes or Acidobacteria, which were identified in soil (6 of the 15 genera, *i.e.* 40 %). In the case of Eukarya, although only 12 genera were retained as being significantly enriched in either soil (4 genera) or roots (8 genera), these two categories sharply differed with respect to their taxonomy, but also to their known patterns of interaction to plants. Among the taxa more abundant in roots, at least 6 are either obligate root-symbionts (one genus affiliated to the Glomeromycota) or known plant-pathogens, or endophytes

affiliated to the Oomycota (in the Peronosporales) or the Fungi (either Basidiomycota or Ascomycota). None of these categories related to plant-microbiome interactions were identified among the four taxa found as more abundant in soil. The two fungal genera (*Podospora* and *Madurella*) are known soil-saprotrophs and the two other taxa are free-living unicellular eukaryotes (here Amoebozoa and Alveolata).

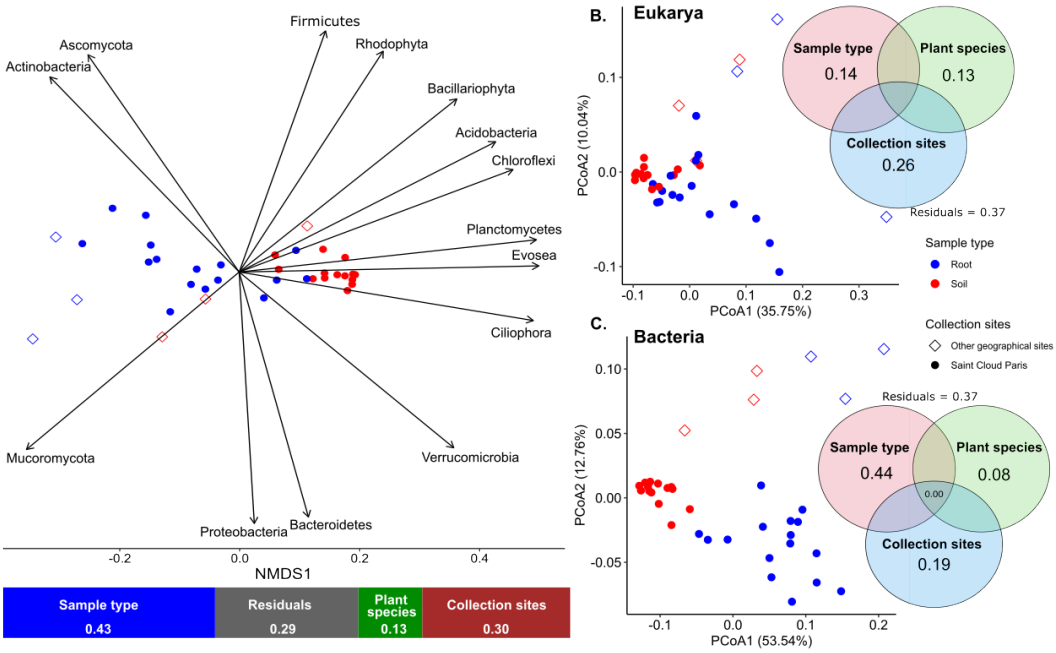


Figure 4. Sampling collection site and sample type (soil *versus* root) affect herbarium-associated communities at both the phylum and genus levels. A. At the phylum level Non-Metric multidimensional Scaling (NMDS, k=3, stress < 0.05) ordination based on Bray-Curtis distances illustrates the distribution of microbial communities according to collection site and sample type (red: soil; blue: root; filled dots: garden of Saint-Cloud, empty triangles: other sites). Explanatory variables (phyla represented by more than 1 % of the reads assigned) that significantly ($P < 0.05$) contributed to the separation of the communities along the two axes are indicated. The lower bar gives the percentage of total variance explained by each variable. B. and C. Principal Coordinates Analysis ordinations based on Bray-Curtis distances illustrate the distribution of eukaryotic (B) and bacterial (C) communities according to collection site and sample type. Venn diagrams give the percentage of total variance explained by each of the variables and their interaction (only explained variance > 0.01 are reported).

Finally, herbarium root and soil microbiomes were also analysed using with simka (Benoit et al., 2020), a method independent of taxonomic assignation. The results (**Fig. 6**) also highlight a significant influence of the matrix (soil, root, and leaf) and host plant species on the microbiomes , leading to the same conclusions as in the case of the taxonomically-assigned sequences (Fig. 5). Separation between plant species can nevertheless be questioned by the presence of highly repetitive plant sequences (such as transposons or non-coding regions) not adequately removed by mapping on plant genomes. It is interesting to observe how different sample types (root and soil) are significantly separated in the analysis (**Table. S7**).

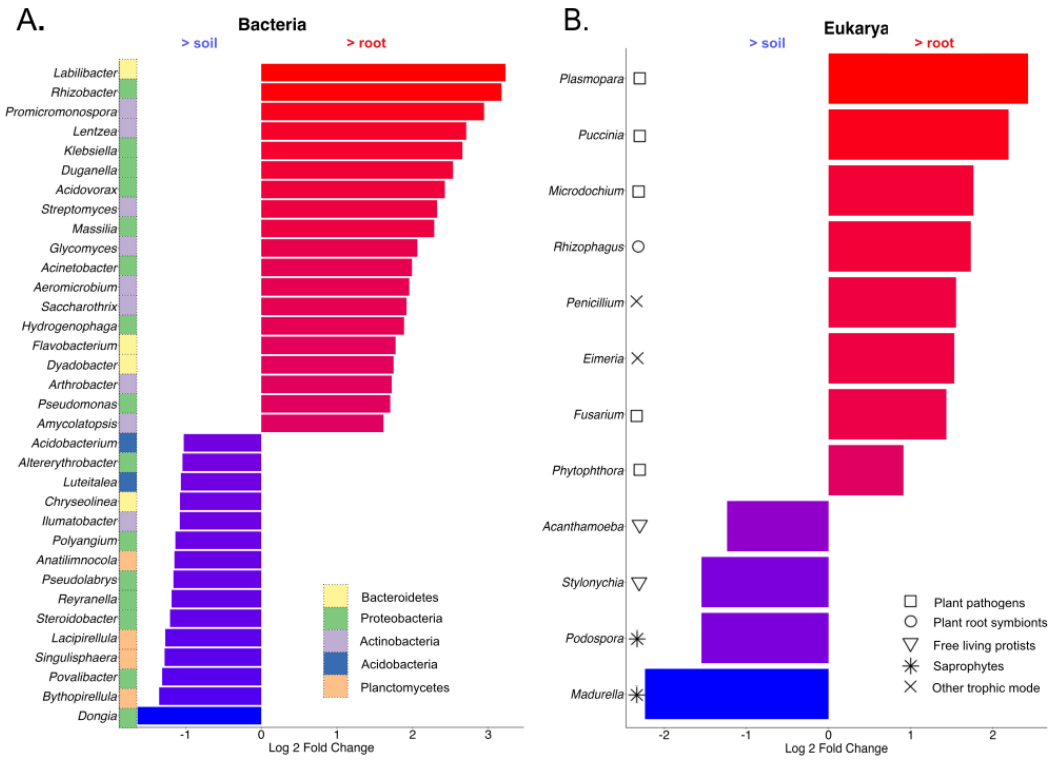


Figure 5. Bacterial and eukaryotic genera that showed differences in abundance between herbarium soil and root samples. Analysis using the DeSeq software allowed identifying taxa that displayed a difference in read abundance between the global root and soil sequence datasets above a threshold of 1.5/-1 for Bacteria and 0.9/-1 for Eukarya (Log_2 scale). A. differentially abundant bacterial taxa (P<0.001). Taxonomic annotation (phylum level) highlights differences between taxa displaying a higher abundance in soil (dominated by Planctomycetes) and those more abundant in roots (dominated by Proteobacteria and Actinomycetes). B. differentially abundant eukaryotic taxa (P<0.05). Functional annotation highlights differences between taxa displaying a higher abundance in soil (unicellular protists and fungal saprotrophs) and those more abundant in roots (dominated by plant pathogens and symbionts). Since reads assigned to eukarya were far less abundant compared to those attributed to bacteria, the analysis was restricted to taxa with more than 50.000 and 5000 reads in the global dataset for bacteria and eukarya respectively

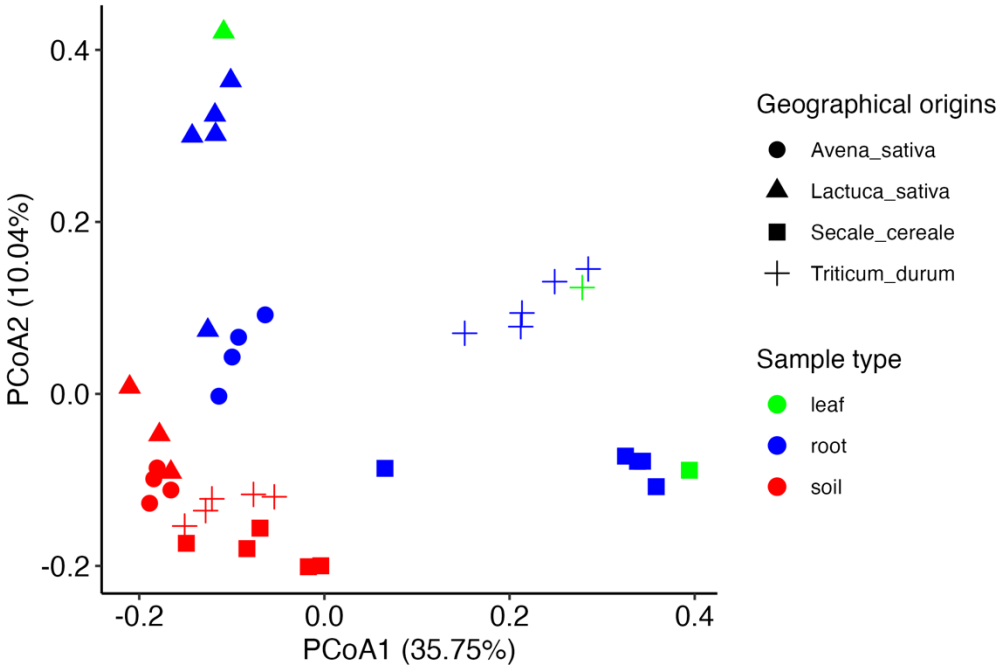


Figure 6. Principal Coordinate analysis of simka output based on k-mer similarity between each herbarium sequence datasets. The impact of both matrix (soil, root-leaf) and host plant species were significant ($P < 0.001$, **Table S7**)

Discussion

Analysis of DNA extracted from herbarium soil and plant roots validates our hypothesis that this material is suitable to evaluate the past diversity of soil microorganisms, for the period under scrutiny (from at least the beginning of the 20th century). The extraction protocol, optimised for sediment samples (Murchie et al., 2021), was effective in recovering DNA present in limited samples of both soil and roots from four collection sites and associated to four different plant species. Enzymatic manipulation of the DNA (with proteinase K during extraction and various enzymes and polymerases during library preparation) was not prevented by the historical treatments by trichloro(nitro)methane (chloropicrin) and Mercury(II)-chloride used for preservation and disinfection (Staats et al., 2011; Herbarium of Lyon, personal communication).

DNA fragment length was on average below 50 bp, which is in the range of what has been reported for herbarium plants (Särkinen et al., 2012; Staats et al., 2011) and one of the typical characteristics of aDNA (Pääbo et al., 2004). However, the average length of the extracted DNA appears to be significantly affected upon the extraction sample type and slightly variable among the taxa considered. In most cases, including for the plant DNA, the fragments extracted from soil are longer than those extracted from plant tissues, either roots or leaves. This may result from a protective effect of the soil environment (salt content, presence of minerals adsorbing DNA, fast desiccation) over DNA fragmentation by depurination, a process known to be affected *in vitro* by different physico-chemical parameters such as low pH, presence of water, metal ions or polycations (An et al., 2014, 2017). Regarding the "microbial taxon effect" on DNA degradation, it seems to be linked to genome GC-content; the higher the GC-content, the higher the degradation (lower average fragment size). This phenomenon could result from a greater depurination sensitivity for guanine compared to adenine residues, as suggested for the human genome (Kunkel, 1984). Alternatively we cannot exclude that it could result from a technical bias that occurs during the preparation or the sequencing of the libraries. Illumina sequencing is known to be negatively affected by uneven GC-content (Chen et al. 2013; Benjamini & Speed 2012), yet no significant effect of this parameter on the produced read length has ever been documented.

Beside the DNA fragmentation, we also highlighted cytosine deamination at both ends of the DNA fragments from either plant or microbial origin. Both these features testify of the ancient origin of the extracted DNA and presumably

of the low contamination level by modern organisms, an expected result for herbarium samples that have been desiccated rapidly after collection and kept dry since then. In addition, no positive amplification was obtained from any of the extraction and library control blank, which excludes contamination of the samples during the laboratory experiments. Fluctuations in hygrometry during conservation that may have allowed occasional secondary microbial colonisation of the soil and roots cannot however be fully ruled out. Post-mortem colonisation of herbarium plants was indeed hypothesised by Bieker et al. (2020) to explain the presence of several fungal taxa, specific to leaves of *Ambrosia* and *Arabidopsis* herbarium plants that were not observed in modern, fresh leaf samples.

Confrontation of–microbial DNA root and soil datasets to numerous microbiomes from highly contrasted environments showed that none of our samples behaved as outliers in the PCoA ordination but almost all clustered with modern soil samples and not with samples from other environments. This is a further indication that secondary contamination of the samples was minimal and that preservation in herbaria did not grossly alter the relative proportions of the microorganisms that characterised the original communities. This procedure is crucial to detect possible microbial contamination during the preservation period and will need to be performed on each new herbarium specimen before microbiome investigations.

We also observed that root microbial communities were always less diverse than their cognate soil ones and that taxa (genera) identified in roots were, for a majority of them, also present in soils. These observations are also in line with what is known of root microbial communities that are mostly recruited horizontally from the surrounding local pool of soil microorganisms (Edwards et al., 2015; Xun et al., 2021).

In terms of community structure, herbarium soil and root microbial communities also follow well-described assembly rules. In the literature, the sampling site, with local specific climate and physico-chemical features, is always the main factor affecting soil community composition (Peiffer et al. 2013). In the present study, although our sampling design, with regard to this factor, was highly unbalanced; all samples from the “Saint-Cloud” site clearly grouped together in a PCoA ordination (genus level) and were clearly separated from the few samples collected in three other sites.

In plant microbiology, the second most important factor controlling microbial community composition is the sample type, in our case soil and root (Xiong et al., 2021). Once again, we observed a clear sample type effect on community composition at both phylum and genus levels. Visualisation of this effect was certainly facilitated by the analysis of herbarium plants that, for most of them, were grown in a single and restricted site (a garden in the town of Saint-Cloud) thus minimising the confounding collection site effect. Furthermore, the very clear sample type effect was observed in spite of a potentially significant level of cross-contamination between soil and root herbarium samples. Herbarium root samples cannot indeed be totally freed of adhering soil particles, as would be the case for fresh ones that are usually cleaned by sonication. At the phylum level, Acidobacteria and Chloroflexi, were consistently more abundant in soils as already observed in other studies that compared different plant species grown in different sites (Tkacz et al., 2020; Xiong et al., 2021). Soils were also globally enriched in Planctomycetes and Acidobacteria, that predominated among genera found enriched in soil. Phylum Planctomycetes is known to actively participate in ammonium oxidation, a key step of the nitrogen cycle, notably in soils. At the genus level, functional roles assigned to eukaryotic taxa enriched in soil or roots were in line with their known associations to plants, with mostly pathogenic and mutualistic taxa among root-enriched taxa that were completely absent among soil-enriched ones that were free-living unicellular eukaryotes and saprotrophic fungi. It is worth emphasising that the root-enriched mutualistic taxon belongs to the Glomeromycota that need a living plant host for their growth and therefore cannot represent post-harvest external secondary contaminants of the samples.

Finally, a third, albeit less pronounced effect, is that of the plant species in the case of root communities. Although communities from all plant species could not be separated, it is worth mentioning that for both Bacteria and Eukarya, *Lactuca*-associated communities differ from communities associated with other plants that belonged to the phylogenetically distant Poaceae family.

Altogether, these different observations support the hypothesis that herbarium material retains the original characteristics of the plant and soil-associated microbiome as it was when the plants were collected. The approach can thus be extended to a larger number of samples to evaluate how cultivated soil microbiomes have been historically affected by the evolution of farming practices. Visual inspection of digitalized herbaria indicates that for several

cultivated plant species, particularly cereals, numerous specimens have been preserved with associated rhizospheric soil. It is thus possible to analyse time series of different plants that cover a period of time that spans almost two centuries, the period during which most of the herbarium collections have been established.

Acknowledgements

Work on ancient plant-associated microbiota at the MNHN was supported by grant ATM2021 (HoloHerbier) and the Emergence program of Sorbonne Université (HerbaSoil). GG was supported by a PhD grant from the University of Turin and a mobility grant from the French-Italian University (program da Vinci 2022). VB was supported by the Short-Term Mobility (2023) program of the CNR. LB acknowledges the Institut Universitaire de France for her 5-year nomination as Junior Member (2020-2025). We thank the Herbarium of Université Lyon 1 and its curator Mélanie Thiebaut for granting access to material from its collections. We thank the Platform P2GM from the MNHN, for granting access to their ancient DNA facilities. We are grateful to the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi: 10.15454/1.5572390655343293E12) for providing advice and computing and storage resources.

Data availability

Sequence datasets analysed during the current study are available in the EMBI-ENA database under Bioproject PRJEB75398 (https://www.ebi.ac.uk/ena/browser/view/PRJEB75398).

References

An, R., Dong, P., Komiyama, M., Pan, X., & Liang, X. (2017). Inhibition of nonenzymatic depurination of nucleic acids by polycations. *FEBS Open Bio*, *7*(11), 1707. https://doi.org/10.1002/2211-5463.12308

An, R., Jia, Y., Wan, B., Zhang, Y., Dong, P., Li, J., & Liang, X. (2014). Non-enzymatic depurination of nucleic acids: factors and mechanisms. *PloS One*, *9*(12). https://doi.org/10.1371/JOURNAL.PONE.0115950

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Anderson, M. J., & Willis, T. J. (2003). CANONICAL ANALYSIS OF PRINCIPAL COORDINATES: A USEFUL METHOD OF CONSTRAINED ORDINATION FOR ECOLOGY. *Ecology*, *84*(2), 511–525. https://doi.org/10.1890/0012-9658(2003)084[0511:CAOPCA]2.0.CO;2

Banerjee, S., & van der Heijden, M. G. A. (2022). Soil microbiomes and one health. *Nature Reviews Microbiology 2022 21:1*, *21*(1), 6–20. https://doi.org/10.1038/s41579-022-00779-w

Bardgett, R. D., & Van Der Putten, W. H. (2014). Belowground biodiversity and ecosystem functioning. *Nature 2014 515:7528*, *515*(7528), 505–511. https://doi.org/10.1038/nature13855

Barouillet, C., Vasselon, V., Keck, F., Millet, L., Etienne, D., Galop, D., Rius, D., & Domaizon, I. (2022). Paleoreconstructions of ciliate communities reveal long-term ecological changes in temperate lakes. *Scientific Reports*, *12*(1), 1–12. https://doi.org/10.1038/s41598-022-12041-7

Benjamini Y, Speed TP. Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. 2012 May;40(10):e72. doi: 10.1093/nar/gks001. Epub 2012 Feb 9. PMID: 22323520; PMCID: PMC3378858.

Benoit, G., Mariadassou, M., Robin, S., Schbath, S., Peterlongo, P., & Lemaitre, C. (2020). SimkaMin: Fast and resource frugal de novo comparative metagenomics. Bioinformatics, 36(4), 1275–1276. https://doi.org/10.1093/bioinformatics/btz685

Berg, G., & Cernava, T. (2022). The plant microbiota signature of the Anthropocene as a challenge for microbiome research. *Microbiome*, *10*(1), 1–12. https://doi.org/10.1186/S40168-021-01224-5/FIGURES/1

Bianciotto, V., Selosse, M. A., Martos, F., & Marmeisse, R. (2022). Herbaria preserve plant microbiota responses to environmental changes. *Trends in Plant Science*, *27*(2), 120–123. https://doi.org/10.1016/J.TPLANTS.2021.11.012

Bieker, V. C., Sánchez Barreiro, F., Rasmussen, J. A., Brunier, M., Wales, N., & Martin, M. D. (2020). Metagenomic analysis of historical herbarium specimens reveals a postmortem microbial community. *Molecular Ecology Resources*, *20*(5), 1206–1219. https://doi.org/10.1111/1755-0998.13174

Bolyen, E., Ram Rideout, J., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Titus Brown, C., Callahan, B. J., Caraballo-Rodramp, M., Chase, J., … Gregory Caporaso, J. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* , *37*, 852–857. https://doi.org/10.1038/s41587-019-0190-3

Brealey, J. C., Leitão, H. G., Hofstede, T., Kalthoff, D. C., & Guschanski, K. (2021). The oral microbiota of wild bears in Sweden reflects the history of antibiotic use by humans. *Current Biology*, *31*(20), 4650-4658.e6. https://doi.org/10.1016/J.CUB.2021.08.010

Hernán A. Burbano, Rafal M. Gutaker ,Ancient DNA genomics and the renaissance of herbaria.Science382,59-63(2023).DOI:10.1126/science.adi1180

Chen YC, Liu T, Yu CH, Chiang TY, Hwang CC. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. PLoS One. 2013 Apr 29;8(4):e62856. doi: 10.1371/journal.pone.0062856. PMID: 23638157; PMCID: PMC3639258.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics (Oxford, England)*, *34*(17), i884–i890. https://doi.org/10.1093/BIOINFORMATICS/BTY560

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1–4. https://doi.org/10.1093/GIGASCIENCE/GIAB008

Daru, B. H., Bowman, E. A., Pfister, D. H., & Arnold, A. E. (2018). A novel proof of concept for capturing the diversity of endophytic fungi preserved in herbarium specimens. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *374*(1763). https://doi.org/10.1098/RSTB.2017.0395

DeLeo, V. L., Menge, D. N. L., Hanks, E. M., Juenger, T. E., & Lasky, J. R. (2020). Effects of two centuries of global environmental variation on phenology and physiology of Arabidopsis thaliana. *Global Change Biology*, *26*(2), 523–538. https://doi.org/10.1111/GCB.14880

Denney, D. A., & Anderson, J. T. (2020). Natural history collections document biological responses to climate change. *Global Change Biology*, *26*(2), 340–342. https://doi.org/10.1111/GCB.14922

Desprez-Loustau, M.-L., Cile Robin, C., Bué, M., Gis Courtecuisse, R., Garbaye, J., Dé Ric Suffert, F., Sache, I., & Rizzo, D. M. (2007). The fungal dimension of biological invasions. *TRENDS in Ecology and Evolution*, *22*(9), 472–480. https://doi.org/10.1016/j.tree.2007.04.005

Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N. K., Bhatnagar, S., Eisen, J. A., Sundaresan, V., & Jeffery, L. D. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(8), E911–E920. https://doi.org/10.1073/PNAS.1414592112/-/DCSUPPLEMENTAL/PNAS.1414592112.SD01.XLSX

Fellows Yates, J. A., Velsko, I. M., Aron, F., Posth, C., Hofman, C. A., Austin, R. M., Parker, C. E., Mann, A. E., Nagele, K., Arthur, K. W., Arthur, J. W., Bauer, C. C., Crevecoeur, I., Cupillard, C., Curtis, M. C., Dalen, L., Bonilla, M. D. Z., Fernandez-Lomana, J. C. D., Drucker, D. G., … Warinner, C. (2021). The evolution and changing ecology of the African hominid oral microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(20), e2021655118. https://doi.org/https://www.pnas.org/doi/abs/10.1073/pnas.2021655118

Gonzalez, A., Navas-Molina, J. A., Kosciolek, T., McDonald, D., Vázquez-Baeza, Y., Ackermann, G., DeReus, J., Janssen, S., Swafford, A. D., Orchanian, S. B., Sanders, J. G., Shorenstein, J., Holste, H., Petrus, S., Robbins-Pianka, A., Brislawn, C. J., Wang, M., Rideout, J. R., Bolyen, E., … Knight, R. (2018). Qiita: rapid, web-enabled microbiome meta-analysis. *Nature Methods*, *15*(10), 796–798. https://doi.org/10.1038/s41592-018-0141-9

Grasso, G., Bianciotto, V., & Marmeisse, R. (2024a). Paleomicrobiology: Tracking the past microbial life from single species to entire microbial communities. *Microbial Biotechnology*, *17*(1). https://doi.org/10.1111/1751-7915.14390

Grasso, G., Rotunno, S., Debruyne, R., Bittner, L., Miozzi, L., Marmeisse, R., & Bianciotto, V. (2024b). Identification of DNA Viruses in Ancient DNA from Herbarium Samples.

Methods in Molecular Biology, *2732*, 221–234. https://doi.org/10.1007/978-1-0716-3515-5_15/COVER

Heberling, J. M., & Burke, D. J. (2019). Utilizing herbarium specimens to quantify historical mycorrhizal communities. *Applications in Plant Sciences*, *7*(4). https://doi.org/10.1002/APS3.1223

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/BIOINFORMATICS/BTT193

Keck, F., Millet, L., Debroas, D., Etienne, D., Galop, D., Rius, D., & Domaizon, I. (2020). Assessing the response of micro-eukaryotic diversity to the Great Acceleration using lake sedimentary DNA. *Nature Communication*, *11*(1), 1–8. https://doi.org/10.1038/s41467-020-17682-8

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1). https://doi.org/10.1093/NAR/GKR771

Kunkel, T. A. (1984). Mutational specificity of depurination . *Proc. Natl. Acad. Sci. USA*, *81*, 1494–1498.

Lang, P. L. M., Erberich, J. M., Lopez, L., Weiß, C. L., Amador, G., & Fung, H. F. (2022). Century-long timelines of herbarium genomes predict plant stomatal response to climate change. *BioRxiv*, *8721*(Ll), 0–3. https://doi.org/10.1101/2022.10.23.513440

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), 1–10. https://doi.org/10.1186/GB-2009-10-3-R25/TABLES/5

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/BIOINFORMATICS/BTP324

Liu, X., Le Roux, X., & Salles, J. F. (2022). The legacy of microbial inoculants in agroecosystems and potential for tackling climate change challenges. *IScience*, *25*(3), 103821. https://doi.org/10.1016/J.ISCI.2022.103821

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9

Maixner, F., Sarhan, M. S., Huang, K. D., Tett, A., Schoenafinger, A., Zingale, S., Blanco-Míguez, A., Manghi, P., Cemper-Kiesslich, J., Rosendahl, W., Kusebauch, U., Morrone, S. R., Hoopmann, M. R., Rota-Stabelli, O., Rattei, T., Moritz, R. L., Oeggl, K., Segata, N., Zink, A., … Kowarik, K. (2021). Hallstatt miners consumed blue cheese and beer during the Iron Age and retained a non-Westernized gut microbiome until the Baroque period. *Current Biology*, *31*(23), 5149-5162.e6. https://doi.org/10.1016/j.cub.2021.09.031

Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications 2016 7:1*, *7*(1), 1–9. https://doi.org/10.1038/ncomms11257

Murchie, T. J., Kuch, M., Duggan, A. T., Ledger, M. L., Roche, K., Klunk, J., Karpinski, E., Hackenberger, D., Sadoway, T., Macphee, R., Froese, D., Poinar, H., & Corresponding, *. (2021). Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. *Quaternary Research*, *99*, 305–328. https://doi.org/10.1017/qua.2020.59

Nic Lughadha, E., Bachman, S. P., Leão, T. C. C., Forest, F., Halley, J. M., Moat, J., Acedo, C., Bacon, K. L., Brewer, R. F. A., Gâteblé, G., Gonçalves, S. C., Govaerts, R., Hollingsworth, P. M., Krisai-Greilhuber, I., de Lirio, E. J., Moore, P. G. P., Negrão, R., Onana, J. M.,

Rajaovelona, L. R., … Walker, B. E. (2020). Extinction risk and threats to plants and fungi. *Plants, People, Planet*, *2*(5), 389–408. https://doi.org/10.1002/PPP3.10146

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., & Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annual Review of Genetics*, *38*, 645–679. https://doi.org/10.1146/ANNUREV.GENET.37.110801.143214

Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J (2022). _vegan: Community Ecology Package_. R package version 2.6-4, <https://CRAN.R-project.org/package=vegan>

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/BIOINFORMATICS/BTG412

Poulton, P. R. (2011). The Rothamsted long-term experiments: Are they still relevant? *Https://Doi.Org/10.4141/Cjps96-103*, *76*(4), 559–571. https://doi.org/10.4141/CJPS96-103

Pyšek, P., Hulme, P. E., Simberloff, D., Bacher, S., Blackburn, T. M., Carlton, J. T., Dawson, W., Essl, F., Foxcroft, L. C., Genovesi, P., Jeschke, J. M., Kühn, I., Liebhold, A. M., Mandrak, N. E., Meyerson, L. A., Pauchard, A., Pergl, J., Roy, H. E., Seebens, H., … Richardson, D. M. (2020). Scientists' warning on invasive alien species. *Biological Reviews*, *95*(6), 1511–1534. https://doi.org/10.1111/BRV.12627

Peiffer, J., A.Spor, A.,Koren, O.,Jin, Z.Tringe, S. G., Dangl, J. L., Buckler, E. S., Ley, R. E (2013). Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proceedings of the National Academy of Sciences,* 6548-6553, 110, 16. https://doi:10.1073/pnas.1302837110

Quagliariello, A., Modi, A., Innocenti, G., Zaro, V., Conati Barbaro, C., Ronchitelli, A., Boschin, F., Cavazzuti, C., Dellù, E., Radina, F., Sperduti, A., Bondioli, L., Ricci, S., Lognoli, M., Belcastro, M. G., Mariotti, V., Caramelli, D., Mariotti Lippi, M., Cristiani, E., … Lari, M. (2022). Ancient oral microbiomes support gradual Neolithic dietary shifts towards agriculture. *Nature Communications 2022 13:1*, *13*(1), 1–14. https://doi.org/10.1038/s41467-022-34416-0

R Core Team (2021). R: A language and environment for statistical ## computing. R Foundation for Statistical Computing, Vienna, Austria. ## URL https://www.R-project.org/.

Roberts David W (2023). Labdsv: Ordination and Multivariate Analysis for Ecology. <https://cran.r-project.org/web/packages/labdsv/labdsv.pdf>

Rull, V. (2021). Biodiversity crisis or sixth mass extinction? *EMBO Reports*, *23*(1). https://doi.org/10.15252/EMBR.202154193

Särkinen, T., Staats, M., Richardson, J. E., Cowan, R. S., & Bakker, F. T. (2012). How to Open the Treasure Chest? Optimising DNA Extraction from Herbarium Specimens. *PLoS ONE*, *7*(8). https://doi.org/10.1371/JOURNAL.PONE.0043808

Siano, R., Lassudrie, M., Cuzin, P., Briant, N., Loizeau, V., Schmidt, S., Ehrhold, A., Mertens, K. N., Lambert, C., Quintric, L., Noël, C., Latimier, M., Quéré, J., Durand, P., & Penaud, A. (2021). Sediment archives reveal irreversible shifts in plankton communities after World War II and agricultural pollution. *Current Biology*, *31*(12), 2682-2689.e7. https://doi.org/10.1016/J.CUB.2021.03.079

Staats, M., Cuenca, A., Richardson, J. E., Vrielink-Van Ginkel, R., & Petersen, G. (2011). DNA Damage in Plant Herbarium Tissue. *PLoS ONE*, *6*(12), 28448. https://doi.org/10.1371/journal.pone.0028448

Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O., & Ludwig, C. (2015). The trajectory of the Anthropocene: The Great Acceleration. The Anthropocene Review, 2(1), 81-98. https://doi.org/10.1177/2053019614564785

Thakur, M. P., van der Putten, W. H., Cobben, M. M. P., van Kleunen, M., & Geisen, S. (2019). Microbial invasions in terrestrial ecosystems. *Nature Reviews Microbiology 2019 17:10*, *17*(10), 621–631. https://doi.org/10.1038/s41579-019-0236-z

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., Prill, R. J., Tripathi, A., Gibbons, S. M., Ackermann, G., Navas-Molina, J. A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J. T., Mirarab, S., Xu, Z. Z., Jiang, L., … Zhao, H. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature 2017 551:7681*, *551*(7681), 457–463. https://doi.org/10.1038/nature24621

Tkacz, A., Bestion, E., Bo, Z., Hortala, M., & Poole, P. S. (2020). Influence of plant fraction, soil, and plant species on microbiota: A multikingdom comparison. *MBio*, *11*(1). https://doi.org/10.1128/MBIO.02785-19/ASSET/C87D6EF3-55B5-45C3-8DA3-B982F8F2C8BE/ASSETS/GRAPHIC/MBIO.02785-19-F0006.JPEG

Vavrek MJ (2011). "fossil: palaeoecological and palaeogeographical analysis tools." *Palaeontologia Electronica*, **14**(1), 1T. R package version 0.4.0.

Wang, L., Li, Q., li, C., Wu, C., Chen, F., Chen, X., & Zhang, F. (2023). Nitrate Nitrogen and pH Correlate with Changes in Rhizosphere Microbial Community Assemblages during Invasion of Ambrosia artemisiifolia and Bidens pilosa. *Microbiology Spectrum*, *11*(1). https://doi.org/10.1128/SPECTRUM.03649-22

Weiß, C. L., Schuenemann, V. J., Devos, J., Shirsekar, G., Reiter, E., Gould, B. A., Stinchcombe, J. R., Krause, J., & Burbano, H. A. (2016). Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *R. Soc. Open Sci*, *3:160239*. https://doi.org/10.1098/rsos.160239

Xiong, C., Singh, B. K., He, J. Z., Han, Y. L., Li, P. P., Wan, L. H., Meng, G. Z., Liu, S. Y., Wang, J. T., Wu, C. F., Ge, A. H., & Zhang, L. M. (2021). Plant developmental stage drives the differentiation in ecological role of the maize microbiome. *Microbiome*, *9*(1), 1–15. https://doi.org/10.1186/S40168-021-01118-6/FIGURES/5

Xun, W., Shao, J., Shen, Q., & Zhang, R. (2021). Rhizosphere microbiome: Functional compensatory assembly for plant fitness. *Computational and Structural Biotechnology Journal*, *19*, 5487–5493. https://doi.org/10.1016/j.csbj.2021.09.035

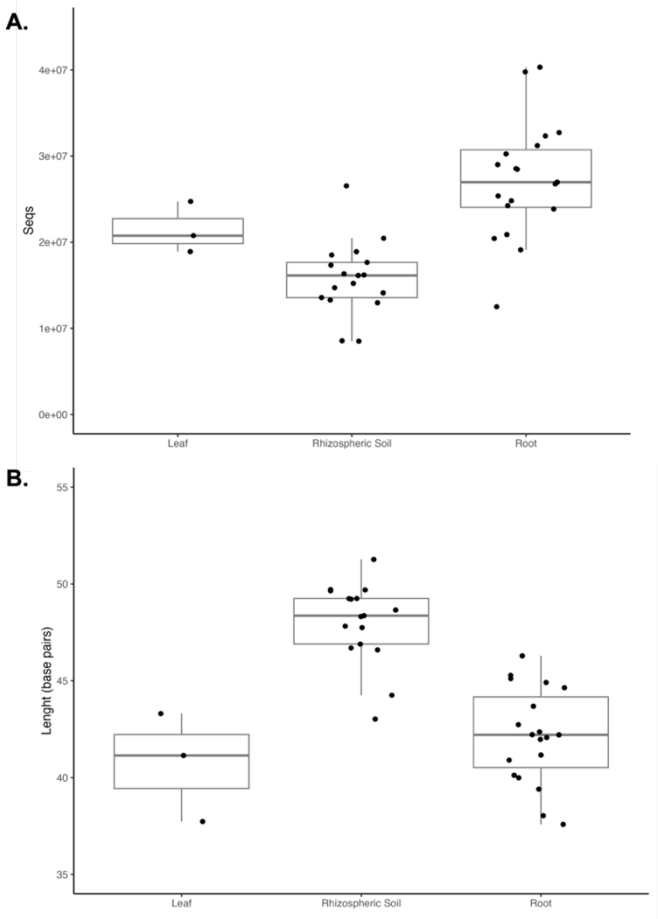
Zhou, Z., Wang, C., & Luo, Y. (2020). Meta-analysis of the impacts of global change factors on soil microbial diversity and functionality. *Nature Communications 2020 11:1*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-16881-7

Zhu, Q., Huang, S., Gonzalez, A., McGrath, I., McDonald, D., Haiminen, N., Armstrong, G., Vázquez-Baeza, Y., Yu, J., Kuczynski, J., Sepich-Poore, G. D., Swafford, A. D., Das, P., Shaffer, J. P., Lejzerowicz, F., Belda-Ferre, P., Havulinna, A. S., Méric, G., Niiranen, T., … Knight, R. (2022). Phylogeny-Aware Analysis of Metagenome Community Ecology Based on Matched Reference Genomes while Bypassing Taxonomy. *MSystems*, *7*(2). https://doi.org/10.1128/MSYSTEMS.00167-22/SUPPL_FILE/MSYSTEMS.00167-22-ST002.XLSX

Zhu, Y. G., Gillings, M., Simonet, P., Stekel, D., Banwart, S., & Penuelas, J. (2017). Microbial mass movements. *Science*, *357*(6356), 1099–1100. https://doi.org/10.1126/SCIENCE.AAO3007/SUPPL_FILE/AAO3007_ZHU_SM.PDF

Supplementary

Figure S1. Summary of Sequencing Data Analysis A. Total number of sequences obtained after demultiplexing, quality controls, merging and eliminations of reads shorter than 25 bp of the raw data. In millions of sequencing reads, the filtered data had an average of 21.8, a standard deviation of 7.7, a minimum of 8.5 for sample Sc707 and Sc-689 rhizospheric soil and a maximum of 40.3 for sample Td-279 root. **B.** Average length (bp) value of the merged reads of the samples. Average values of the different samples are 40.72 for leaf, 48.02



for rhizospheric soil and 42.14 for root samples.

Figure S2. Effect of plant species on bacterial and eukaryotic root microbiomes. Principal Coordinates Analysis ordinations based on Bray-Curtis indices illustrate the distribution of eukaryotic (A) and bacterial (B) root communities according to the plant species (P< 0.05, PerMANOVA, **Table S6**).

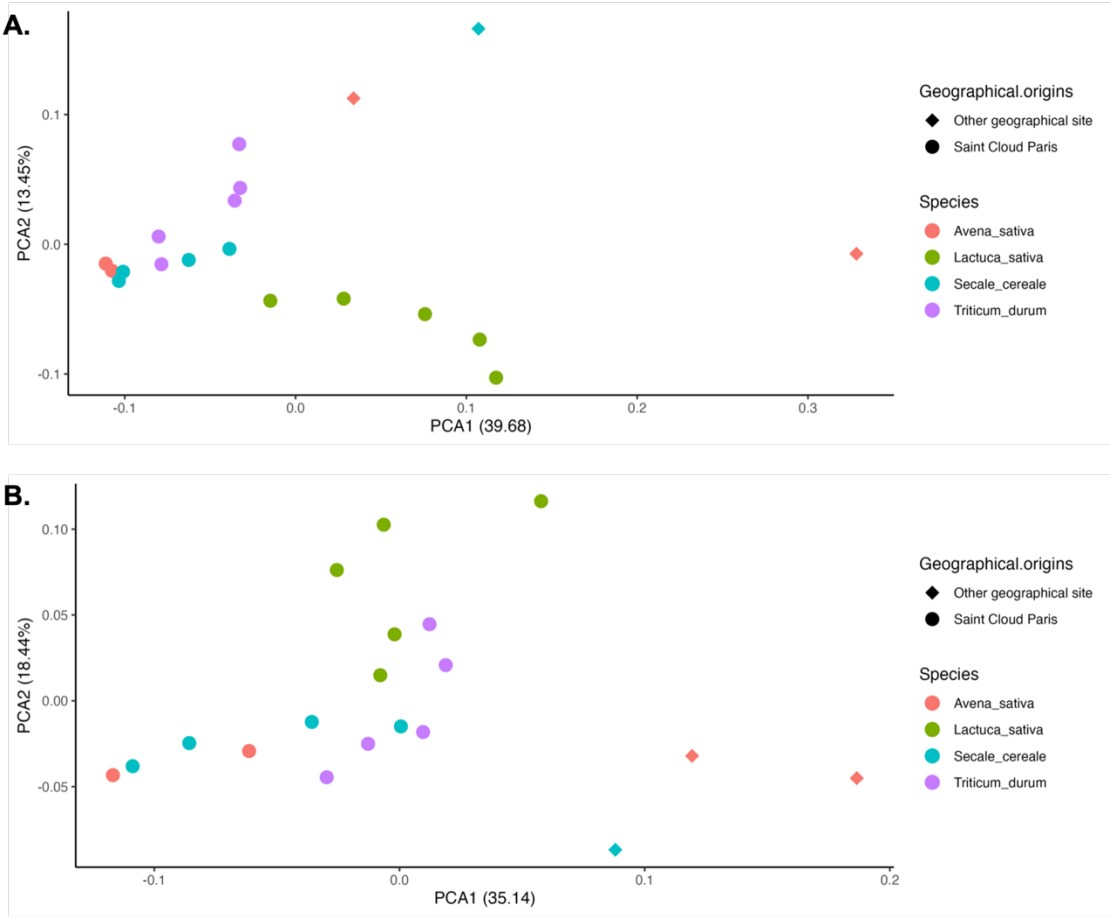


Figure S3. Effect of plant species on bacterial and eukaryotic root microbiomes. At the genus level, in the case of bacteria (**A**) and the samples from Saint-Cloud, a large proportion of the genera were identified in both soil and roots (53 % and 71 % for genera represented by at least 0.05 % (left) or 0.3 % (right) of the total number of reads, respectively). Genera identified only in soil outnumbered those identified only in roots (39 *versus* 25 or 104 *versus* 42 for genera represented by at least 0.3 % or 0.05 % of the total number of reads, respectively). A similar pattern of distribution of taxa between roots and soil was also observed for Eukarya for which the relaxed and stringent filtering thresholds were set to 0.1 (right) and 0.5 % (left) of the total number of reads"(**B**).

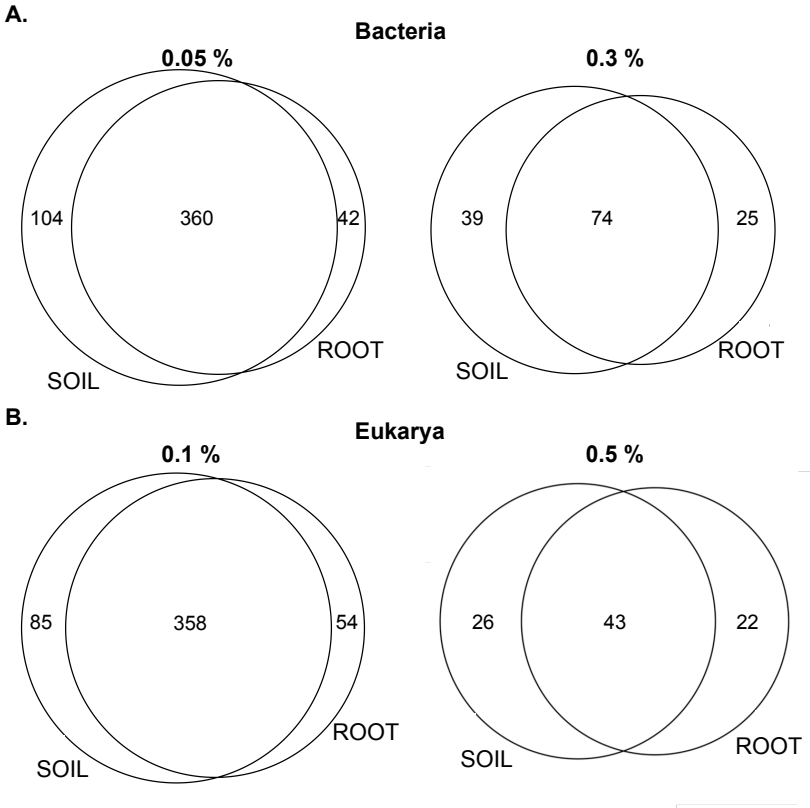


Table S1. Studied herbarium material. All samples came from the herbarium of the University Lyon 1 (LY), France. “species names”, “poisoned”, “variety”, “collection date” and “collection site” refer to information written on the herbarium sheets; “ni”, not indicated. Plants from Saint-Cloud were grown in a common garden that no longer exists due to urbanization of the area. “Specimen accession No.” refers to the Recolnat accession Nos. that allow visualizing the corresponding material (https://explore.recolnat.org/search/botanique/type=index).

Species	Specimen acession No.	Poisoned	Variety	Collection date	Collection site	Geographic coordinate	Sample name	Sample type
Avena sativa (Oat)	LY0642166	ni	ni	1907	Aisne	N 49°27'11.8'' E 3°36'24.8''	As_0166R	Root
							As_0166S	Rhizospheric soil
	LY0642205	ni	Noire de Brie	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	As_205R	Root
							As_205S	Rhizospheric soil
	LY0642225	ni	ni	1903	Royat	N 45°45'51.7'' E 3°2'52.0''	As_225R	Root
							As_225S	Rhizospheric soil
	LY0642206	ni	Noire de Brie	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	As_FHR	Root
							As_FHS	Rhizospheric soil
Lactuca sativa (Lettuce)	LY0411321	yes	Laitue Frisée d'Amérique	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Ls_321R	Root
	LY0411358	yes	Laitue pommée grosse blonde Parisienne	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Ls_358R	Root
	LY0411378	yes	petite Laitue crêpe	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Ls_378R	Root
							Ls_378S	Rhizospheric soil
	LY0411389	yes	Laitue Romaine blonde maraichère	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Ls_389R	Root
							Ls_389S	Rhizospheric soil
	LY0411390	yes	Laitue Romaine blonde maraichère	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Ls_390F	Leaf
							Ls_390R	Root
							Ls_390S	Rhizospheric soil
Secale cereale (Rye)	LY0662637	ni	Seigle de Mars ordinaire	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Sc_637R	Root
							Sc_637S	Rhizospheric soil
	LY0662680	ni	Seigle d'hiver de Schlanstedt	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Sc_680R	Root
							Sc_680S	Rhizospheric soil
	LY0662663	ni	Seigle géant d'hiver	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Sc_684F	Leaf
							Sc_684R	Root
							Sc_684S	Rhizospheric soil
	LY0662689	ni	Seigle de Mars ordinaire	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Sc_689R	Root
							Sc_689S	Rhizospheric soil
	LY0662707	ni	ni	1901	Port-la-Nouvelle	N 43°1'13.3'' E 3°2'47.3''	Sc_707S	Root
							Sc_707R	Rhizospheric soil
Triticum durum (durum wheat)	LY0664270	ni	Blé de Xérès	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Td_270R	Root
							Td_270S	Rhizospheric soil
	LY0664279	ni	Blé trimenia barbu de Sicile	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Td_279R	Root
							Td_279S	Rhizospheric soil
	LY0664283	ni	Blé dur de Médéah	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Td_283F	Leaf
							Td_283R	Root
							Td_283S	Rhizospheric soil
	LY0664284	ni	Blé dur de Médéah	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Td_284R	Root
							Td_284S	Rhizospheric soil
	LY0664297	ni	Blé Beloturka	1903	Saint-Cloud	N 48°50'37.4'' E 2°13'9.6''	Td_297R	Root
							Td_297S	Rhizospheric soil

Table S2. References nuclear, mitochondrial, and plastid plant genomes . All plant reference genomes were downloaded from the NCBI database. For species whose complete mitochondrial or chloroplast genomes were not available, genomes of phylogenetically-close plant species were selected.

Specie	Organelle	Accession NCBI
Avena longiglumis	Mitochondrion	OP649434.1
Avena longiglumis	Mitochondrion	OP649436
Avena longiglumis	Mitochondrion	OP649437.1
Avena longiglumis	Mitochondrion	OP649435.1
Avena sativa	Chloroplast	MG687313.1
Avena sativa	Chloroplast	MK336398.1
Avena sativa	Nuclear	GCA_022788535.1
Hordeum vulgare	Mitochondrion	AP017300.1
Hordeum vulgare	Mitochondrion	AP017301.1
Lactuca sativa	Chloroplast	AP007232.1
Lactuca sativa	Chloroplast	NC_007578.1
Lactuca sativa	Mitochondrion	NC_042756.1
Lactuca sativa	Mitochondrion	MZ159955.1
Lactuca sativa	Nuclear	GCA_002870075.4
Secale cereale	Chloroplast	NC_021761
Secale cereale	Chloroplast	LC645358.1
Secale cereale	Nuclear	GCA_016097815.1
Triticum aestivum	Mitochondrion	EU534409.1
Triticum turgidum	Chloroplast	KJ614397.1
Triticum turgidum	Chloroplast	KJ614400.1
Triticum turgidum	Chloroplast	KJ614399.1
Triticum turgidum	Chloroplast	KJ614401.1
Triticum turgidum	Chloroplast	KJ614402.1
Triticum turgidum	Chloroplast	KJ614398.1
Triticum turgidum	Chloroplast	MG958545.1
Triticum turgidum	Mitochondrion	KJ078650.1
Triticum turgidum	Mitochondrion	KJ078649
Triticum turgidum	Nuclear	GCA_900231445.1

Table S3. Reference genomes of the most abundant bacterial genera used to run MapDamage. MapDamage was run using the metagenomics sequences that did not map on plant genomes using as reference database all of the reference genomes (RefSeq present in the NCBI database of April 2023) affiliated to the 10 most abundant bacterial genera identified following taxonomic affiliation of the sequences using Kaiju.

Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession
Bradyrhizobium acaciae	GCF_020889785.1	Bradyrhizobium frederickii	GCF_004570865.1	Bradyrhizobium retamae	GCF_001440415.1	Devosia indica	GCF_003056405.1
Bradyrhizobium aeschynomenes	GCF_013178945.1	Bradyrhizobium glycinis	GCF_016031655.1	Bradyrhizobium rifense	GCF_008123425.1	Devosia insulae DS-56	GCF_000970465.2
Bradyrhizobium agreste	GCF_016031625.1	Bradyrhizobium guangdongense	GCF_004114975.1	Bradyrhizobium sacchari	GCF_007828095.1	Devosia limi DSM 17137	GCF_900128975.1
Bradyrhizobium algeriense	GCF_003062295.1	Bradyrhizobium guangxiense	GCF_004114915.1	Bradyrhizobium sediminis	GCF_018736085.1	Devosia lucknowensis	GCF_900177655.1
Bradyrhizobium altum	GCF_020889705.1	Bradyrhizobium guangzhouense	GCF_004114955.1	Bradyrhizobium semiaridum	GCF_020329505.1	Devosia marina	GCF_009758415.1
Bradyrhizobium amphicarpaeae	GCF_002266435.2	Bradyrhizobium hereditatis	GCF_020329435.1	Bradyrhizobium septentrionale	GCF_011516645.4	Devosia naphthalenivorans	GCF_003056355.1
Bradyrhizobium arachidis	GCF_015291705.1	Bradyrhizobium hipponense	GCF_008123965.1	Bradyrhizobium shewense	GCF_900094605.1	Devosia neptuniae	GCF_027214225.1
Bradyrhizobium archetypum	GCF_013114835.1	Bradyrhizobium huanghuaihaiense	GCF_007830635.1	Bradyrhizobium stylosanthis	GCF_001641335.1	Devosia oryzaradicis	GCF_016698645.1
Bradyrhizobium australafricanum	GCF_020329495.1	Bradyrhizobium icense	GCF_001693385.1	Bradyrhizobium symbiodeficiens	GCF_011604665.1	Devosia oryzisoli	GCF_014837245.1
Bradyrhizobium australiense	GCF_013114825.1	Bradyrhizobium iriomotense	GCF_018130765.1	Bradyrhizobium tropiciagri	GCF_001189845.1	Devosia pacifica	GCF_014652635.1
Bradyrhizobium barranii subsp. barranii	GCF_011068405.5	Bradyrhizobium ivorense	GCF_900696085.1	Bradyrhizobium uaiense	GCF_010811875.1	Devosia psychrophila	GCF_900112505.1
Bradyrhizobium betae	GCF_008932115.1	Bradyrhizobium japonicum	GCF_013752735.1	Bradyrhizobium valentinum	GCF_001440385.1	Devosia rhizoryzae	GCF_016698665.1
Bradyrhizobium brasilense	GCF_900101405.1	Bradyrhizobium jicamae	GCF_001440395.1	Bradyrhizobium vignae	GCF_900324035.1	Devosia riboflavina	GCF_000743575.1
Bradyrhizobium cajani	GCF_009759665.1	Bradyrhizobium lablabi	GCF_001440475.1	Bradyrhizobium viridifuturi	GCF_001238275.1	Devosia salina	GCF_019504385.1
Bradyrhizobium campsiandrae	GCF_014530645.1	Bradyrhizobium liaoningense	GCF_018130825.1	Bradyrhizobium yuanmingense	GCF_029167075.1	Devosia sediminis	GCF_016411825.1
Bradyrhizobium canariense	GCF_019402665.1	Bradyrhizobium macuxiense	GCF_001542415.1	Bradyrhizobium zhanjiangense	GCF_004114935.1	Devosia soli	GCF_000970455.1
Bradyrhizobium cenepequi	GCF_020329485.1	Bradyrhizobium manausense	GCF_001440035.1	Bradyrhizobium zhengyangense	GCF_022012555.1	Devosia subaequoris	GCF_014197055.1
Bradyrhizobium centrolobii	GCF_001641635.1	Bradyrhizobium mercantei	GCF_001982635.1	Candidatus Pseudomonas adelgestsugas	GCF_004124255.1	Devosia submarina	GCF_003056345.1
Bradyrhizobium centrosematis	GCF_024809475.1	Bradyrhizobium murdochi	GCF_000472965.1	Devosia aurantiaca	GCF_011058215.1	Devosia ureilytica	GCF_024273205.1
Bradyrhizobium cosmicum	GCF_000284275.1	Bradyrhizobium nanningense	GCF_004114535.1	Devosia beringensis	GCF_014926585.1	Flavobacterium aciduliphilum	GCF_003268855.1
Bradyrhizobium cytisi	GCF_008123515.1	Bradyrhizobium neotropicale	GCF_001641695.1	Devosia chinhatensis	GCF_000969445.1	Flavobacterium agri	GCF_013366395.1
Bradyrhizobium daqingense	GCF_021044685.1	Bradyrhizobium niftali	GCF_004571025.1	Devosia crocina	GCF_900116545.1	Flavobacterium agrisoli	GCF_016522065.1
Bradyrhizobium denitrificans	GCF_018129525.1	Bradyrhizobium nitroreducens	GCF_002776695.1	Devosia elaeis	GCF_001650025.1	Flavobacterium ajazii	GCF_010614725.1
Bradyrhizobium diazoefficiens	GCF_004359355.1	Bradyrhizobium oligotrophicum S58	GCF_000344805.1	Devosia enhydra	GCF_900119845.1	Flavobacterium akiainvivens	GCF_001278115.1
Bradyrhizobium diversitatis	GCF_016031635.1	Bradyrhizobium oropedii	GCF_020889685.1	Devosia epidermidihirudinis	GCF_000971295.1	Flavobacterium album	GCF_003096035.1
Bradyrhizobium elkanii USDA 61	GCF_012871055.1	Bradyrhizobium ottawaense	GCF_009176665.1	Devosia equisanguinis	GCF_900631955.1	Flavobacterium algicola	GCF_022399565.1
Bradyrhizobium embrapense	GCF_001189235.2	Bradyrhizobium pachyrhizi	GCF_009759865.1	Devosia faecipullorum	GCF_015158295.1	Flavobacterium alkalisoli	GCF_008000935.1
Bradyrhizobium erythrophlei	GCF_900142985.1	Bradyrhizobium paxllaeri	GCF_001693515.2	Devosia geojensis	GCF_000969415.1	Flavobacterium alvei	GCF_002920895.1
Bradyrhizobium forestalis	GCF_002795245.1	Bradyrhizobium quebecense	GCF_017493175.3	Devosia ginsengisoli	GCF_007859655.1	Flavobacterium ammonificans	GCF_020885975.1
Organism Scientific Name	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**
Flavobacterium ammoniigenes	GCF_020886055.1	Flavobacterium cheniae	GCF_007830415.1	Flavobacterium endophyticum	GCF_003634455.1	Flavobacterium hibernum	GCF_000832125.1
Flavobacterium amnicola	GCF_004122165.1	Flavobacterium cheongpyeongense	GCF_003202435.1	Flavobacterium enshiense DK69	GCF_000498495.1	Flavobacterium hibisci	GCF_020026925.1
Flavobacterium amniphilum	GCF_023634845.1	Flavobacterium chilense	GCF_900142685.1	Flavobacterium erciyesense	GCF_018069925.1	Flavobacterium hiemivividum	GCF_004349145.1
Flavobacterium anhuiense	GCF_001705175.1	Flavobacterium chungangense LMG 26729	GCF_000735715.2	Flavobacterium faecale	GCF_003076455.1	Flavobacterium humi	GCF_004745595.1
Flavobacterium antarcticum DSM 19726	GCF_000419685.1	Flavobacterium chungbukense	GCF_020805785.1	Flavobacterium filum DSM 17961	GCF_000425465.1	Flavobacterium humidisoli	GCF_023272795.1
Flavobacterium aquariorum	GCF_003254745.1	Flavobacterium circumlabens	GCF_004345565.1	Flavobacterium flabelliforme	GCF_017948675.1	Flavobacterium hydatis	GCF_002217335.1
Flavobacterium aquaticum	GCF_003259835.1	Flavobacterium collinsii	GCF_946903285.1	Flavobacterium flevense	GCF_900142775.1	Flavobacterium hydrocarbonoxydans	GCF_009789235.1
vobacterium aquatile LMG 4008 = ATCC 1	GCF_000757385.1	Flavobacterium columnare	GCF_014844295.1	Flavobacterium fluviale	GCF_003312915.1	Flavobacterium hydrophilum	GCF_003202405.1
Flavobacterium aquicola	GCF_003385115.1	Flavobacterium commune	GCF_001857965.1	Flavobacterium fluviatile	GCF_010645065.1	Flavobacterium ichthyis	GCF_009915155.1
Flavobacterium aquidurense	GCF_002217195.1	Flavobacterium coralii	GCF_019891545.1	Flavobacterium fluvii	GCF_900129545.1	obacterium indicum GPTSA100-9 = DSM 1	GCF_000455605.1
Flavobacterium araucananum	GCF_003148525.1	Flavobacterium covae	GCF_001534645.1	Flavobacterium foetidum	GCF_004634245.1	Flavobacterium inviolabile	GCF_013389455.1
Flavobacterium arcticum	GCF_003344925.1	Flavobacterium crassostreae	GCF_001831475.1	Flavobacterium fontis	GCF_900129405.1	Flavobacterium jejuense	GCF_006491595.2
Flavobacterium aurantiibacter	GCF_002251775.1	Flavobacterium croceum DSM 17960	GCF_002917885.1	Flavobacterium franklandianum	GCF_007097365.1	Flavobacterium johnsoniae UW101	GCF_000016645.1
Flavobacterium azooxidireducens	GCF_023195775.1	Flavobacterium crocinum	GCF_003122385.1	Flavobacterium frigidarium DSM 17623	GCF_000425505.1	Flavobacterium jumunjinense	GCF_021650975.2
Flavobacterium baculatum	GCF_008629655.1	Flavobacterium cucumis	GCF_900148835.1	Flavobacterium frigidimaris	GCF_002217275.1	Flavobacterium kayseriense	GCF_014305095.1
Flavobacterium beibuense F44-8	GCF_000769915.1	Flavobacterium cupreum	GCF_003996965.1	Flavobacterium frigoris	GCF_900111075.1	Flavobacterium kingsejongi	GCF_003076475.1
Flavobacterium bernardetii	GCF_011305415.1	Flavobacterium cutihirudinis	GCF_003385895.1	Flavobacterium fryxellicola	GCF_900143245.1	Flavobacterium lacus	GCF_003268815.1
Flavobacterium bizetiae	GCF_904425415.1	Flavobacterium cyanobacteriorum	GCF_002251835.1	Flavobacterium gawalongense	GCF_007097285.1	Flavobacterium laiguense	GCF_003097655.1
Flavobacterium bomense	GCF_003970355.1	Flavobacterium cydinae	GCF_021172145.1	Flavobacterium gelidilacus DSM 15343	GCF_000422685.1	Flavobacterium limi	GCF_010645075.1
Flavobacterium branchiicola	GCF_018383905.1	Flavobacterium daejeonense DSM 17708	GCF_000425425.1	Flavobacterium geliluteum	GCF_017948595.1	Flavobacterium limicola	GCF_003634755.1
Flavobacterium branchiophilum FL-15	GCF_000253275.1	Flavobacterium daemonense	GCF_007341385.1	Flavobacterium gillisiae	GCF_900107635.1	Flavobacterium limnosediminis JC2902	GCF_000498535.1
Flavobacterium buctense	GCF_014486695.1	Flavobacterium dankookense	GCF_004362665.1	Flavobacterium gilvum	GCF_001761465.1	Flavobacterium lindanitolerans	GCF_002846575.1
Flavobacterium caeni	GCF_900101895.1	Flavobacterium dauae	GCF_004151275.1	Flavobacterium ginsenosidimutans	GCF_003254625.1	Flavobacterium litorale	GCF_019613795.1
Flavobacterium caseinilyticum	GCF_004349085.1	Flavobacterium davisii	GCF_002204815.1	Flavobacterium glaciei	GCF_003350545.1	Flavobacterium lotistagni	GCF_011392125.1
Flavobacterium cauense R2A-7	GCF_000498475.1	Flavobacterium defluvii	GCF_900129555.1	Flavobacterium glycines	GCF_900100165.1	Flavobacterium luteum	GCF_008806775.1
Flavobacterium celericrescens	GCF_011392075.1	Flavobacterium degerlachei	GCF_900106645.1	Flavobacterium gossypii	GCF_014138495.1	Flavobacterium macacae	GCF_003865365.1
Flavobacterium cellulosilyticum	GCF_004349355.1	Flavobacterium denitrificans DSM 15936	GCF_000425445.1	Flavobacterium granuli	GCF_003003155.1	Flavobacterium macrobrachii	GCF_014486675.1
Flavobacterium cerinum	GCF_004028155.1	Flavobacterium difficile	GCF_011392115.1	Flavobacterium haoranii	GCF_900142055.1	Flavobacterium magnum	GCF_003055625.1
Flavobacterium channae	GCF_021172165.1	Flavobacterium endoglycinae	GCF_017352115.1	Flavobacterium hercynium	GCF_002217285.1	Flavobacterium microcysteis	GCF_006385255.1

Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession
Flavobacterium micromati	GCF_900129585.1	Flavobacterium quisquiliarum	GCF_019308285.1	Flavobacterium suaedae	GCF_014642915.1	Flavobacterium zepuense	GCF_007097145.1
Flavobacterium muglaense	GCF_014305155.1	Flavobacterium rakeshii	GCF_009741375.1	bacterium subsaxonicum WB 4.1-42 = DSM	GCF_000422725.1	Flavobacterium zhairuonense	GCF_004352915.1
Flavobacterium nackdongense	GCF_004355225.1	Flavobacterium ranwuense	GCF_004349315.1	Flavobacterium succinicans	GCF_900114945.1	Lentzea aerocolonigenes	GCF_000974445.1
Flavobacterium nitrogenifigens	GCF_008369745.1	Flavobacterium reichenbachii	GCF_000737685.1	Flavobacterium sufflavum	GCF_004016525.1	Lentzea alba	GCF_011067745.1
Flavobacterium noncentrifugens	GCF_900100375.1	Flavobacterium resistens	GCF_900182645.1	bacterium suncheonense GH29-5 = DSM 1	GCF_000430025.1	Lentzea albida	GCF_900111005.1
Flavobacterium notoginsengisoli	GCF_014207905.1	Flavobacterium restrictum	GCF_007097245.1	Flavobacterium supellecticarium	GCF_004801375.1	Lentzea albidocapillata	GCF_900176525.1
Flavobacterium omnivorum	GCF_900099915.1	Flavobacterium rhamnosiphilum	GCF_004349195.1	Flavobacterium swingsii	GCF_900111965.1	Lentzea atacamensis	GCF_003148865.1
Flavobacterium oncorhynchi	GCF_002217355.1	avobacterium rivuli WB 3.3-2 = DSM 2178	GCF_000378485.1	Flavobacterium tagetis	GCF_016735615.1	Lentzea californiensis	GCF_024648825.1
Flavobacterium oreochromis	GCF_002204835.1	Flavobacterium saccharophilum	GCF_900142735.1	Flavobacterium taihuense	GCF_019351435.1	Lentzea cavernae	GCF_014653695.1
Flavobacterium orientale	GCF_014638005.1	Flavobacterium salilacus subsp. salilacus	GCF_004634195.1	Flavobacterium tangerinum	GCF_003865405.1	Lentzea flava	GCF_014648475.1
Flavobacterium pallidum	GCF_003097535.1	Flavobacterium saliperosum	GCF_900100625.1	Flavobacterium tegetincola DSM 22377	GCF_000425485.1	Lentzea flaviverrucosa	GCF_003350445.1
Flavobacterium palustre	GCF_014639535.1	Flavobacterium salmonis	GCF_903819435.1	Flavobacterium terrae	GCF_900142035.1	Lentzea fradiae	GCF_900100955.1
Flavobacterium panacis	GCF_024649945.1	Flavobacterium sandaracinum	GCF_004349135.1	Flavobacterium terrigena	GCF_900108955.1	Lentzea guizhouensis	GCF_001701025.1
Flavobacterium panici	GCF_024499525.1	Flavobacterium sangjuense	GCF_004797125.1	Flavobacterium tiangeerense	GCF_007830355.1	Lentzea indica	GCF_012184385.1
Flavobacterium pectinovorum	GCF_900142715.1	Flavobacterium sasangense DSM 21067	GCF_000686885.1	Flavobacterium tibetense	GCF_003293845.1	Lentzea jiangxiensis	GCF_900104175.1
Flavobacterium petrolei	GCF_003314435.2	Flavobacterium sediminilitoris	GCF_023008245.1	Flavobacterium tistrianum	GCF_003254545.1	Lentzea kentuckyensis	GCF_002150765.1
Flavobacterium phragmitis	GCF_900112575.1	Flavobacterium sediminis	GCF_003148385.1	Flavobacterium tructae	GCF_001853485.1	Lentzea nigeriaca	GCF_016907955.1
Flavobacterium phycosphaerae	GCF_010119235.1	Flavobacterium segetis	GCF_900129575.1	Flavobacterium turcicum	GCF_014305215.1	Lentzea pudingi	GCF_014646255.1
Flavobacterium piscinae	GCF_004122145.1	Flavobacterium selenitireducens	GCF_014748335.1	Flavobacterium tyrosinilyticum	GCF_023656565.1	Lentzea terrae	GCF_003265345.1
Flavobacterium piscis	GCF_001686925.1	Flavobacterium seoulense	GCF_000695795.1	Flavobacterium ummariense	GCF_900115115.1	Lentzea tibetensis	GCF_007845675.1
Flavobacterium plurextorum	GCF_002217395.1	Flavobacterium sharifuzzamanii	GCF_003254585.1	Flavobacterium undicola	GCF_009909155.2	Lentzea waywayandensis	GCF_900115955.1
Flavobacterium pokkalii	GCF_014596575.1	Flavobacterium silvaticum	GCF_012911565.1	Flavobacterium urocanicphilum	GCF_900110615.1	Lentzea xinjiangensis	GCF_900110955.1
Flavobacterium potami	GCF_020026915.1	Flavobacterium silvisoli	GCF_004329815.1	Flavobacterium urumqiense	GCF_900108015.1	Massilia agilis	GCF_024756255.1
Flavobacterium profundi	GCF_006491645.1	Flavobacterium sinopsychrotolerans	GCF_900110375.1	Flavobacterium ustbae	GCF_003946915.1	Massilia agri	GCF_024753255.1
Flavobacterium proteolyticum	GCF_015223105.1	Flavobacterium soli DSM 19725	GCF_000422705.1	Flavobacterium viscosum	GCF_003858535.1	Massilia alkalitolerans DSM 17462	GCF_000427785.1
Flavobacterium psychrolimnae	GCF_003312425.1	Flavobacterium solisilvae	GCF_012911605.1	Flavobacterium weaverense	GCF_003688495.1	Massilia antarctica	GCF_015689335.1
Flavobacterium psychrophilum	GCF_013343195.2	Flavobacterium soyae	GCF_021245985.1	Flavobacterium xanthum	GCF_900142695.1	Massilia antibiotica	GCF_011682055.1
Flavobacterium psychroterrae	GCF_018380615.1	Flavobacterium soyangense	GCF_015277675.1	Flavobacterium xinjiangense	GCF_900142885.1	Massilia aquatica	GCF_011682045.1
Flavobacterium psychrotolerans	GCF_003097635.1	Flavobacterium stagni	GCF_004122105.1	Flavobacterium xueshanense	GCF_900112975.1	Massilia arenae	GCF_008014745.1
Organism Scientific Name	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**
Massilia arenosa	GCF_004614185.1	Microbacterium aerolatum	GCF_007988825.1	Microbacterium flavum	GCF_018717645.1	Microbacterium lacus	GCF_002812805.1
Massilia aromaticivorans	GCF_013003915.1	Microbacterium agarici	GCF_002563955.1	Microbacterium fluvii	GCF_025515205.1	Microbacterium laevaniformans	GCF_001584605.1
Massilia atriviolacea	GCF_003953935.1	Microbacterium album	GCF_014643695.1	Microbacterium foliorum	GCF_003367705.1	Microbacterium lemovicicum	GCF_003991875.1
Massilia aurea	GCF_024753205.1	Microbacterium algeriense	GCF_008868005.1	Microbacterium gallinarum	GCF_014837165.1	Microbacterium lindanitolerans	GCF_006788875.1
Massilia cavernae	GCF_003590855.1	Microbacterium allomyrinae	GCF_020682705.1	Microbacterium ginsengisoli	GCF_000956535.1	Microbacterium liquefaciens	GCF_024362265.1
Massilia eburnea	GCF_009720745.1	Microbacterium amylolyticum	GCF_011046975.1	Microbacterium ginsengiterrae	GCF_014205075.1	Microbacterium lushaniae	GCF_008727775.1
Massilia eurypsychrophila	GCF_002760655.1	Microbacterium aoyamense	GCF_023155855.1	Microbacterium gorillae	GCF_000826185.2	Microbacterium luteum	GCF_015277875.1
Massilia forsythiae	GCF_012849555.1	Microbacterium aquimaris	GCF_015278225.1	Microbacterium gubbeenense DSM 15944	GCF_000422745.1	Microbacterium luticocti DSM 19459	GCF_000422405.1
Massilia frigida	GCF_011682175.1	Microbacterium arborescens	GCF_001662775.1	Microbacterium halimionae	GCF_011761265.1	Microbacterium mangrovi	GCF_000802305.1
Massilia glaciei	GCF_003011895.2	Microbacterium atlanticum	GCF_015277815.1	Microbacterium halophytorum	GCF_002970975.1	Microbacterium marinilacus	GCF_019753765.1
Massilia horti	GCF_004614195.1	Microbacterium aurantiacum	GCF_021654035.1	Microbacterium halotolerans	GCF_003569805.1	Microbacterium marinum	GCF_014204835.1
Massilia mucilaginosa	GCF_011682145.1	Microbacterium aurum	GCF_001974985.1	Microbacterium hatanonis	GCF_008017415.1	Microbacterium maritypicum MF109	GCF_000455825.1
Massilia niastensis DSM 21313	GCF_000382345.1	Microbacterium azadirachtae	GCF_900114965.1	Microbacterium helvum	GCF_014779795.1	Microbacterium mitrae	GCF_008017445.1
Massilia norwichensis	GCF_024753245.1	Microbacterium barkeri	GCF_027921765.1	Microbacterium hibisci	GCF_015278255.1	Microbacterium murale	GCF_014635185.1
Massilia oculi	GCF_003143515.1	Microbacterium binotii	GCF_021398715.1	Microbacterium hominis	GCF_014725695.1	Microbacterium nanhaiense	GCF_014646015.1
Massilia phyllostachyos	GCF_021044825.1	Microbacterium bovistercoris	GCF_003387575.1	Microbacterium humi	GCF_900105715.1	Microbacterium oleivorans	GCF_001975955.2
Massilia pinisoli	GCF_024753285.1	Microbacterium caowuchunii	GCF_008727755.1	obacterium hydrocarbonoxydans NBRC 10	GCF_001552455.1	Microbacterium oryzae	GCF_009735645.1
Massilia polaris	GCF_012927275.1	Microbacterium chengjingii	GCF_011751765.2	Microbacterium hydrothermale	GCF_004854025.1	Microbacterium oxydans	GCF_006540085.1
Massilia psychrophila	GCF_002760665.1	Microbacterium chocolatum	GCF_001278365.1	Microbacterium immunditiarum	GCF_013409785.1	Microbacterium paludicola	GCF_001887285.1
Massilia puerhi	GCF_009861485.1	Microbacterium commune	GCF_014836945.1	Microbacterium imperiale	GCF_027921865.1	Microbacterium paraoxydans	GCF_900105335.1
Massilia putida	GCF_001941825.1	Microbacterium cremeum	GCF_015277855.1	Microbacterium indicum DSM 19969	GCF_000422385.1	Microbacterium paulum	GCF_015707995.1
Massilia rhizosphaerae	GCF_016093545.1	Microbacterium dextranolyticum	GCF_027921925.1	Microbacterium invictum	GCF_023155715.1	Microbacterium phyllosphaerae	GCF_024809135.1
Massilia rubra	GCF_011682065.1	Microbacterium endense	GCF_900096885.1	Microbacterium jejuense	GCF_019511665.1	Microbacterium profundi	GCF_000763375.1
Massilia soli	GCF_016809835.1	Microbacterium endophyticum	GCF_011047135.1	Microbacterium keratanolyticum	GCF_027921965.1	Microbacterium protaetiae	GCF_004135285.1
Massilia solisilvae	GCF_024753215.1	Microbacterium esteraromaticum	GCF_016907315.1	Microbacterium ketosireducens	GCF_000956575.1	Microbacterium proteolyticum	GCF_014192415.1
Massilia terrae	GCF_024753145.1	Microbacterium excoecariae	GCF_011326725.1	Microbacterium kitamiense	GCF_002872075.1	Microbacterium pseudoresistens	GCF_013409745.1
Massilia timonae	GCF_001866515.1	Microbacterium faecale	GCF_014640975.1	Microbacterium kunmingense	GCF_023155745.1	Microbacterium pullorum	GCF_014836535.1
Massilia violaceinigra	GCF_002752675.1	Microbacterium fandaimingii	GCF_011751745.2	Microbacterium kyungheense	GCF_006783905.1	Microbacterium pygmaeum	GCF_900100885.1
Massilia yuzhufengensis	GCF_900112225.1	Microbacterium flavescens	GCF_018588945.1	Microbacterium lacticum	GCF_006716815.1	Microbacterium radiodurans	GCF_008710705.1

Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession
Microbacterium resistens	GCF_021172025.1	Nocardioides allogilvus	GCF_003047285.1	Nocardioides exalbidus	GCF_900105585.1	Nocardioides luteus	GCF_027921805.1
Microbacterium rhizomatis	GCF_008710745.1	Nocardioides alpinus	GCF_900112035.1	Nocardioides ferulae	GCF_003660455.1	Nocardioides luti	GCF_014212315.1
Microbacterium saccharophilum	GCF_007992455.1	Nocardioides anomalus	GCF_011046535.1	Nocardioides flavescens	GCF_009823805.1	Nocardioides malaquae	GCF_015070385.1
Microbacterium saperdae	GCF_006716345.1	Nocardioides antri	GCF_008373765.1	Nocardioides flavus Wang et al. 2016	GCF_014653115.1	Nocardioides mangrovi	GCF_020073815.1
Microbacterium schleiferi	GCF_015565955.1	Nocardioides aquaticus	GCF_018459925.1	Nocardioides furvisabuli	GCF_021083185.1	Nocardioides marinisabuli	GCF_013466785.1
Microbacterium sediminis	GCF_004564075.1	Nocardioides aromaticivorans	GCF_017309995.1	Nocardioides ganghwensis	GCF_004137235.1	Nocardioides marinus	GCF_013408145.1
Microbacterium sorbitolivorans	GCF_003327285.1	Nocardioides astragali	GCF_024436375.1	Nocardioides gansuensis	GCF_003076135.1	Nocardioides massiliensis	GCF_900067705.2
Microbacterium stercoris	GCF_017592545.1	Nocardioides baculatus	GCF_016735675.1	Nocardioides gilvus	GCF_003194605.1	Nocardioides mesophilus	GCF_014395785.1
Microbacterium suaedae	GCF_003254645.1	Nocardioides baekrokdamisoli	GCF_003945325.1	Nocardioides ginsengisegetis	GCF_014138045.1	Nocardioides nematodiphilus	GCF_020216525.1
Microbacterium sulfonylureivorans	GCF_003999995.1	Nocardioides bruguierae	GCF_023656415.1	Nocardioides glacieisoli	GCF_004137245.1	Nocardioides nitrophenolicus	GCF_016907515.1
Microbacterium telephonicum	GCF_003651225.1	Nocardioides caeni	GCF_004912195.1	Nocardioides guangzhouensis	GCF_004519545.1	Nocardioides ochotonae	GCF_011420305.2
Microbacterium tenebrionis	GCF_020682665.1	Nocardioides campestrisoli	GCF_013624435.2	Nocardioides halotolerans DSM 19273	GCF_000422805.1	Nocardioides okcheonensis	GCF_020991065.1
Microbacterium terrae	GCF_027921945.1	Nocardioides carbamazepini	GCF_024614185.1	Nocardioides houyundeii	GCF_003333285.1	Nocardioides oleivorans	GCF_004137255.1
Microbacterium terricola	GCF_025758395.1	Nocardioides cavernae	GCF_016907475.1	Nocardioides humi	GCF_006494775.1	Nocardioides palaemonis	GCF_018275325.1
Microbacterium testaceum	GCF_006539145.1	Nocardioides cavernaquae	GCF_003600895.1	Nocardioides humilatus	GCF_008373755.1	Nocardioides panacis	GCF_019039255.1
Microbacterium thalassium	GCF_027922065.1	Nocardioides coralli	GCF_019880385.1	Nocardioides hwasunensis	GCF_014779665.1	Nocardioides panacisoli	GCF_019448235.1
Microbacterium timonense	GCF_900292075.1	Nocardioides currus	GCF_003057875.1	Nocardioides immobilis	GCF_003515065.1	Nocardioides panaciterrulae	GCF_013409645.1
Microbacterium trichothecenolyticum	GCF_000956465.1	Nocardioides cynanchi	GCF_008761635.1	Nocardioides insulae DSM 17944	GCF_000422825.1	Nocardioides panthologis	GCF_003710085.1
Microbacterium ulmi	GCF_011759705.1	Nocardioides daedukensis	GCF_013408415.1	Nocardioides iriomotensis	GCF_004168035.1	Nocardioides panzhihuensis	GCF_013408335.1
Microbacterium ureisolvens	GCF_015278315.1	Nocardioides daeguensis	GCF_024612375.1	Nocardioides islandensis	GCF_015352455.1	Nocardioides pelophilus	GCF_014180685.1
Microbacterium wangchenii	GCF_004564355.1	Nocardioides daejeonensis	GCF_003194585.1	Nocardioides jejuensis	GCF_004337475.1	Nocardioides perillae	GCF_013409425.1
Microbacterium yannicii	GCF_024055635.1	Nocardioides daphniae	GCF_014635145.1	ocardioides jensenii JCM 1364 = NBRC 147	GCF_001552535.1	Nocardioides phosphati	GCF_014646375.1
Nocardioides acrostichi	GCF_015319165.1	Nocardioides deserti	GCF_014646035.1	Nocardioides jishulii	GCF_005217835.1	Nocardioides piscis	GCF_011300215.1
Nocardioides agariphilus	GCF_015352445.1	Nocardioides dokdonensis FR1436	GCF_001653335.1	Nocardioides kongjuensis	GCF_013409625.1	Nocardioides plantarum	GCF_006346395.1
Nocardioides albertanoniae	GCF_006716315.1	Nocardioides donggukensis	GCF_014842875.1	Nocardioides kribbensis	GCF_015070375.1	Nocardioides psychrotolerans	GCF_007992415.1
Nocardioides albidus	GCF_006335005.1	Nocardioides dongkuii	GCF_014127485.1	Nocardioides lacusdianchii	GCF_020102855.1	Nocardioides rotundus	GCF_019931675.1
Nocardioides albus	GCF_014648255.1	Nocardioides dongxiaopingii	GCF_005250675.1	Nocardioides lianchengensis	GCF_900101465.1	Nocardioides rubriscoriae	GCF_008087085.1
Nocardioides alcanivorans	GCF_021555175.1	Nocardioides eburneiflavus	GCF_004785795.1	Nocardioides lijunqiniae	GCF_015024265.1	Nocardioides salarius	GCF_016907435.1
Nocardioides alkalitolerans DSM 16699	GCF_000426525.1	Nocardioides euryhalodurans	GCF_004564375.1	Nocardioides litoris	GCF_006346315.1	Nocardioides sambongensis	GCF_006494815.1
Organism Scientific Name	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**
Nocardioides sediminis	GCF_003047295.1	Pseudomonas alvandae	GCF_019141525.1	Pseudomonas bijieensis	GCF_023330585.1	Pseudomonas citronellolis	GCF_001586155.1
Nocardioides seonyuensis	GCF_004683965.1	domonas amygdali pv. tabaci str. ATCC 11	GCF_000145945.2	Pseudomonas boanensis	GCF_018704125.1	Pseudomonas coleopterorum	GCF_900105555.1
Nocardioides silvaticus	GCF_003160695.1	Pseudomonas anatoliensis	GCF_017976235.1	Pseudomonas bohemica	GCF_002934685.1	Pseudomonas composti	GCF_900115475.1
Nocardioides soli	GCF_014191495.1	Pseudomonas anguilliseptica	GCF_900105355.1	Pseudomonas borbori	GCF_900115555.1	Pseudomonas congelans	GCF_900103225.1
Nocardioides solisilvae	GCF_003194625.1	Pseudomonas antarctica	GCF_900103795.1	Pseudomonas botevensis	GCF_019145475.1	eudomonas coronafaciens pv. oryzae str. 1	GCF_000156995.2
Nocardioides speluncae	GCF_003121585.1	Pseudomonas anuradhapurensis	GCF_014269225.2	Pseudomonas brassicacearum	GCF_008370715.1	Pseudomonas corrugata	GCF_001708425.1
Nocardioides stalactiti	GCF_014180715.1	Pseudomonas arcuscaelestis	GCF_016865555.1	Pseudomonas brassicae	GCF_010671725.1	Pseudomonas costantinii	GCF_900105935.1
Nocardioides szechwanensis	GCF_900103935.1	Pseudomonas argentinensis	GCF_900113905.1	Pseudomonas brenneri	GCF_900105815.1	monas cremoricolorata DSM 17059 = NBR	GCF_000425745.1
Nocardioides terrae	GCF_900112345.1	Pseudomonas aromaticivorans	GCF_019097855.1	Pseudomonas bubulae	GCF_900618535.1	Pseudomonas cremoris	GCF_014230465.1
Nocardioides terrigena	GCF_003047265.1	Pseudomonas arsenicoxydans	GCF_900103875.1	Pseudomonas californiensis	GCF_021147765.1	Pseudomonas crudilactis	GCF_013433315.1
Nocardioides thalensis	GCF_013410655.1	Pseudomonas asgharzadehiana	GCF_019139815.1	Pseudomonas campi	GCF_013200955.2	Pseudomonas cuatrocienegasensis	GCF_900110925.1
Nocardioides ungokensis	GCF_013466985.1	Pseudomonas asiatica	GCF_009932335.1	Pseudomonas canadensis	GCF_000503215.1	Pseudomonas cyclaminis	GCF_015163715.1
Nocardioides yefusunii	GCF_004023645.1	Pseudomonas asplenii	GCF_900105475.1	Pseudomonas canavaninivorans	GCF_024749365.1	Pseudomonas daroniae	GCF_004327275.1
Nocardioides zeae	GCF_010686755.1	Pseudomonas asturiensis	GCF_900143095.1	Pseudomonas cannabina pv. alisalensis	GCF_016599635.1	Pseudomonas deceptionensis	GCF_900106095.1
Nocardioides zhouii	GCF_004137345.1	Pseudomonas asuensis	GCF_014646755.1	Pseudomonas capeferrum	GCF_000731675.1	Pseudomonas defluvii	GCF_001695625.1
Pseudomonas abietaniphila	GCF_900100795.1	Pseudomonas atacamensis	GCF_017167965.1	Pseudomonas capsici	GCF_025792415.1	Pseudomonas delhiensis	GCF_900187975.1
Pseudomonas abyssi	GCF_002307495.1	Pseudomonas atagonensis	GCF_011369485.1	Pseudomonas carbonaria	GCF_904061905.1	Pseudomonas donghuensis	GCF_000696345.2
Pseudomonas aegrilactucae	GCF_019168305.1	Pseudomonas avellanae	GCF_000452845.1	Pseudomonas caricapapayae	GCF_001400735.1	Pseudomonas dryadis	GCF_004327225.1
Pseudomonas aeruginosa PAO1	GCF_000006765.1	Pseudomonas aylmerensis	GCF_001702265.1	Pseudomonas carnis	GCF_024106375.1	Pseudomonas duriflava	GCF_007830155.1
Pseudomonas agarici	GCF_001543125.1	Pseudomonas azadiae	GCF_019145355.1	Pseudomonas caspiana	GCF_002158995.1	Pseudomonas edaphica	GCF_013385965.1
Pseudomonas agronomica	GCF_025917275.1	Pseudomonas azerbaijanoccidens	GCF_019145495.1	Pseudomonas cavernae	GCF_003595175.1	Pseudomonas ekonensis	GCF_019145435.1
Pseudomonas akapageensis	GCF_011355085.1	Pseudomonas azerbaijanoriens	GCF_019139795.1	Pseudomonas cavernicola	GCF_003596405.1	Pseudomonas endophytica	GCF_001411475.1
Pseudomonas alcaligenes	GCF_001597285.1	Pseudomonas azotoformans	GCF_900103345.1	Pseudomonas cedrina	GCF_900104915.1	Pseudomonas entomophila L48	GCF_000026105.1
Pseudomonas alcaliphila NBRC 102411	GCF_002091495.1	Pseudomonas baetica	GCF_002813455.1	Pseudomonas cerasi	GCF_027983475.1	Pseudomonas eucalypticola	GCF_013374995.1
Pseudomonas alkylphenolica	GCF_009755645.1	Pseudomonas baltica	GCF_014235765.1	Pseudomonas ceruminis	GCF_021602025.1	Pseudomonas extremaustralis	GCF_900102035.1
Pseudomonas allii	GCF_013392005.1	Pseudomonas bananamidigenes	GCF_001679645.1	Pseudomonas chengduensis	GCF_000465575.1	Pseudomonas extremorientalis	GCF_900104365.1
Pseudomonas alliivorans	GCF_023955735.1	Pseudomonas baturnici	GCF_000820515.1	Pseudomonas chlororaphis	GCF_014524625.1	Pseudomonas fakonensis	GCF_019139895.1
Pseudomonas allokribbensis	GCF_014863605.1	Pseudomonas benzenivorans	GCF_900100495.1	Pseudomonas cichorii JBC1	GCF_000517305.1	Pseudomonas farris	GCF_019145235.1
Pseudomonas alloputida	GCF_021282585.1	Pseudomonas bharatica CSV86	GCF_000319305.2	Pseudomonas citri	GCF_025336485.1	Pseudomonas farsensis	GCF_014268805.2

Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession
Pseudomonas faucium	GCF_021602585.1	Pseudomonas haemolytica	GCF_016405625.1	Pseudomonas knackmussii	GCF_009911755.1	Pseudomonas massiliensis	GCF_000826105.1
Pseudomonas ficuserectae	GCF_024584585.1	Pseudomonas hamedanensis	GCF_014268595.2	Pseudomonas koreensis	GCF_900101415.1	Pseudomonas matsuisoli	GCF_014647635.1
Pseudomonas fildesensis	GCF_001050345.1	Pseudomonas helleri	GCF_001043025.1	Pseudomonas kribbensis	GCF_003352185.1	Pseudomonas maumuensis	GCF_019139675.1
Pseudomonas fitomaticsae	GCF_021018765.1	Pseudomonas helmanticensis	GCF_004365765.1	Pseudomonas kurunegalensis	GCF_024520475.1	Pseudomonas mediterranea	GCF_900106005.1
Pseudomonas flavescens NBRC 103044	GCF_002091575.1	Pseudomonas huaxiensis	GCF_003231275.1	Pseudomonas kuykendallii	GCF_900106975.1	Pseudomonas meliae	GCF_001400515.1
Pseudomonas flexibilis	GCF_000806415.1	Pseudomonas humi	GCF_001748265.1	Pseudomonas lactis	GCF_000263675.1	Pseudomonas mendocina S5.2	GCF_000733715.2
Pseudomonas floridensis	GCF_002087235.1	Pseudomonas hunanensis	GCF_013393325.1	Pseudomonas lactucae	GCF_016937555.1	Pseudomonas mercuritolerans	GCF_025706525.1
Pseudomonas fluorescens	GCF_900215245.1	Pseudomonas hydrolytica	GCF_024205225.1	Pseudomonas lalkuanensis	GCF_008807375.1	Pseudomonas migulae	GCF_900106025.1
Pseudomonas fluvialis	GCF_014635625.1	Pseudomonas indica NBRC 103045	GCF_002091635.1	Pseudomonas lalucatii	GCF_018398425.1	Pseudomonas mohnii	GCF_900105115.1
Pseudomonas foliumensis	GCF_014357575.1	Pseudomonas indoloxydans	GCF_003052605.1	Pseudomonas laurentiana	GCF_014648275.1	Pseudomonas monsensis	GCF_014268495.2
Pseudomonas fragi	GCF_015476275.1	Pseudomonas inefficax	GCF_900277125.1	Pseudomonas laurylsulfatiphila	GCF_002934665.1	Pseudomonas monteilii	GCF_003671975.1
Pseudomonas frederiksbergensis	GCF_001874645.1	Pseudomonas insulae	GCF_016901015.1	Pseudomonas laurylsulfativorans	GCF_002906155.1	Pseudomonas moorei	GCF_900102045.1
Pseudomonas fulva	GCF_001186195.1	Pseudomonas iranensis	GCF_014268585.2	Pseudomonas leptonychotis	GCF_004920405.1	Pseudomonas moraviensis	GCF_900105805.1
Pseudomonas furukawaii	GCF_002355475.1	Pseudomonas iridis	GCF_025397885.1	Pseudomonas libanensis	GCF_001439685.1	Pseudomonas morbosilactucae	GCF_023241665.1
Pseudomonas fuscovaginae	GCF_900108595.1	Pseudomonas izuensis	GCF_002356535.2	Pseudomonas lijiangensis	GCF_018968705.1	Pseudomonas mosselii	GCF_002736065.1
Pseudomonas germanica	GCF_019614655.1	domonas japonica NBRC 103040 = DSM 22	GCF_000730585.1	Pseudomonas lini	GCF_900104735.1	Pseudomonas mucidolens	GCF_900106045.1
Pseudomonas gessardii	GCF_900101185.1	Pseudomonas jessenii	GCF_900104905.1	Pseudomonas linyingensis	GCF_900109175.1	Pseudomonas mucoides	GCF_015461845.1
Pseudomonas gingeri	GCF_013386115.1	Pseudomonas jilinensis	GCF_003586265.1	Pseudomonas lopnurensis	GCF_015070855.1	Pseudomonas multiresinivorans	GCF_012971725.1
Pseudomonas glycinae	GCF_001594225.2	Pseudomonas jinjuensis	GCF_900103845.1	Pseudomonas lundensis	GCF_015377145.2	Pseudomonas muyukensis	GCF_019139535.1
Pseudomonas gozinkensis	GCF_014863585.1	Pseudomonas juntendi	GCF_021560075.1	Pseudomonas lurida	GCF_001708485.1	Pseudomonas nabeulensis	GCF_004682045.1
Pseudomonas graminis	GCF_900111735.1	Pseudomonas kairouanensis	GCF_004682055.1	Pseudomonas lutea	GCF_000759445.1	Pseudomonas neuropathica	GCF_015461835.1
Pseudomonas granadensis	GCF_900105485.1	Pseudomonas karstica	GCF_009707515.1	Pseudomonas luteola	GCF_008693965.1	Pseudomonas neustonica	GCF_003797945.1
Pseudomonas gregormendelii	GCF_017114825.1	Pseudomonas kermanshahensis	GCF_029215585.1	Pseudomonas mandelii	GCF_019693255.1	Pseudomonas nicosulfuronedens	GCF_005877905.1
Pseudomonas grimontii	GCF_900101085.1	Pseudomonas khavaziana	GCF_019145205.1	Pseudomonas mangiferae	GCF_007109405.1	Pseudomonas nitritireducens	GCF_010994165.1
Pseudomonas guangdongensis	GCF_900105885.1	Pseudomonas khazarica	GCF_017915135.1	Pseudomonas mangrovi	GCF_003052585.1	Pseudomonas nitroreducens	GCF_029457395.1
Pseudomonas guariconensis	GCF_002864915.1	Pseudomonas khorasanensis	GCF_014268505.2	Pseudomonas marginalis	GCF_008692035.1	Pseudomonas nunensis	GCF_024296925.1
Pseudomonas guguanensis	GCF_900104265.1	Pseudomonas kielensis	GCF_014236655.1	Pseudomonas marianensis	GCF_022810315.1	Pseudomonas ogarae	GCF_002879715.1
Pseudomonas guineae	GCF_900113745.1	Pseudomonas kilonensis	GCF_900105635.1	Pseudomonas marincola	GCF_900682675.2	Pseudomonas oleovorans CECT 5344	GCF_000297075.2
Pseudomonas guryensis	GCF_014164785.1	Pseudomonas kitaguniensis	GCF_009296165.1	Pseudomonas marvdashtae	GCF_014268655.2	Pseudomonas oligotrophica	GCF_021726475.1
Organism Scientific Name	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**
Pseudomonas orientalis	GCF_003852045.1	Pseudomonas piscis	GCF_009380155.1	Pseudomonas saliphila	GCF_008638345.1	udomonas syringae pv. coryli str. NCPPB 4	GCF_000972175.1
Pseudomonas oryzae	GCF_900104805.1	Pseudomonas piscis	GCF_025628765.1	Pseudomonas salmasensis	GCF_014268375.2	Pseudomonas syringae pv. tagetis	GCF_022557255.1
Pseudomonas oryzagri	GCF_020831405.1	Pseudomonas plecoglossicida	GCF_003391255.1	Pseudomonas salomonii	GCF_013386885.1	eudomonas syringae pv. tomato str. DC30	GCF_000007805.1
Pseudomonas oryzicola	GCF_014269185.2	Pseudomonas poae	GCF_900103965.1	Pseudomonas saponiphila	GCF_900105185.1	Pseudomonas taeanensis MS-3	GCF_000498575.2
Pseudomonas oryzihabitans	GCF_007665635.1	Pseudomonas pohangensis	GCF_900105995.1	Pseudomonas saudimassiliensis	GCF_000939975.1	Pseudomonas taetrolens	GCF_900104825.1
Pseudomonas oryzihila	GCF_003940825.1	Pseudomonas profundi	GCF_008638305.1	Pseudomonas saudiphocaensis	GCF_003935525.1	Pseudomonas taiwanensis DSM 21245	GCF_000425785.1
Pseudomonas otitidis	GCF_011397855.1	Pseudomonas promysalinigenes	GCF_014269025.2	Pseudomonas savastanoi	GCF_020917325.1	Pseudomonas tehranensis	GCF_014268615.1
Pseudomonas ovata	GCF_003131185.1	Pseudomonas prosekii	GCF_900105155.1	Pseudomonas saxonica	GCF_007858365.1	Pseudomonas tensinigenes	GCF_014268445.2
Pseudomonas palleroniana	GCF_900105975.1	Pseudomonas protegens CHA0	GCF_900560965.1	Pseudomonas schmalbachii	GCF_017589465.1	Pseudomonas thermotolerans J53	GCF_000513835.1
Pseudomonas palmensis	GCF_017848315.1	Pseudomonas proteolytica	GCF_900105955.1	Pseudomonas sediminis	GCF_013409125.2	Pseudomonas thivervalensis	GCF_900102295.1
Pseudomonas panipatensis	GCF_900099785.1	Pseudomonas pseudonitroreducens	GCF_021603645.1	Pseudomonas segetis	GCF_900188155.1	Pseudomonas tohonis	GCF_012767755.2
Pseudomonas paracarnis	GCF_904063075.1	Pseudomonas psychrophila	GCF_011040435.1	Pseudomonas seleniipraecipitans	GCF_900102335.1	Pseudomonas tolaasii NCPPB 2192	GCF_002813445.1
Pseudomonas paraeruginosa	GCF_003025345.2	Pseudomonas psychrotolerans	GCF_001991015.1	Pseudomonas sessilinigenes	GCF_003850565.1	Pseudomonas toyotomiensis	GCF_017167985.1
udomonas parafulva NBRC 16636 = DSM 1	GCF_000730645.1	Pseudomonas pudica	GCF_015680405.1	Pseudomonas shahriarae	GCF_014268455.2	Pseudomonas tremae	GCF_023278125.1
Pseudomonas paraglycinae	GCF_021601625.1	Pseudomonas punonensis	GCF_900142655.1	Pseudomonas shirazensis	GCF_014268785.2	Pseudomonas tritici	GCF_014268275.3
Pseudomonas parakoreensis	GCF_021602155.1	Pseudomonas putida NBRC 14164	GCF_000412675.1	Pseudomonas shirazica	GCF_015168335.1	Pseudomonas triticicola	GCF_019145375.1
Pseudomonas paralactis	GCF_001439735.1	Pseudomonas qingdaonensis	GCF_018398475.1	Pseudomonas sichuanensis	GCF_024749065.1	Pseudomonas triticumensis	GCF_014358015.1
Pseudomonas paralcaligenes	GCF_019655855.2	Pseudomonas quasicaspiana	GCF_021147805.1	Pseudomonas sihuiensis	GCF_900106015.1	Pseudomonas trivialis	GCF_900104885.1
Pseudomonas parasichuanensis	GCF_021601385.1	Pseudomonas quercus	GCF_012033695.1	Pseudomonas silesiensis	GCF_001661075.1	Pseudomonas tructae	GCF_004214895.1
Pseudomonas paraversuta	GCF_904063085.1	Pseudomonas reactans	GCF_013385825.1	Pseudomonas siliginis	GCF_024072275.1	Pseudomonas tumuqii	GCF_013184545.1
Pseudomonas peli	GCF_900099645.1	Pseudomonas reidholzensis	GCF_900536025.1	Pseudomonas simiae	GCF_000934565.1	Pseudomonas turukhanskensis	GCF_027922365.1
Pseudomonas peradeniyensis	GCF_014268935.2	Pseudomonas reinekei	GCF_900104125.1	Pseudomonas sivasensis	GCF_021391435.1	Pseudomonas typographi	GCF_014694385.1
Pseudomonas pergaminensis	GCF_024112395.1	Pseudomonas resinovorans DSM 21078	GCF_000423545.1	Pseudomonas soli	GCF_000498975.2	Pseudomonas ullengensis	GCF_014174475.1
Pseudomonas petroselini	GCF_021166595.1	Pseudomonas rhizophila	GCF_003033885.1	Pseudomonas songnenensis	GCF_003696315.1	Pseudomonas umsongensis	GCF_012647205.1
Pseudomonas pharmacofabricae	GCF_002835605.1	Pseudomonas rhizoryzae	GCF_005250605.1	Pseudomonas spelaei	GCF_009724245.1	Pseudomonas urethralis	GCF_021602465.1
Pseudomonas pharyngis	GCF_021602345.1	Pseudomonas rhizosphaerae	GCF_000761155.1	Pseudomonas sputi	GCF_021603585.1	Pseudomonas urmiensis	GCF_014268815.2
Pseudomonas phenolilytica	GCF_021432765.1	Pseudomonas rhodesiae	GCF_013285305.1	Pseudomonas straminea	GCF_900112645.1	Pseudomonas uvaldensis	GCF_021271205.1
Pseudomonas phragmitis	GCF_002056295.1	Pseudomonas rustica	GCF_018336155.1	Pseudomonas synxantha	GCF_003851555.1	Pseudomonas vancouverensis	GCF_900105825.1
Pseudomonas pisciculturae	GCF_015461805.1	Pseudomonas sagittaria	GCF_900115715.1	Pseudomonas syringae	GCF_018394375.1	Pseudomonas vanderleydeniana	GCF_014268755.2

Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession	Organism Scientific Name	Assembly Accession
Pseudomonas veronii	GCF_002028325.1	Sphingomonas asaccharolytica NBRC 15495	GCF_001598355.1	Sphingomonas flavalba	GCF_004796535.1	Sphingomonas laterariae	GCF_900188165.1
Pseudomonas versuta	GCF_001294575.1	Sphingomonas astaxanthinifaciens DSM 222	GCF_000711715.1	Sphingomonas folli	GCF_019429525.1	Sphingomonas leidyi	GCF_011761945.1
Pseudomonas viciae	GCF_004786035.1	Sphingomonas aurantiaca	GCF_003050705.1	Sphingomonas formosensis	GCF_009755815.1	Sphingomonas lenta	GCF_002288825.1
Pseudomonas viridiflava	GCF_900184295.1	Sphingomonas azotifigens NBRC 15497	GCF_002091475.1	Sphingomonas gei	GCF_004792685.1	Sphingomonas lutea	GCF_021497585.1
Pseudomonas vlassakiae	GCF_025447535.1	Sphingomonas baiyangensis	GCF_005144715.1	Sphingomonas gellani	GCF_900110035.1	Sphingomonas mali NBRC 15500	GCF_001598415.1
Pseudomonas vranovensis DSM 16006	GCF_000425805.1	Sphingomonas bisphenolicum	GCF_024349785.1	Sphingomonas gilva	GCF_003515075.1	Sphingomonas melonis C3	GCF_000371765.1
Pseudomonas wadenswilerensis	GCF_900497695.1	Sphingomonas canadensis	GCF_026013525.1	Sphingomonas ginkgonis	GCF_003979025.1	Sphingomonas mesophila	GCF_003499275.1
Pseudomonas weihenstephanensis	GCF_016812185.1	Sphingomonas cannabina	GCF_021391395.1	Sphingomonas ginsengisoli An et al. 2013	GCF_009363895.1	Sphingomonas metalli	GCF_014641735.1
Pseudomonas wenzhouensis	GCF_021029445.1	Sphingomonas carotinifaciens	GCF_900102275.1	Sphingomonas ginsenosidimutans	GCF_002374835.1	Sphingomonas montana	GCF_001956315.1
Pseudomonas xantholysinigenes	GCF_014268885.2	Sphingomonas cavernae	GCF_003590775.1	Sphingomonas ginsenosidivorax	GCF_007995065.1	Sphingomonas montanisoli	GCF_008274695.1
Pseudomonas xanthosomae	GCF_019139835.1	Sphingomonas changbaiensis NBRC 104936	GCF_000974765.1	Sphingomonas glacialis	GCF_014653575.1	Sphingomonas morindae	GCF_023822065.1
Pseudomonas xionganensis	GCF_009763245.1	Sphingomonas changnyeongensis	GCF_009913435.1	Sphingomonas glaciei	GCF_023380025.1	Sphingomonas mucoissima	GCF_002197665.1
Pseudomonas yamanorum	GCF_001612705.2	Sphingomonas chungangi	GCF_013778325.1	Sphingomonas guangdongensis	GCF_900199185.1	Sphingomonas naasensis	GCF_004792695.1
Pseudomonas yangonensis	GCF_009932815.1	Sphingomonas citri	GCF_019429485.1	Sphingomonas hankookensis	GCF_022664465.1	Sphingomonas naphthae	GCF_028607085.1
Pseudomonas zanjanensis	GCF_014268745.2	Sphingomonas citricola	GCF_019429535.1	Sphingomonas hengshuiensis	GCF_000935025.1	Sphingomonas nostoxanthinifaciens	GCF_019930585.1
Pseudomonas zarinae	GCF_014268695.2	Sphingomonas colocasiae	GCF_019880585.1	Sphingomonas hominis	GCF_013328205.1	Sphingomonas olei	GCF_004801655.1
Pseudomonas zeae	GCF_014268485.2	Sphingomonas corticis	GCF_012035195.1	Sphingomonas horti	GCF_009753715.1	Sphingomonas oleivorans	GCF_003050615.1
Pseudomonas zeshuii	GCF_900141925.1	Sphingomonas crocodyli	GCF_004005865.1	Sphingomonas hylomeconis	GCF_025370105.1	Sphingomonas oligoaromativorans	GCF_011762195.1
Serpens gallinarum	GCF_014836765.1	Sphingomonas crusticola	GCF_003391115.1	Sphingomonas insulae	GCF_010450875.1	Sphingomonas oligophenolica	GCF_006438955.1
Sphingomonas abaci	GCF_014199625.1	Sphingomonas daechungensis	GCF_014489495.1	Sphingomonas japonica	GCF_006346325.1	Sphingomonas paeninsulae	GCF_003660165.1
Sphingomonas adhaesiva	GCF_002374855.1	Sphingomonas deserti	GCF_003012735.1	Sphingomonas jaspsi DSM 18422	GCF_000585415.1	Sphingomonas palmae	GCF_900109565.1
Sphingomonas aerolata	GCF_003046295.1	Sphingomonas desiccabilis	GCF_004135605.1	Sphingomonas jatrophae	GCF_900113315.1	Sphingomonas panacis	GCF_001717955.1
Sphingomonas aerophila	GCF_014199305.1	Sphingomonas dokdonensis	GCF_002197685.1	Sphingomonas jeddahensis	GCF_001981525.1	Sphingomonas panacisoli	GCF_007859635.1
Sphingomonas albertensis	GCF_014358075.1	Sphingomonas echinoides ATCC 14820	GCF_000241465.1	Sphingomonas jejuensis	GCF_011927695.1	Sphingomonas panni	GCF_022664435.1
Sphingomonas aliaeris	GCF_016743815.1	Sphingomonas edaphi	GCF_003583725.1	Sphingomonas jinjuensis	GCF_014197105.1	Sphingomonas parapaucimobilis NBRC 15100	GCF_000787715.1
Sphingomonas alpina	GCF_025370095.1	Sphingomonas elodea ATCC 31461	GCF_000226955.1	Sphingomonas kaistensis	GCF_011927725.1	Sphingomonas parva	GCF_004564275.1
Sphingomonas aquatilis	GCF_022760365.1	Sphingomonas endophytica	GCF_014199415.1	Sphingomonas koreensis	GCF_001922385.1	Sphingomonas paucimobilis	GCF_016027095.1
Sphingomonas aracearum	GCF_003345355.1	Sphingomonas faeni	GCF_003053745.1	Sphingomonas kyeonggiensis	GCF_014196745.1	Sphingomonas phyllosphaerae FA2	GCF_000427645.1
Sphingomonas arenae	GCF_016924655.1	Sphingomonas fennica	GCF_003034225.1	Sphingomonas lacunae	GCF_012979535.1	Sphingomonas piscis	GCF_011300455.1
Organism Scientific Name	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**	**Organism Scientific Name**	**Assembly Accession**
Sphingomonas pituitosa NBRC 102491	GCF_001598435.1	Sphingomonas xinjiangensis	GCF_014199255.1	Streptomyces cinnabarinus	GCF_027270315.1	Streptomyces harbinensis	GCF_013364095.1
Sphingomonas pokkalii	GCF_003096275.1	Sphingomonas yabuchiae	GCF_014199595.1	Streptomyces clavuligerus	GCF_015767775.1	Streptomyces hawaiiensis	GCF_004803895.1
Sphingomonas populi	GCF_004208535.1	Sphingomonas yantingensis	GCF_014199325.1	Streptomyces coeruleorubidus	GCF_008705135.1	Streptomyces huasconensis	GCF_021028635.1
Sphingomonas prati	GCF_009755805.1	Sphingomonas yunnanensis	GCF_019898765.1	Streptomyces collinus Tu 365	GCF_000444875.1	Streptomyces hundungensis	GCF_003627815.1
Sphingomonas profundi	GCF_009739515.1	Sphingomonas zeae	GCF_014197135.1	Streptomyces cyaneochromogenes	GCF_003963535.1	Streptomyces inhibens	GCF_022221585.1
Sphingomonas pruni NBRC 15498	GCF_001598455.1	Streptomyces akebiae	GCF_019599145.1	Streptomyces cyanogenus	GCF_017526105.1	Streptomyces kanamyceticus	GCF_008704495.1
Sphingomonas pseudosanguinis	GCF_014196255.1	Streptomyces albidoflavus	GCF_000359525.1	Streptomyces dangxiongensis	GCF_003675325.1	Streptomyces koyangensis	GCF_016803985.1
Sphingomonas psychrolutea	GCF_014636175.1	Streptomyces alboniger	GCF_008704395.1	Streptomyces deccanensis	GCF_022385335.1	Streptomyces lavendulae subsp. lavendulae	GCF_018255875.1
Sphingomonas psychrotolerans	GCF_002796605.1	Streptomyces albus	GCF_024752535.1	Streptomyces decoyicus	GCF_019880305.1	Streptomyces leeuwenhoekii	GCF_001013905.1
Sphingomonas qomolangmaensis	GCF_024496245.1	Streptomyces alfalfae	GCF_018141485.1	Streptomyces dengpaensis	GCF_002946835.1	Streptomyces liangshanensis	GCF_011694815.1
Sphingomonas radiodurans	GCF_020866845.1	Streptomyces ambofaciens ATCC 23877	GCF_001267885.1	Streptomyces durmitorensis	GCF_023498005.1	Streptomyces lienomycini	GCF_027947595.1
Sphingomonas rhizophila	GCF_014396585.1	Streptomyces angustmyceticus	GCF_019933235.1	Streptomyces fagopyri	GCF_009498275.1	Streptomyces lilliifuscus	GCF_016598615.1
Sphingomonas rubra	GCF_900115745.1	Streptomyces anulatus	GCF_020783455.1	Streptomyces ferrugineus	GCF_015160855.1	Streptomyces lincolnensis	GCF_003444445.1
Sphingomonas sabuli	GCF_014352855.1	Streptomyces aquilus	GCF_003955715.1	Streptomyces ficellus	GCF_009739905.1	Streptomyces lunaelactis	GCF_003054555.1
Sphingomonas sanguinis NBRC 13937	GCF_001591005.1	Streptomyces atratus	GCF_003330865.1	Streptomyces fodineus	GCF_001735805.1	Streptomyces lusitanus	GCF_025402955.1
ingomonas sanxanigenens DSM 19645 = N	GCF_000512205.2	Streptomyces aurantiacus	GCF_014701115.1	Streptomyces formicae	GCF_022647665.1	Streptomyces lydicus	GCF_001984445.1
Sphingomonas sediminicola	GCF_014489515.1	Streptomyces auratus AGR0001	GCF_000280865.2	eptomyces fradiae ATCC 10745 = DSM 40	GCF_008704425.1	Streptomyces marincola	GCF_002128305.1
Sphingomonas segetis	GCF_009720245.1	eptomyces avermitilis MA-4680 = NBRC 14	GCF_000009765.2	Streptomyces fungicidicus	GCF_003665435.1	Streptomyces microflavus	GCF_013364315.1
Sphingomonas sinipercae	GCF_011302055.1	Streptomyces bathyalis	GCF_015910445.1	Streptomyces galilaeus	GCF_008704575.1	Streptomyces milbemycinicus	GCF_025908395.1
Sphingomonas solaris	GCF_007785815.1	Streptomyces broussonetiae	GCF_009796285.1	Streptomyces gardneri	GCF_015710995.1	Streptomyces mobaraensis	GCF_020099395.1
Sphingomonas soli NBRC 100801	GCF_001591025.1	Streptomyces buecherae	GCF_014295035.1	Streptomyces genisteinicus	GCF_014489615.1	Streptomyces murinus	GCF_016467295.1
Sphingomonas spermidinifaciens	GCF_002351485.1	Streptomyces cadmiisoli	GCF_003261055.1	Streptomyces gilvosporeus	GCF_002082195.1	Streptomyces nigra	GCF_003074055.1
Sphingomonas suaedae	GCF_007833215.1	Streptomyces californicus	GCF_016906185.1	Streptomyces globosus	GCF_003325375.1	Streptomyces nigrescens	GCF_027626975.1
Sphingomonas taxi	GCF_000764535.1	Streptomyces calvus	GCF_006716135.1	Streptomyces gobiensis	GCF_021216675.1	Streptomyces nitrosporeus	GCF_008704555.1
Sphingomonas trueperi	GCF_011927635.1	Streptomyces caniscabiei	GCF_014852565.2	Streptomyces griseochromogenes	GCF_001542625.1	Streptomyces nodosus	GCF_017352335.1
Sphingomonas turrisvirgatae	GCF_001721295.1	Streptomyces cavourensis	GCF_002804165.1	Streptomyces griseofuscus	GCF_014673495.1	Streptomyces nojiriensis	GCF_017639205.1
Sphingomonas ursincola	GCF_013607875.1	Streptomyces chartreusis	GCF_008704715.1	Streptomyces griseorubiginosus	GCF_003595235.1	Streptomyces noursei	GCF_016917755.1
Sphingomonas vulcanisoli	GCF_011761305.1	Streptomyces chromofuscus	GCF_015160875.1	Streptomyces griseoviridis	GCF_003994395.1	Streptomyces olivaceus	GCF_015244315.1
Sphingomonas xanthus	GCF_007998985.1	Streptomyces cinereoruber	GCF_009299385.1	eptomyces griseus subsp. griseus NBRC 13	GCF_000010605.1	reptomyces olivoreticuli subsp. olivoretic	GCF_003391135.1
Streptomyces paludis	GCF_003344965.1	Streptomyces qaidamensis	GCF_001611795.1	Streptomyces spectabilis	GCF_008704795.1	Streptomyces violaceoruber	GCF_002082175.1
Streptomyces parvulus	GCF_001660045.1	Streptomyces qinzhouensis	GCF_007856155.1	Streptomyces spongiicola	GCF_003122365.1	Streptomyces viridosporus T7A	GCF_008704515.1
Streptomyces phaeolivaceus	GCF_009184865.1	Streptomyces rectiverticillatus	GCF_013394065.1	Streptomyces subrutilus	GCF_008704535.1	Streptomyces xanthii	GCF_014621695.1
Streptomyces phaeoluteigriseus	GCF_023923245.1	ptomyces rimosus subsp. rimosus ATCC 10	GCF_000331185.2	Streptomyces tanashiensis	GCF_026427415.1	Streptomyces xanthophaeus	GCF_028421465.1
Streptomyces platensis	GCF_008704855.1	Streptomyces roseirectus	GCF_014489635.1	Streptomyces tirandamycinicus	GCF_003097515.1	Streptomyces xiamenensis	GCF_000993785.3
Streptomyces pluripotens	GCF_000802245.2	Streptomyces rutgersensis	GCF_014216335.1	Streptomyces tsukubensis	GCF_003932715.1	Streptomyces xinghaiensis S187	GCF_000220705.2
Streptomyces prasinus	GCF_008704445.1	Streptomyces scabiei 87.22	GCF_000091305.1	Streptomyces tubbatahanensis	GCF_022699385.1	Streptomyces yanglinensis	GCF_024666385.1
Streptomyces pratensis	GCF_016804005.1	Streptomyces seoulensis	GCF_004328625.1	Streptomyces tuirus	GCF_014701095.1	Streptomyces yatensis	GCF_018069625.1
Streptomyces pristinaespiralis	GCF_001278075.1	Streptomyces solisilvae	GCF_015689475.1	Streptomyces venezuelae	GCF_001886595.1	Streptomyces vietnamensis	GCF_000830005.1

Table S4. Detailed description of the processing of the sequence datasets. Table includes information such as sample type, the number of merged reads, indexing PCR cycles, mapped sequences on plant genomes (Table S2), mapped reads on the h28 human reference genome, total numbers, percentages of mapped reads, and unmapped reads.

ID	Matrix	Merge seqs (before mapping)	Plant mapped Seqs	Human mapped Seqs	Total mapped Seqs	Total unmapped Seqs	Percentage mapped	Percentage unmapped
As_0166	Rhizospheric Soil	17336361	661262	8574	669836	16666525	3.863763566	96.13623643
As_0166	Root	31221262	15239658	14126	15253784	15967478	48.85703851	51.14296149
As_205	Rhizospheric Soil	20463286	6386169	35607	6421776	14041510	31.38193934	68.61806066
As_205	Root	23850577	18205374	5224	18210598	5639979	76.3528614	23.6471386
As_225	Rhizospheric Soil	26541264	3974705	8557	3983262	22558002	15.00780822	84.99219178
As_225	Root	19116070	7330070	12455	7342525	11773545	38.41022239	61.58977761
As_FH	Rhizospheric Soil	13293613	6621323	5938	6627261	6666352	49.85297075	50.14702925
As_FH	Root	26765683	23001060	2762	23003822	3761861	85.94520827	14.05479173
Ls_321	Root	39777561	21247151	29552	21276703	18500858	53.48920966	46.51079034
Ls_358	Root	24240906	12835256	17972	12853228	11387678	53.02288619	46.97711381
Ls_378	Rhizospheric Soil	14114408	5146825	13243	5160068	8954340	36.55886949	63.44113051
Ls_378	Root	20886175	5031499	44988	5076487	15809688	24.30548916	75.69451084
Ls_389	Rhizospheric Soil	17659989	1792955	16249	1809204	15850785	10.24464964	89.75535036
Ls_389	Root	20436865	14906923	9925	14916848	5520017	72.98990329	27.01009671
Ls_390	Leaf	24731271	21903879	10917	21914796	2816475	88.61168518	11.38831482
Ls_390	Rhizospheric Soil	16335140	2900464	161800	3062264	13272876	18.74648151	81.25351849
Ls_390	Root	25372586	15110630	15708	15126338	10246248	59.61685577	40.38314423
Sc_637	Rhizospheric Soil	18514951	7030561	4662	7035223	11479728	37.99752427	62.00247573
Sc_637	Root	24809146	17305371	5702	17311073	7498073	69.77698063	30.22301937
Sc_680	Rhizospheric Soil	13570350	4284689	6942	4291631	9278719	31.62505757	68.37494243
Sc_680	Root	29012106	20678504	4060	20682564	8329542	71.28942656	28.71057344
Sc_684	Leaf	20767518	18031055	1571	18032626	2734892	86.83091547	13.16908453
Sc_684	Rhizospheric Soil	15222638	1596660	11532	1608192	13614446	10.56447641	89.43552359
Sc_684	Root	28472297	18517131	9294	18526425	9945872	65.06824862	34.93175138
Sc_689	Rhizospheric Soil	8566991	1862634	386484	2249118	6317873	26.25330177	73.74669823
Sc_689	Root	30248982	24674832	4712	24679544	5569438	81.58801509	18.41198491
Sc_707	Rhizospheric Soil	8515366	2665406	4423	2669829	5845537	31.35307396	68.64692604
Sc_707	Root	12507885	5120958	5466	5126424	7381461	40.98553832	59.01446168
Td_270	Rhizospheric Soil	12972221	1827549	3743	1831292	11140929	14.11702745	85.88297255
Td_270	Root	26976280	19709544	9021	19718565	7257715	73.09593836	26.90406164
Td_279	Rhizospheric Soil	16195887	4457170	7612	4464782	11731105	27.56738177	72.43261823
Td_279	Root	40313091	28567253	13778	28581031	11732060	70.89764216	29.10235784
Td_283	Leaf	18918573	16504141	3249	16507390	2411183	87.25494254	12.74505746
Td_283	Rhizospheric Soil	18908106	2135956	8824	2144780	16763326	11.34317737	88.65682263
Td_283	Root	28551986	19085380	7306	19092686	9459300	66.8699053	33.1300947
Td_284	Rhizospheric Soil	16131606	1937337	11744	1949081	14182525	12.08237419	87.91762581
Td_284	Root	32734421	19311361	17268	19328629	13405792	59.04680275	40.95319725
Td_297	Rhizospheric Soil	14721660	5597409	71688	5669097	9052563	38.50854455	61.49145545
Td_297	Root	32341450	21766796	8948	21775744	10565706	67.33075975	32.66924025

Table S5. PerMANOVA test of Eukarya and Bacteria beta diversity (phylum and genus levels). The table presents all the parameters obtained from PerMANOVA tests on Eukarya and Bacteria beta diversity matrices. Microbial communities, both at genus and phylum level, were significantly diverse among sample types (root and rhizospheric soil), geographic origins and plant species. In the tables are reported degree of freedom (df), sum of squares, coefficients of determination (R2), F statistic value (F), P-value (Pr(>F)).

Bacteira + Eukarya - Phylum level						
	Df	**Sum of squares**	**R2**	**F**	**Pr(>F)**	
Species	3	0.05114	0.13594	5.41	0.001	***
Matrix	1	0.13773	0.36611	43.71	0.001	***
Origins	3	0.11043	0.29356	11.638	0.001	***
Residual	28	0.08823	0.23452			
Total	35	0.37619	1			
Bacteria - Genus level						
	Df	**Sum of squares**	**R2**	**F**	**Pr(>F)**	
Species	3	0.06611	0.10404	3.2938	0.004	**
Matrix	1	0.23884	0.37585	35.6964	0.001	***
Origins	3	0.12899	0.20299	6.4265	0.001	***
Residual	28	0.18734	0.29482			
Total	35	0.63545	1			
Eukarya - Genus level						
	Df	**Sum of squares**	**R2**	**F**	**Pr(>F)**	
Species	3	0.06611	0.10404	3.2938	0.004	**
Matrix	1	0.23884	0.37585	35.6964	0.001	***
Origins	3	0.12899	0.20299	6.4265	0.001	***
Residual	28	0.18734	0.29482			
Total	35	0.63545	1			

Table S6. PerMANOVA test of Eukarya and Bacteria beta diversity of root microbiomes.

The table presents all the parameters obtained from PerMANOVA tests on Eukarya and Bacteria beta diversity matrices reduced to root microbiome samples. Root microbiome communities, at genus level, were significantly different among sample types (root and rhizospheric soil) geographic, origins and plant species. In the tables are reported degree of freedom (df), sum of squares, coefficients of determination (R2), F statistic value (F), P-value (Pr(>F)).

Genus level - Bacteria - Root samples						
	Df	**Sum of squares**	**R2**	**F**	**Pr(>F)**	
Species	3	0.075207	0.26325	2.7259	0.002	**
Origins	3	0.100787	0.35279	3.6531	0.001	***
Residual	12	0.110359	0.38629			
Total	18	0.285685	1			
Genus level - Eukarya - Root samples						
	Df	**Sum of squares**	**R2**	**F**	**Pr(>F)**	
Species	3	0.15843	0.28732	3.9712	0.001	***
Origins	3	0.2461	0.44629	6.1685	0.001	***
Residual	12	0.15958	0.2894			
Total	18	0.55143	1			

Table S7. PerMANOVA test on simka results of herbarium metagenomics datasets.

The table presents all the parameters obtained from PerMANOVA tests of output file obtained from simka. In the tables are reported degree of freedom (df), sum of squares, coefficients of determination (R2), F statistic value (F), P-value (Pr(>F)).

	Df	Sum of squares	R2	F	Pr(>F)	
Species	3	2.6318	0.15945	2.3758	0.001	***
Matrix	2	1.3408	0.08123	1.8155	0.001	***
Origins	3	1.379	0.08354	1.2448	0.002	**
Residual	30	11.0776	0.67113			
Total	38	16.5059	1			

Chapter 6. Future perspectives: herbarium soils to investigate the consequences of agriculture intensification on the soil microbiome

In the previous chapters of this manuscript, we demonstrated that soil and herbarium roots represent an exceptional resource for describing ancient soil microbial communities (up to *ca.* two centuries-old). In this final chapter, we will present future perspectives that could be addressed using this material.

A central theme will be the functional annotation of historical DNA and the assembly of sequences to reconstruct the entire genomes of microbial microorganisms of the past (so-called MAGs, Metagenome assembled genomes), thus paving the way for the functional analysis of soil communities over time. Subsequently, we will discuss the "HerbaSoil" project, which aims to extend the work of the present thesis by using herbarium soil samples collected over approximately 250 years to analyze temporal changes in microbial communities. The project will focus on identifying the impacts of intensive farming on microorganisms, using both herbarium soil samples and modern agricultural ones from the same geographic areas where the herbarium plants were collected several decades earlier.

These approaches will not only provide a better understanding of the evolution of microbial communities in response to agricultural practices but may also provide insights into the sustainable management of agricultural resources in the future.

Assembly of historical microbial genomes associated to herbarium plants

Molecular paleomicrobiology is revolutionizing microbiology by inserting microorganisms into a historical framework, what classical microbiology could not do. This groundbreaking discipline not only unveils the mysteries of the past evolution of microorganisms but also serves as an entry point to exploring uncharted territories of the microbial realm. In this respect, the *de novo* assembly of microbial genomes from historical samples allows for the investigation of species within a more complex microbial community (such as pathogens or symbionts) and the exploration of their functions and characteristics. This has already been performed for pathogens associated to herbarium plants; this resulted in the partial assembly of the genomes of the Oomycete *P. infestans*, the causal agent of Potato late blight and in the assembly of the full-length genome of the bacterium *Xanthomonas citri*, agent of the Citrus canker disease (Campos

et al., 2023; Hartung et al., 2015; Yoshida et al., 2013). Besides these examples that targeted the cited pathogens, applying this approach to dental calculus samples (not yet to plants) enabled unveiling species previously unknown to science, representing a new milestone for the discipline. One particularly fascinating result is the discovery of a secondary metabolite biosynthetic gene cluster present in the oral microbiomes of both Neanderthals and modern humans who lived during the Middle and Upper Paleolithic periods (Klapper et al., 2023). Additionally, a large-scale *de novo* assembly of microbial genomes from palaeofaeces has led to the discovery and characterization of previously undescribed gut microorganisms from ancient microbiomes and the investigation of the evolutionary history of the human gut microbiota (Wibowo et al., 2021).

In herbarium specimens, the use of assembly could be very useful for studying the structure and evolution of the plant genome (identifying genetic modifications or specific cultivars) and also for studying individual microorganisms in the associated microbiome. In the case of plants, studies that indeed use herbarium plants or archeological plant remains to reconstruct plant genomes to trace plant diversity and recent plant evolution and adaptation to new environments have been published (Goloubinoff et al., 1993; Gutaker et al., 2017; Mazo et al., 2012). Using the data obtained in Chapter 5, an assembly of DNA sequences was performed and resulted in the partial or full-length assembly of chloroplast genomes of *Avena, Secale, Lactuca,* and *Triticum*. Complete reconstruction of the chloroplast genomes was achieved for leaf and for several, but not all, root DNA extracts (**Fig. 1**). In the case of soil DNA, reconstruction proved more complex, resulting in only fragmented genome assemblies. This is likely due to the presence of fewer plant DNA sequences in soil and root samples. Regarding microorganisms, a de novo assembly was also attempted to obtain genomic fragments of microbial origin. However, the analysis only allowed to obtain small genomic fragments of microorganisms without the possibility of obtaining whole microbes genomes or genes.

In the future, deeper sequencing of herbarium soil samples may allow for the reconstruction of genomic fragments, or even entire genomes, from the most abundant microbial species. However, the high microbial richness of soils and the fragmentation of DNA during the conservation period in herbaria will make this process challenging. It will be necessary to optimize methods to obtain intact microbial genomes and rely on curated and comprehensive genomic databases.

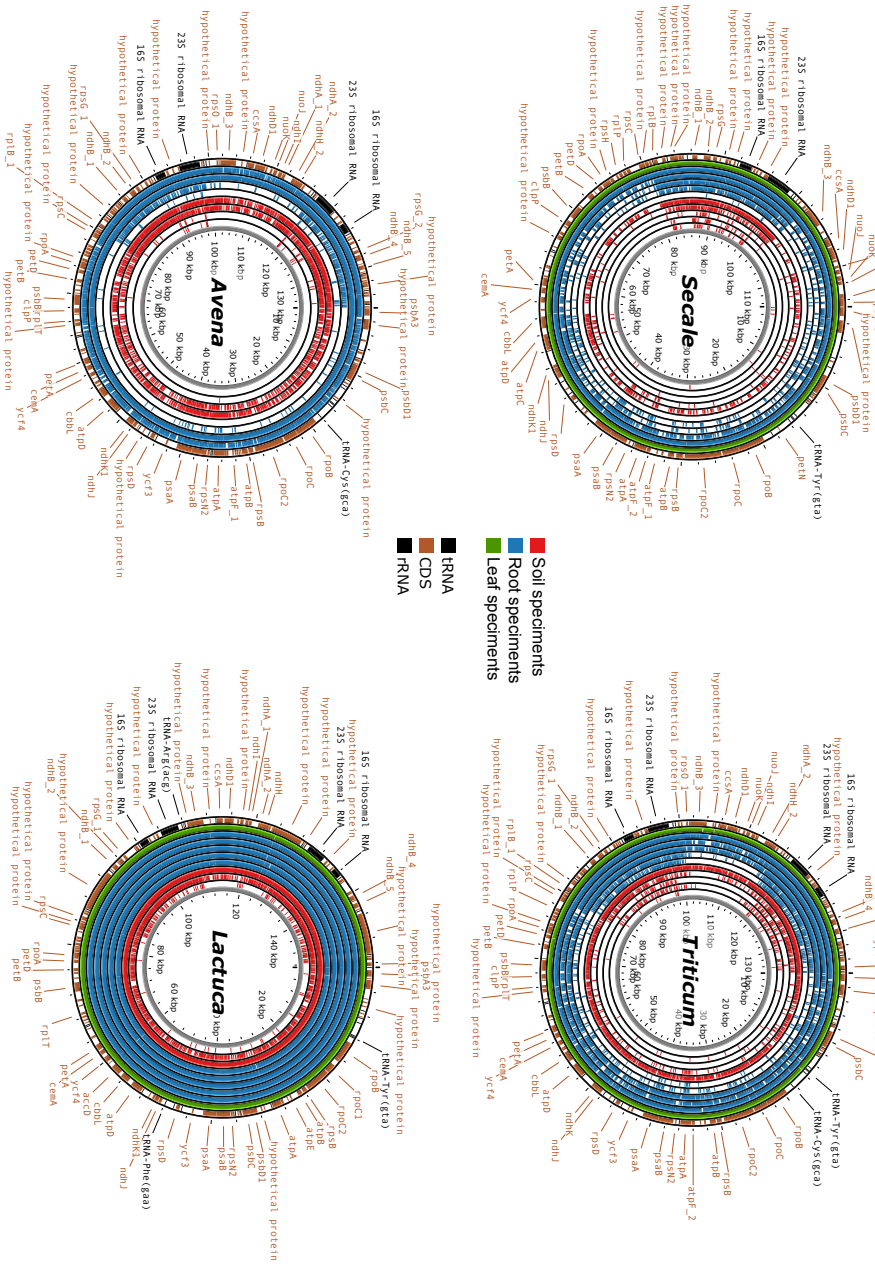


Figure 3. Chloroplast genome reconstruction from herbarium leaf, root and soil samples. The image shows circular plot of the chloroplast genomes of *Avena, Secale, Triticum,* and *Lactuca* assembled from each of the different metagenome datasets described in Chapter 3. Maps demonstrate that contigs assembled from leaf and root samples (green and blue, respectively) are often more complete compared to those from soil samples, which are characterized by shorter and more scattered fragments (red). The reference chloroplast genomes used for *Secale, Triticum, Avena*, and I are NC_021761.1, KM352501.1, NC_027468.1, and NC_0075781.1, respectively, available in GenBank. Assembly was performed using the same methods describe in chapter 3 and visualized using proksee tools (Grant et al., 2023).

“HerbaSoil”, a project to investigate the impacts of farming intensification on soils and soil microbiomes

The “HerbaSoil” project, funded by Sorbonne Université (2023-2025), aims at highlighting recent temporal changes in agricultural soils and soil microbial communities. The project implements an innovative diachronic experimental approach based on the analysis of soils associated with the roots of herbarium crop plants. These historical soils will be characterized for both their physico-chemical properties and their microbial communities after the extraction, sequencing, and annotation of their ancient DNA (aDNA) content. The diversity of historical soil microbial communities will then be compared to the diversity of current communities present in soils associated with the same crop plant species. The project aims at ***investigating changes and shifts that have characterized the soil microbiome over the last 200 years, a period marked by the acceleration of global changes and the intensification of agricultural practices***. Intensive farming, promoted after around the end of the Second World War (WWII), is characterized by the massive use of synthetic chemicals (fertilizers and pesticides), of new high-yielding crop varieties and by mechanization and homogenization of agricultural landscapes. In the absence of historical records of soil microorganism diversity, a direct answer to this question would require access to collections of "ancient" (pre-WWII) agricultural soils that have preserved signatures, in the form of ancient DNA (aDNA) fragments of their original microbial communities. The following hypotheses have been formulated:

(i) “Ancient” agricultural soils differ from “modern” ones with respect to macronutrient composition and N isotopic signature and metallic contaminants because of prolonged application of chemical fertilizers.

(ii) “Ancient” soils differ from “modern” ones with respect to the taxonomic and phylogenetic diversity of microbial species, with *e.g.* more copiotrophic taxa resulting from higher productivity.

(iii) “Ancient” soils differ from “modern” ones with respect to the functional diversity of microbial species, especially for functions related to nutrient cycling and resistance to antimicrobial compounds.

“HerbaSoil” focuses on the rhizosphere soils of two major cereals, wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*). Both species are well represented in herbaria and are still widely cultivated in France. For each species, soils (ca. 100 mg) were sampled from *ca*. 60 herbarium specimens (**Fig 2**) collected between *ca*. 1820 and 1980, thus covering as wide temporal period but also geographical area (**Fig.2**).

With regard to modern agricultural soils, about 50 farmers were contacted to sample wheat and barley plants in their field during the 2024 growing season. For comparability between ancient and modern samples, most farmers were close to the place where ancient herbarium samples had historically been collected. Modern plants will be air-dried and adhering rhizospheric soil collected a few months later to mimic herbarium plant preparation.

Soils sampled from herbarium and modern plants will be subjected chemico-physical measurements such as pH, total C, N and ^{13}C, ^{15}N contents by EA-MS; macro and microelements (Al, Ca, Cd, Cu, K, Mg, Mn, Fe, Si, Zn, ...) contents by ICP-MS and X-ray fluorescence. DNA from herbarium soil samples will be extracted as described in Chapter 3. Soil DNA will be sequenced to obtain at least 40 million of 150bp reads per sample. Sequences will be analyzed using the bioinformatics protocols described in Chapter 5 for the study of microbiomes taxonomic diversity. In addition, microbial metagenomic reads will also be subjected to functional annotation to relate them to protein sequences of known functions. To that aim, reads will be mapped to reference bacterial and fungal genomes and to curated protein databases. *De novo* assembly of "Metagenome Assembled Genomes" (MAGs) followed by taxonomical and functional annotations will also be attempted. In this way, through a comparison of modern and historical soils, we expect to highlight taxonomic and functional shifts that have characterized the microbial communities over the last 200/250 years.

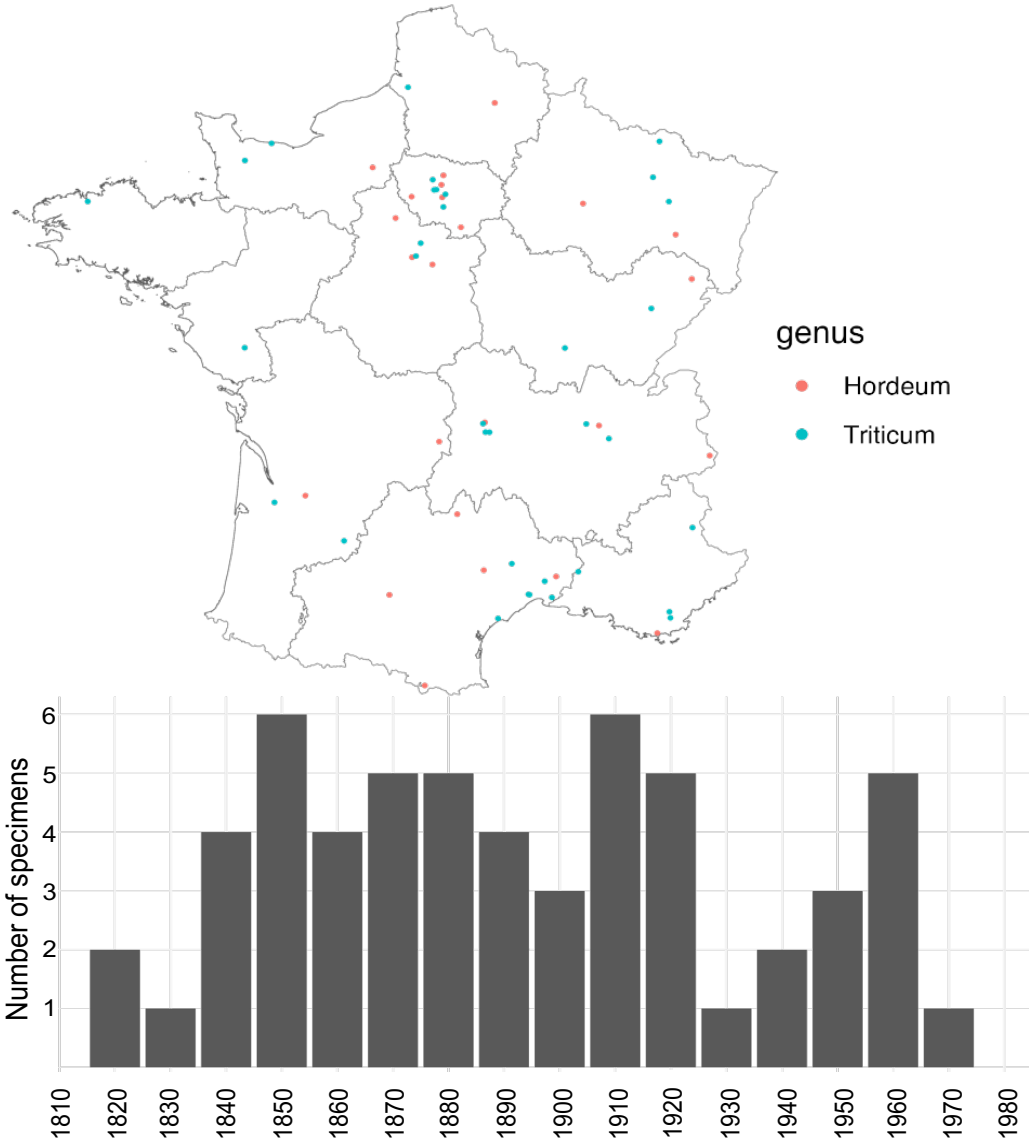


Figure 4 Origin of soil herbarium samples selected for the "HerbaSoil" project. Map of France showing the geographical distribution of the 60 herbarium soil samples. The bar plot indicates the temporal distribution of the soil sample from the 1820 to 1980.

Conclusion

The numerous herbarium plant specimens worldwide, many of which unfortunately still not digitized and buried within immense collections, represent a multifunctional tool that allows the investigation of various aspects of plant biology, ecology in the historical and cultural context of the collection period.

In the previous chapters, we have shown how the soils present in the samples add a new dimension to herbaria by allowing exploration of plant-associated past soil microbial and viral communities. Soil is a material associated to various herbarium crop plant genera (especially those with complex and extensive root systems like in the *Poaceae* and *Polygonaceae*) and DNA extraction has been successfully performed on several samples of *Hordeum* and *Triticum*, showing that factors such as collection period and drying processes do not dramatically affect DNA preservation. Preservation appears however to be affected by several physico-chemical characteristics of the original soil (Chapter 3).

The analysis of microbial DNA of herbarium soil, as shown in Chapter 5, allowed us to suggest that herbarium soil preserve the original characteristics of their microbiomes that are comparable to modern ones. Additionally, through the analysis of DNA extracted from herbarium specimens, it was possible to assemble viral genomes, as illustrated in Chapter 3.

Herbarium soil samples can thus be considered true repositories of soil and plant microbial diversity over the last *ca* three centuries. In the future, their study may highlight the changes that the Anthropocene has caused to soil-associated microorganisms. Furthermore, optimizing protocols and methods will refine the analysis of these samples, allowing for better characterization of the taxonomic composition of past microbial communities and a deeper understanding of their original functionalities.

References

Campos, P. E., Pruvost, O., Boyer, K., Chiroleu, F., Cao, T. T., Gaudeul, M., Baider, C., Utteridge, T. M. A., Becker, N., Rieux, A., & Gagnevin, L. (2023). Herbarium specimen sequencing allows precise dating of Xanthomonas citri pv. citri diversification history. *Nature Communications 2023 14:1*, *14*(1), 1–13. https://doi.org/10.1038/s41467-023-39950-z

Goloubinoff, P., Pääbo, S., & Wilson, A. C. (1993). Evolution of maize inferred from sequence diversity of an Adh2 gene segment from archaeological specimens. *Proceedings of the National Academy of Sciences*, *90*(5), 1997–2001. https://doi.org/10.1073/PNAS.90.5.1997

Gutaker, R. M., Reiter, E., Furtwängler, A., Schuenemann, V. J., & Burbano, H. A. (2017). Extraction of ultrashort DNA molecules from herbarium specimens. *BioTechniques*, *62*(2), 76–79. https://doi.org/10.2144/000114517/SUPPL_FILE/GUTAKER_SUPPLEMENTA_254083A.ZIP

Hartung, J. S., Roy, A., Fu, S., Shao, J., Schneider, W. L., & Brlansky, R. H. (2015). History and diversity of Citrus Leprosis virus recorded in herbarium specimens. *Phytopathology*, *105*(9), 1277–1284. https://doi.org/10.1094/PHYTO-03-15-0064-R

Klapper, M., Hübner, A., Ibrahim, A., Wasmuth, I., Borry, M., Haensch, V. G., Zhang, S., Al-Jammal, W. K., Suma, H., Yates, J. A. F., Frangenberg, J., Velsko, I. M., Chowdhury, S., Herbst, R., Bratovanov, E. V, Dahse, H.-M., Horch, T., Hertweck, C., Morales, M. R. G., … Stallforth, P. (2023). Natural products from reconstructed bacterial genomes of the Middle and Upper Paleolithic. *Science*, *380*(6645), 619–624. https://doi.org/10.1126/science.adf5300

Mazo, L. C., Gómez, A., Quintanilla, S. R., Bernal, J. E., & Valdivieso, P. O. (2012). Extraction and amplification of DNA from orchid exsiccates conserved for more than half a century in a herbarium in Bogotá, Colombia. *Lankesteriana: International Journal on Orchidology*. https://doi.org/10.15517/LANK.V0I0.11762

Wibowo, M. C., Yang, Z., Borry, M., Hübner, A., Huang, K. D., Tierney, B. T., Zimmerman, S., Barajas-Olmos, F., Contreras-Cubas, C., García-Ortiz, H., Martínez-Hernández, A., Luber, J. M., Kirstahler, P., Blohm, T., Smiley, F. E., Arnold, R., Ballal, S. A., Pamp, S. J., Russ, J., … Kostic, A. D. (2021). Reconstruction of ancient microbial genomes from the human gut. *Nature*, *594*(7862), 234–239. https://doi.org/10.1038/s41586-021-03532-0

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., Thines, M., Weigel, D., & Burbano, H. A. (2013). The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. *ELife*, *2*(2), 731. https://doi.org/10.7554/ELIFE.00731

Acknowledgments

As I come to the end of these years, I would like to conclude by thanking all the people who have helped me during these three years of work.

First of all, I would like to thank my supervisors, Dr. Valeria Bianciotto and Dr. Roland Marmeisse, who have accompanied me over the past three years with attention and interest, providing me with all the necessary tools and advice that have allowed me to face this doctorate in the best possible way. My thanks also go to Professor Silvia Perotto for the guidance she always gave me during my thesis.

I am deeply grateful to Prof. Lucie Bittner and the entire ABI team for warmly welcoming me into their laboratories and for the important knowledge and scientific exchanges I participated in; I also thank the entire INEVEF team for their hospitality and support.

Thank you to Dr. Regis Debruyne for introducing me to the world of paleogenetics and for generously sharing his extensive expertise.

If I can call myself "a little bioinformatic" I thank Martino Adamo, Oliver Ruè (and all the Migale platform) and Franck Lejzerowicz for their teachings.

A sincere thank you to the numerous herbarium collections I visited and the various curators I met. Thank you for the time you dedicated to me and for the chance to sample from your collections. I thank Mélanie Thiebaut, who allowed us to take the first samples from the Lyon herbarium, which initiated this thesis work.

I would then like to thank all all my friends and colleagues: Miriana, Beatrice and Sonia, Teresa, Cristina, Martina, Andrea, Davide, Serena, Simon, Lorenzo, Francesco, Alessandro, William and everyone with whom I shared moments at the "Orto" in Turin. Thanks to you, the time spent there was lighter and more pleasant, and sharing our efforts, joys, and failures was truly enriching.

I also thank Gaspard, Karen, Thomas, and Jeremy because arriving in Paris and finding such wonderful people allowed me to enjoy this period fully and share beautiful moments with them both in the office and outside.

Vorrei ringraziare di cuore gli amici di sempre, che, da vicino o da lontano, mi sono stati accanto in ogni momento. Un ringraziamento speciale va a Ire, amica di sempre con la quale il confronto e il supporto reciproco sono stati

fondamentali in questo percorso. Un grazie anche a Clock e Cate, che non solo mi regalano la loro splendida amicizia, ma hanno anche condiviso la casa con me in questi anni (con Bruno, per nostra fortuna o sfortuna!). Grazie anche a Ema, per il legame che ci unisce. Non posso dimenticare il mondo scout e la mia Co.Ca., che in questi anni mi hanno permesso di crescere e mettermi alla prova, offrendomi spesso momenti di felicità e svago al di fuori della mia quotidianità. Un enorme grazie va infine a Matteo, su cui so di poter sempre contare.

Infine, per ultimi ma non di importanza, ringrazio la mia famiglia: i miei genitori, Fiora e Pierangelo e mio fratello Massi insieme a Sara, Elia e Iago. Mi hanno sopportato da molto più di questi tre anni di dottorato e mi hanno in gran parte permesso di essere chi sono adesso con i loro insegnamenti (spesso silenziosi e fatti di gesti) ma immancabili e pieni di cura. Un grazie va anche a Chiara, che anche se non è più qui, so che sarebbe fiera di me.