



OPEN

Metagenomics reveals contrasted responses of microbial communities to wheat straw amendment in cropland and grassland soils

Domitille Jarrige¹, Vincent Tardy¹, Valentin Loux^{2,3}, Olivier Rué^{2,3}, Abad Chabbi^{4,5}, Sébastien Terrat^{1✉} & Pierre-Alain Maron¹

Soil microbial communities respond quickly to natural and/or anthropic-induced changes in environmental conditions. Metagenomics allows studying taxa that are often overlooked in microbiota studies, such as protists or viruses. Here, we employed metagenomics to characterise microbial successions after wheat straw input in a 4-month in-situ field study. We compared microbial successions patterns with those obtained by high throughput amplicon sequencing on the same soil samples to validate metagenomics as a tool for the fine analysis of microbial population dynamics in situ. Taxonomic patterns were concordant between the two methodologies but metagenomics allowed studying all the microbial groups simultaneously. Notably, our results evidenced that each domain displayed a specific dynamic pattern after wheat straw amendment. For instance, viral sequences multiplied in the early phase of straw decomposition, in parallel to copiotrophic bacteria, suggesting a “kill-the-winner” pattern that, to our knowledge, had not been observed before in soil. Altogether, our results highlighted that both inter and intra-domain trophic interactions were impacted by wheat amendment and these patterns depended on the land use history. Our study highlights that top-down regulation by microbial predators or viruses might play a key role in soil microbiota dynamics and structure.

Soil ecosystems harbour complex and highly diverse microbial communities, comprising microorganisms from all domains of life: *Archaea*, *Bacteria*, *Eukaryota* (comprising *Fungi* and protists, i.e., a paraphyletic group including non-fungal, non-metazoan and non-land plant eukaryotes¹) and *Viruses*. This microbiome responds strongly to land-use (e.g. forest, grassland, cropland)² and agricultural practices such as tillage, fertilization, crop rotation and protection strategies^{3,4}. These anthropogenic parameters significantly impact soil microbial diversity, with consequences in terms of soil sustainability and productivity, since soil functioning results from the activity of the huge diversity of these interacting microbial communities^{5–8}. Such observations have stimulated much research to decipher the dynamic patterns of soil microbiota. Our ability to better manage microbial diversity and associated functions depends on our capacity to understand these dynamics.

The addition of crop residues to soils is a widespread agricultural practice aimed at enhancing soil organic matter (SOM) content, and improving soil structure^{9,10}. As such, much effort was dedicated to determine the extent to which it changes the diversity and abundance of soil microbial communities. Studies have focussed on different questions such as the impact of residue quality^{11,12}, quantity and localisation in the soil¹³. All these studies considerably expanded our knowledge of the ecology and dynamics of the heterotrophic microbial communities involved in C-transformations in soil. However, most of these studies targeted specific groups, mainly *Bacteria* and/or *Fungi*. Thus, they lacked exhaustiveness and did not provide comprehensive understanding of the dynamics of the whole soil microbiome. Such holistic data are however needed to gain insights into the interactions between taxonomic groups, such as competition, predation, parasitism and symbiosis, especially

¹Agroécologie, INRAE, Université de Bourgogne, Institut Agro, Université de Bourgogne Franche-Comté, Dijon F-21000, France. ²Université Paris Saclay, INRAE, Jouy-en-Josas, MalAGE, France. ³INRAE, Bioinformatics, MIGALE bioinformatics facility, Université Paris-Saclay, Jouy-en-Josas, France. ⁴INRAE, Poitou-Charentes, URP3F, Lusignan 86600, France. ⁵UMR-ECOSYS Joint research unit INRAE, Université Paris-Saclay, AgroParisTech, Paris, France. ✉email: sebastien.terrat@inrae.fr

since less studied microbial predators belonging to protists and *Viruses* might be crucial in regulating the soil microbiome and associated activity^{14,15}.

Technical limitations impeding access to the diversity of the whole soil microbiota may in part explain this knowledge gap. Indeed, the main methodology used in the literature to study soil microbial diversity remains targeted metagenomics (i.e., high throughput amplicon sequencing, also called metabarcoding)¹⁶. It provides resource and cost-effective data on the ecology of indigenous microbial communities, but it is limited to the genes that are targeted, and the taxonomic groups that include them (the most commonly surveyed groups are *Bacteria*, *Archaea* and *Fungi*¹⁶). This makes it unsuitable for obtaining the whole picture of the microorganisms representing every domain of life interacting together in soils. In contrast, the metagenomics approach (i.e., whole metagenome sequencing), which does not rely on amplifying specific markers, allows achieving a global assessment of the complete soil microbiome (i.e., not restricted to a given domain). It has already been applied successfully for studying the ecology of the soil microbial community in several studies^{17–19}. However, applications to decipher the dynamic patterns of the microbial communities in situ remain scarce, particularly in the soil environment and comparisons between targeted and global metagenomic approaches are still required to fully validate the performances and/or limitations of metagenomics for such assessments. To date, a few attempts have been made to compare targeted and global metagenomic approaches. They involved diverse experimental designs and materials (polar biocrusts, environmental DNA, sedimentary ancient DNA, pollen)^{20–24}. Some studies showed large differences in the taxonomic profiles obtained with the two methods^{20,22,24}, whereas others showed closer results^{21,23}. Given the disagreement between studies so far, making conclusions about the congruence of both approaches remains uncertain and additional studies are needed.

In this paper, we present the results of a comprehensive in-situ field study conducted to evaluate the influence of contrasted land use history (20 years cropland vs. 17 years grassland) on soil microbial successions occurring after wheat straw input. For each land use history, two treatments were applied: without (control) and with wheat straw amendment. Four sampling dates were analysed for a total of 48 soil samples (3 replicates × 2 treatments × 2 plots × 4 dates). We employed metagenomics to characterize the whole soil microbial diversity (*Archaea*, *Bacteria*, *Eukaryota* and *Viruses*), and compared the results obtained for *Bacteria*, *Archaea* and *Fungi* with those previously obtained using high throughput amplicon sequencing on the same soil samples²⁵. We expected the results obtained by amplicon sequencing and global metagenomic to display similar conclusions in terms of bacterial and fungal dynamics following wheat straw input. It was a prerequisite to validate the suitability of metagenomics for characterising the dynamics of the whole soil microbiome diversity in situ. We hypothesized that due to differences in ecological strategies between the different organisms, response dynamics following wheat straw amendment would vary between *archaea*, *bacteria*, *fungi*, protists and viruses. In addition, we expected dynamic patterns to depend on the contrasted land use histories of the soil. Altogether, these results would provide information about the drivers of each group in terms of trophic attributes and/or biotic interactions.

Results and discussion

Amplicon sequencing and metagenomics depict similar pictures of the structure of bacterial and fungal communities

The structures of bacterial and fungal communities obtained with amplicon sequencing²⁵ and metagenomics displayed similar patterns in both control and amended treatments regardless of land-use history (Fig. 1). In addition, these two approaches provided similar results in terms of reproducibility, with the three field replicates showing close similarity (Fig. 1; Supplementary Fig. 1). Control grassland and cropland communities were discriminated on the factorial map, evidencing the impact of land-use history on the diversity of soil microbial communities^{25,26}. In addition, while bacterial and fungal communities did not show any significant differences in their dynamics between the four sampling dates in the control, highlighting their stability along the seasons, the input of wheat residues strongly impacted these two communities, regardless of land-use history (Permutational Multivariate Analysis of Variance [PERMANOVA], p -value < 0.001) (Supplementary Fig. 1). Thus, a strong population shift was observed after three days for both fungi and bacteria (Fig. 1). In the latter phase (days 51 and 125), bacterial communities but not fungal communities were resilient (Fig. 1).

Compared to *Bacteria* and *Fungi*, patterns of archaeal communities differed between the two sequencing approaches (Fig. 1). This discrepancy could be due to the lack of archaeal sequence recovered in the amplicon sequencing dataset. For this specific group, metagenomics proved to be far more resolutive than the metabarcoding dataset (probably due to its limited 4000 bacterial and archaeal sequences) and was able to detect a significant difference between cropland and grassland archaeal communities with PERMANOVA and Redundancy Analysis (RDA) (Supplementary Fig. 1 & Supplementary Table 1).

When looking at the relative abundance of the most abundant *phyla*, there was no major difference for bacterial and fungal *phyla* between amplicon sequencing and metagenomics (Supplementary Fig. 2), except for *Cyanobacteria* and *Basidiomycota* that were underrepresented in the amplicon sequencing datasets. In contrast, Becker and Pushkareva²⁰ found that *Cyanobacteria* were overrepresented with their amplicon sequencing primers. The poor detection of *Cyanobacteria* in the amplicon sequencing dataset was also in apparent contradiction with previous in silico analyses²⁷ of the PCR primers used in²⁵, which predicted high hit frequency for *Cyanobacteria*. This could be explained by the change in taxonomic classification since the in silico analysis or by PCR biases in the in vivo application. This illustrates the potential drawback of amplicon sequencing whose representativeness for specific groups is highly dependent on the primers used and the PCR conditions. Moreover, thanks to its considerable depth of coverage, metagenomics identified many low abundance bacterial *phyla* that could not be detected with amplicons, among which were numerous candidate *phyla* (Supplementary Fig. 3).

At a finer taxonomic level, we compared the variations of raw count abundances of *genera* previously highlighted in²⁵ during the experiment, with those obtained with the global metagenomic approach (Supplementary Fig. 4).

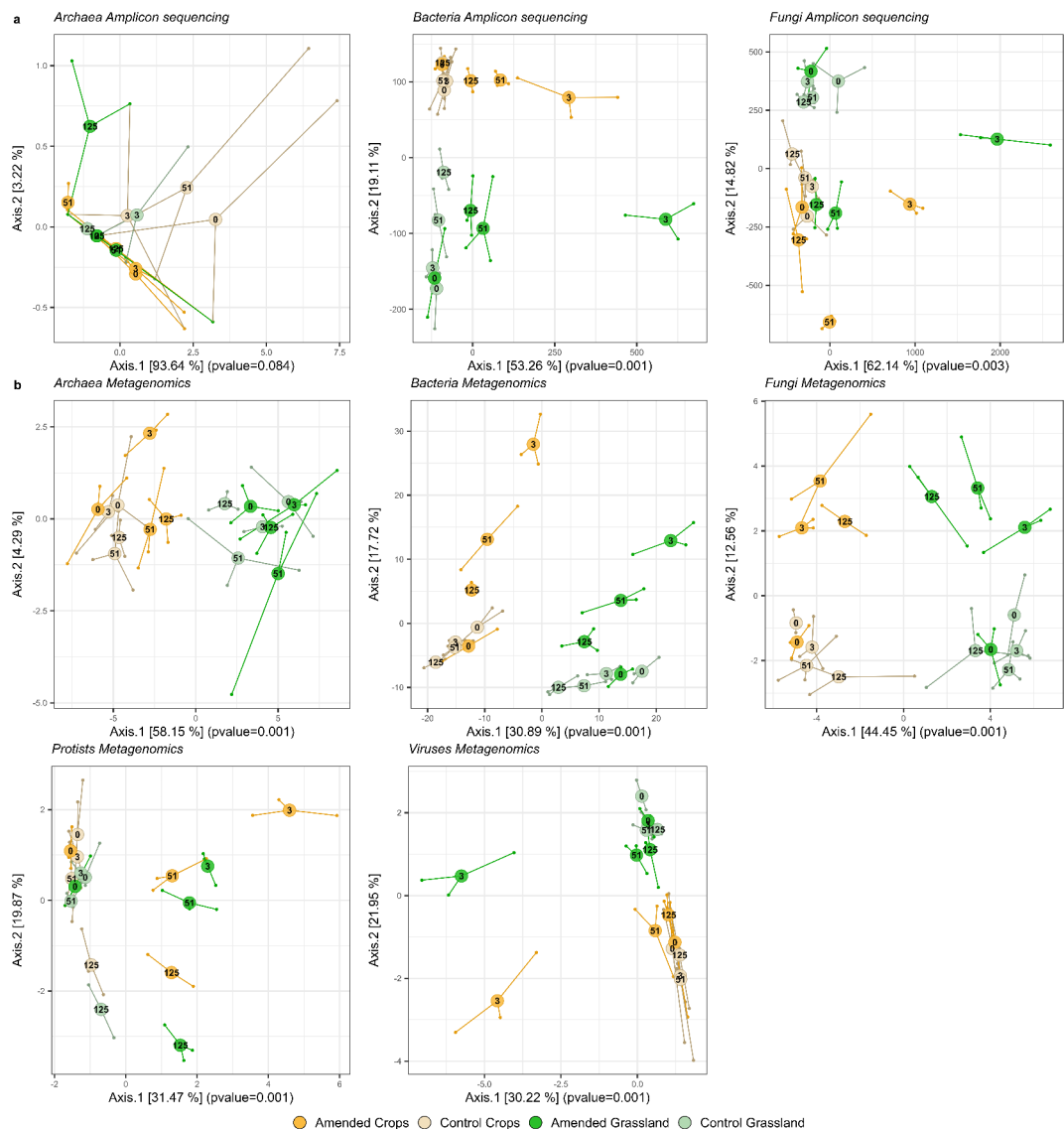


Fig. 1. PCoA of communities separated by taxonomic group. Both metabarcoding (a) and metagenomics (b) separate communities according to the land-use history, and bacterial and fungal communities display similar patterns with both approaches. (a) PCoA of metabarcoding CLR transformed data. (b) PCoA of metagenomics CLR transformed taxonomic data. Green samples denote grassland, yellow crops, pale colours controls, deep colours amended samples. The percentage of explained variability for each axis is indicated in square brackets. The p-values of PERMANOVAs computed between cropland and grassland samples are indicated in parentheses on Axis.1.

For bacterial *genera*, the abundance patterns were consistent between the two methods (Supplementary Fig. 4a. and c.), with *genera* such as *Pseudomonas* and *Massilia* highly stimulated by straw amendment, regardless of land-use history. However, metagenomics appeared more sensitive than amplicon sequencing for some *genera* such as *Bulkholderia* and *Lysobacter*. Indeed, it showed a significant influence of straw amendment in both land-use histories for these *taxa*, whereas amplicon sequencing only detected a significant influence in one soil management history (Supplementary Fig. 4a. and c.). For fungal *genera*, both approaches also exhibited close abundance profiles (Supplementary Fig. 4b. and d.). Again, metagenomics appeared more sensitive than amplicon sequencing for certain *genera* such as *Rhizopus* and *Mucor*, for which hardly any sequences were detected in the control samples with amplicon sequencing. This could be explained easily by the much higher sequencing depth in metagenomics than in amplicon sequencing, as previously suggested²⁰.

Overall, these results provide evidence that amplicon sequencing and metagenomics identified similar taxonomic patterns at the community, *phyla* and *genus* levels, highlighting that metagenomics is suitable for assessing the dynamics of the soil microbial communities in situ. Moreover, metagenomics was more sensitive than the amplicon dataset, showing the huge potential of this approach for the fine analysis of population dynamics. The similarity of the results observed between the two approaches for bacteria and fungi makes us

confident about the robustness of the patterns observed for the other microbial groups (i.e., *Viruses* and protists) through metagenomics.

Metagenomics revealed the temporal response to wheat input of the whole soil microbiota

Metagenomics gave access to the whole soil microbiota, including *Viruses*, which is difficult to achieve with amplicon sequencing²⁸. As with *Archaea*, *Bacteria* and *Fungi*, viruses and protists presented distinct structures between grasslands and croplands (Fig. 1b., PERMANOVA, p -value < 0.001), showing a strong influence of land-use history. In addition to determining the structure of all microbial groups, land-use history also impacted the complexity of the whole soil microbial community, as evidenced by the results of the multigroup cooccurrence network analysis that revealed an increase of the total number of links and connectivity in cropland control compared to grassland control (Fig. 2). This result contradicts previous reports from national²⁹ and regional scale studies³⁰ that showed higher complexity of microbial networks in semi-natural systems such as forest and grassland compared to cropland. However, it must be kept in mind that our study deals with the effect of land-use history rather than with the land-use itself, since plant cover was removed and the soil was left bare by manual weeding for 5 months to stabilise before applying wheat residue inputs (see material and methods section). It is likely that such elimination of plant cover may have deeply affected the complexity of the soil microbial network, in agreement with³¹ who observed a high overall network complexity in the rhizosphere of oat that increased as plants grew. In other respects, this highlights the importance of plant cover for stimulating the complexity of soil microbial interactions and suggests that this effect is transient and disappears quickly following plant removal. In absence of plant cover, the lasting differences in soil properties (i.e. lower pH and phosphorous content; higher values of SOM, soil organic carbon content, total nitrogen, Soil C/N and cation exchange capacity in grassland soil) probably accounted for a large part of the microbial community's differentiation between grassland and cropland soils (25).

As observed in the previous section for *Bacteria* and *Fungi*, wheat residue input strongly impacted the structure of *Viruses* and protists (Fig. 1 & Supplementary Table 1). Interestingly, it also significantly increased the complexity of the whole soil microbial community, as evidenced by the higher number of links and higher connectance of the multigroup microbial co-occurrence networks observed in both land-use histories (Fig. 2a, b.). This increase in the complexity of the soil microbial community by crop residue input had been reported previously for *Bacteria* and *Fungi*³², but to our knowledge, this is the first time it has been evidenced at the whole soil microbial community level. It points to the overall high network complexity that increases as C-inputs feed the soil microbiota, reflecting extensive interactions such as mutualism, competition, predation and parasitism that occur among microbial groups during the decomposition process.

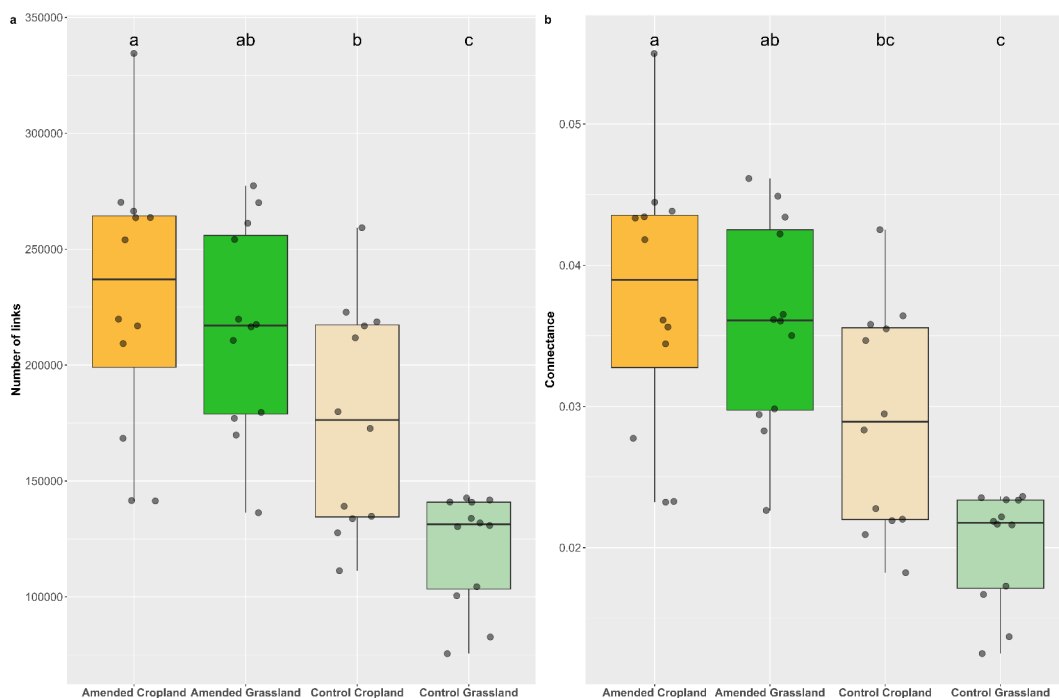


Fig. 2. Cooccurrence networks statistics according to treatment and land-use history. Complexity of the whole microbial communities was impacted by land-use history and amendment, with more complexity in the cropland soil than in the grassland soil and a significative increase in complexity for both land-uses after wheat straw amendment. **(a)** Number of links. **(b)** Connectance (proportion of links formed out of all possible links). Kruskal Wallis tests with Bonferroni corrections (α 0.05) were performed on the metrics to assess their difference between different treatments and land-uses.

The rapid and highly significant changes in the structure of both bacterial and fungal communities on day 3 after straw input (Fig. 1) reflected that not only bacteria but also fungi can act as pioneer decomposers of wheat straw^{25,33}. The resilience of bacterial populations in the late phase of straw decomposition, where easily decomposable C-substrate was depleted, could be attributed to the decrease of copiotrophic bacteria populations (Fig. 1). In contrast, the different trajectories of the fungal communities implied that other fungal populations rose at days 51 and 135. Indeed, in this latter stage of straw decomposition, mainly recalcitrant organic carbon remained, thus fungal oligotrophs were likely preponderant, as was suggested before^{25,33} due to improved abilities to decompose more complex C-substrates compared to bacteria.

In amended plots, viral communities experienced transient but strong modifications on day 3 (Supplementary Fig. 1). Since *Viruses* are obligate parasites, such modifications can result only from the increase of their hosts abundance. Indeed, Virus would multiply through lysogenic or lytic infections. In oceans, viral lysis is estimated to turnover ~ 20% of microbial biomass every day and is a major actor of biogeochemical cycles by liberating nutrients and carbon³⁴. In soil, *Viruses* have been studied less so far and very little is known about their roles in regulation of other microbial groups and carbon decomposition³⁵. In our study, their short-lived response following wheat straw input might indicate a strong link with copiotrophic populations. They might be in part responsible for the rapid resilience of bacterial population structures. As such, these results suggest that *Viruses* could act as moderating agents by targeting the most active populations of soil microorganisms in a “kill-the-winner” strategy³⁶.

As for the other microbial groups; the wheat residue input still had a clear impact on the structure of the protist community on day 3, suggesting that copiotrophic protists (or perhaps predators of copiotrophic microorganisms) might multiply quickly following C-input into the soil. Moreover, we observed a shift of both control and amended communities from the initial state at day 125 (Fig. 1), implying that factors other than amendment might impact this group. It could be, for example, an influence of seasonality, as day 125 samples were collected in January. Indeed, RDA analyses including land use, amendment and meteorological data (i.e. we chose soil temperature at 10 cm depth, as soil humidity was strongly anticorrelated to temperature) demonstrated that the protist community was the only microbial group majorly influenced by climatic conditions (Supplementary Fig. 5 & Supplementary Table 1). These results are in agreement with a global distribution study that evidenced that meteorological data were good predictors of the composition of the protist soil community³⁷.

Microbial heterotrophic successions revealed common and specific patterns according to land-use history

To elucidate the abundance patterns of *genera* within the reactive part of the community, i.e., *genera* whose abundance changed significantly following wheat addition, differential analyses were performed on microbial *genera* with DESeq2, between control and amended conditions. We identified 351 Differentially Abundant *Genera* (or DAGs) from all domains of life (Supplementary Table 2). Using these DAGs, hierarchical biclustering was performed to identify the succession of populations during the whole kinetics linked with land management history (Fig. 3). In parallel, Principal Component Analyses (PCAs) and PERMANOVAs were performed on each of the *genera* clusters defined with the hierarchical biclustering to assess their role in community response³⁸. This analysis evidenced different features of the response of the whole soil microbial community to wheat straw residue input.

Firstly, the clustering of the samples above all showed a separation between the amended samples 3 days after the straw input from all other samples, regardless of land management history (Fig. 3). This highlighted the strong impact of the wheat straw input on soil microbial communities in the early decomposition phase. Notably, many over-represented microbial *genera* were common to both land use histories during this early phase of degradation. These generalist *genera*, like *Massilia*, a bacterial *genera* which was also found to proliferate in other straw amendment studies^{25,39}, and the fungus *Mortierella*, were mainly gathered in DAGs cluster 9. Interestingly, cluster 9 also encompassed protists such as *Pythium* and several viral *genera* of the class *Caudoviricetes* (which groups bacterial and archaeal phages) like *Amigovirus*.

Secondly, amended soils on day 3 were discriminated between croplands and grasslands, suggesting differences between the early responses of the microbial communities to the straw addition according to land-use history. Put simply, these differences stemmed from early responder *genera* reacting specifically in one land-use history with *genera* belonging to cluster 11, e.g., *Leifsonia* and *Paraburkholderia*, specific to grassland; or to cluster 7, e.g., *Serratia* and *Enterobacter*, specific to cropland).

Thirdly, all the other samples were separated according to their land-use history, with unamended soil microbiota remaining close together throughout the whole kinetic. This separation was mainly driven by cluster 1 *genera* which comprised all the archaeal DAGs which were more represented in croplands, and cluster 8 *genera* that were more abundant in grasslands. Interestingly, in these clusters, amended day 0 samples were clustered with the control samples. It was obviously due to a lack of time for the communities to respond to wheat addition, but it confirms the robustness of our field experiment and sampling approach.

Finally, days 51 and 125 amended soil samples were clustered together within their respective land-use histories. This can be linked with the response of *genera* stimulated during the late phase of wheat straw decomposition (days 51 and 125) belonging to DAGs clusters 4, 5, and 10. Again, some responding *genera* were common to both land-use histories (e.g., *Chitinophaga*, *Fusarium*, *Syncephalis*) while others were specific to grasslands (e.g., *Arenimonas*, *Cytrophaga*) or croplands (e.g., *Hypoxylon*, *Lysobacter*, *Rhizopus*, *Acanthamoeba*).

Archaeal *genera*, did not react strongly to the amendment, suggesting that they were not major players in carbon decomposition. *Bacteria* represented the bulk of *genera* (240 out of 351) influenced by wheat straw input. The variety of their responses in the late or early phase, in one land-use or in both, illustrated the considerable breadth of life strategies amongst *Bacteria*. Fungal *genera* likewise showed a broad variety of responses after straw input, showing their ability to act as labile or recalcitrant organic matter degraders. Regarding protists,

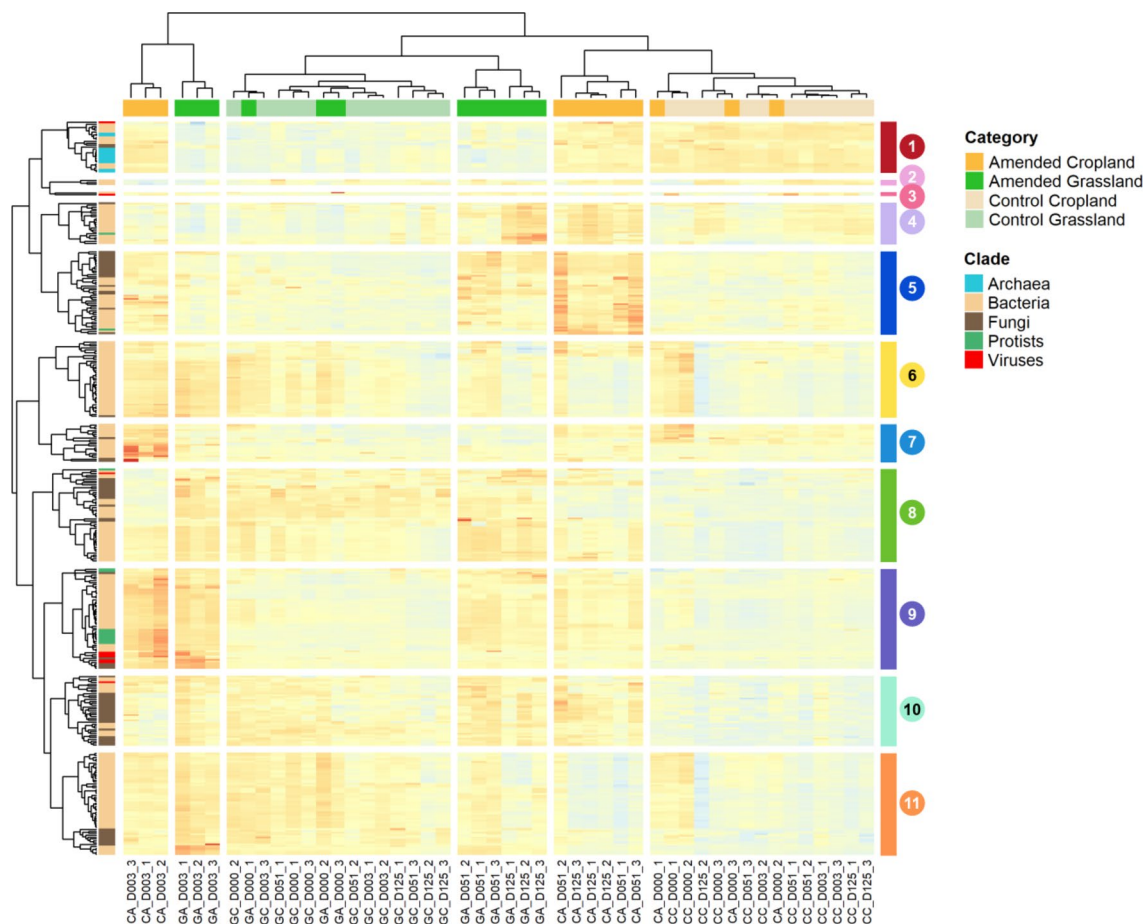


Fig. 3. DAG and sample hierarchical biclustering. Clusters of *genera*, some specific to one land-use history, some generalist, that multiply in either the early phase or late phase after amendment are observed. Sample hierarchical clustering displayed at top, DAG hierarchical clustering on the left. DAGs clusters numbers are indicated in colours next to the *genera*, as well as their taxonomic assignment on the left.

Oomycetes reacted in the early phase of decomposition in both land-use histories, showing their similarities with fungal lifestyles, while other protists did not react quickly to carbon addition. Lastly, most viral *genera* reacting to the amendment belonged to the *Caudoviricetes* phages. They responded positively and quickly after the straw input in both land-use histories. This suggests either that their hosts multiplied in the early stage of straw decomposition in both land-uses, or that they might infect multiple hosts.

While soil bacterial and fungal heterotrophic successions have already been described in previous studies^{25,39}, our metagenomics survey provides a powerful tool for examining the concerted succession of all soil microbial actors. It allows looking further than the regulation of bacterial and fungal populations by resource availability, and investigating potential trophic regulations between all taxonomic groups¹⁶.

Biotic interactions suggested

Since we looked at every microbial entity at the same time, we could hypothesise about potential interactions between taxonomic groups and how they might participate in regulating the growth of wheat straw consumers. For instance, the bacterial facultative predator *genus* *Lysobacter* was most abundant at day 51. *Lysobacter* species can lyse *Bacteria*, *Fungi* and *Oomycetes* cells⁴⁰. Interestingly, all the reactive *Oomycetes* *genera* were found in cluster 9 (Supplementary Table 2) and were most abundant on day 3 and then decreased, perhaps in part because of predators such as *Lysobacter*. Overall, several bacterial *genera* previously characterised as facultative predator (e.g., *Cupriavidus*⁴¹, members of the *Cytophagales*⁴², *Ensifer*⁴³, *Stenotrophomonas*⁴⁴) were influenced positively by the straw input. However, because of their omnivorous diet, the question of whether their multiplication was fuelled mostly by wheat straw decomposition, predation or by a combination of both processes cannot be answered. In contrast, the situation was more straightforward for *Acanthamoeba*, a ubiquitous amoebozoan *genus* that feeds on bacterial and eukaryotic prey⁴⁵. It was significantly more abundant in cropland amended soils at days 51 and 125, suggesting greater nutrient uptake by predation and gradual multiplication (Fig. 4). Predation by this *genus* has also been observed in a leaf litter decomposition experiment¹⁵. Protozoan grazers notably release the ammonium contained by their prey⁴⁶. They have been suggested to play an important role in nutrient release from *Bacteria* to soils, where the nutrients could be exploited by other microorganisms, feeding the “microbial loop”⁴⁷.

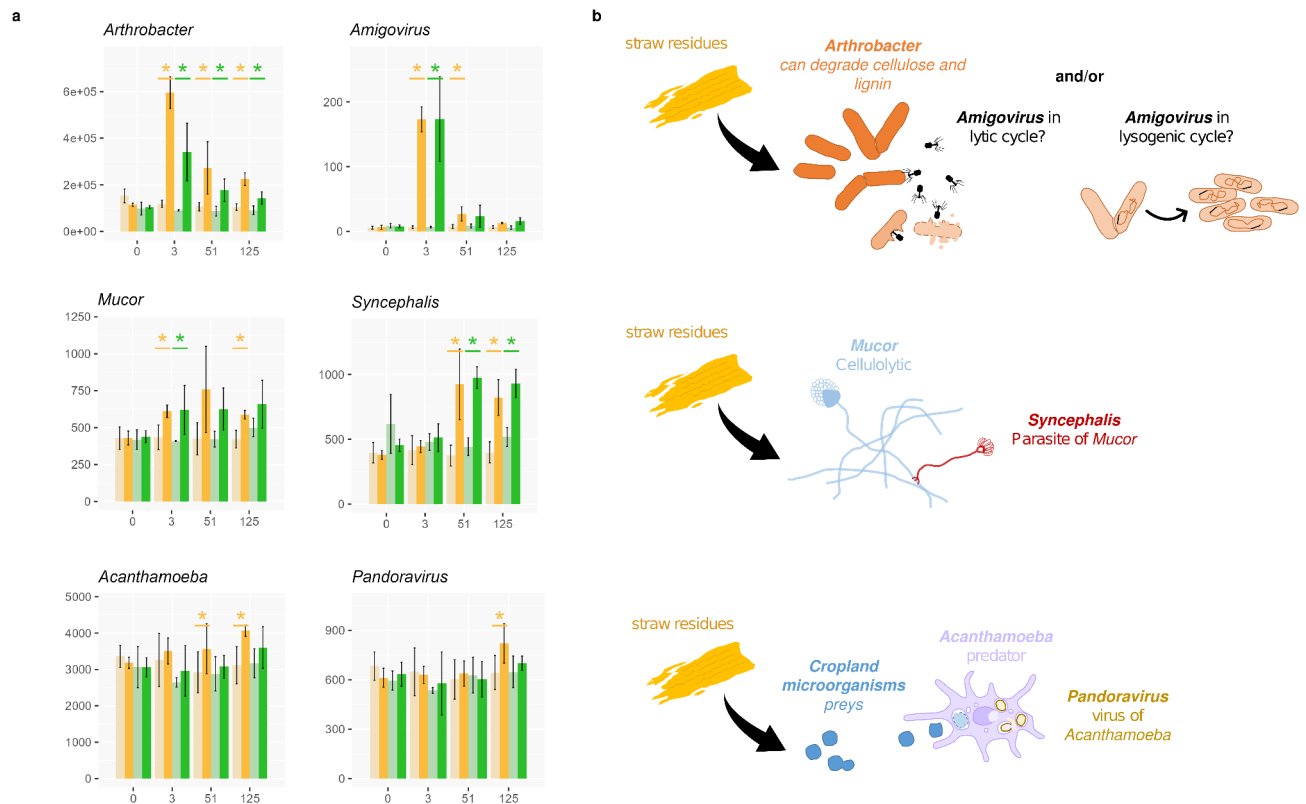


Fig. 4. Abundance patterns of selected *genera* and theorized trophic links. Abundance profiles (a) of some *genera* with trophic interactions described in the literature suggest potential top-down regulation models (b) in our dataset. (a) Raw abundances of several selected *genera* along the time series. Stars denote differentially abundant counts between the control and the amended treatment at a given time point (b) Models of potential trophic interactions between the selected *genera*.

In DAGs cluster 9, sequences affiliated to several bacteriophages displayed significant enrichment: *Amigovirus*, *R4virus* and *Slashvirus*. Notably, *Amigovirus* is a known predator of *Arthrobacter*⁴⁸ and sequences affiliated to this virus increased in parallel with *Arthrobacter* sequences. As *Viruses* have no means of directly exploiting wheat residues, this could suggest either a passive multiplication of viral sequences in the bacterial host, or an active lytic cycle infection (Fig. 4). Interestingly, another type of virus was found in the DAGs, namely *Pandoravirus*, a giant virus *genus* known to infect *Acanthamoeba* organisms⁴⁹. These two organisms were most abundant specifically in croplands on day 125. We could hypothesise that populations of the predator *Acanthamoeba* could in turn be regulated by viral infection (Fig. 4). The haustorial obligate mycoparasite *genus* *Syncephalis* that targets *Mucorales*⁵⁰ was found in DAGs cluster 5. This parasite might be implicated in the reduction of sequences affiliated to its potential hosts *Mucor* or *Mortierella* at days 51 and/or 125 in both land-use histories (Fig. 4). This parasitic *genus* was also previously shown to incorporate carbon from plant residue, alongside *Mortierella*⁵¹.

In conclusion, in this study we have shown that amplicon sequencing and metagenomics conducted years apart, with different technologies and with large sequencing depth differences, still provide similar assessments of in situ bacterial and fungal communities. Owing to its greater sequencing depth and its without a priori approach, metagenomics was more resolutive and allowed simultaneously assessing the whole soil microbial community (*Archaea*, *Bacteria*, *Eukaryota* and *Viruses*), hence providing precious data to decipher the interactions between groups in response to fresh plant residue inputs. Organisms from every taxonomic compartment were affected either positively or negatively by the amendment. Land-use histories impacted the soil communities and their responses to amendment, with different microorganisms responding in the two types of soils. Lastly, both inter and intra-domain trophic interactions implicating known consumer *genera* could be suggested from our dataset. Microbial predators, parasites and *Viruses* likely regulated soil communities' response due to their generalist or specific prey/host range, thereby constituting top-down regulation and counterbalancing the bottom-up regulation of plant inputs (exudates, or plant litter)¹⁴.

Going further in establishing the functional potential of the soil microbiota as a whole will enable the exploration of carbon cycling pathways. In addition, by analysing the functional repertoire of the microbial community, we might gain insight into why some *genera* had different responses in the two land-use histories. For instance, it might allow distinguishing different strains with different capacities for carbon decomposition or different defence mechanisms against predators or competitors. Lastly, by analysing viral sequences more thoroughly, it could be possible to predict their potential hosts and assess whether phages are induced or

integrated in their hosts' genome. Thus, we could confirm whether, as in marine environments⁵², *Viruses* take part in a “kill-the-winner” strategy in soil environments.

Materials and methods

Field experiment description and soil sampling

The field experiment was conducted at Lusignan long term observatory (France), on two adjacent plots with strongly contrasted land use-histories: 20 years croplands, with crop rotation, annual tillage, herbicide crop treatment and nitrogen fertilization, or 17 years grasslands, fertilized and harvested for annual forage. The two plots' soils had similar loamy sand textures. But some chemical properties were different between each land use: grassland soil had a mean pH of 6.02, a mean SOM of 24.05 and a mean C/N ratio of 11.38, cropland soil had a higher mean pH of 6.61, a lower mean SOM of 18.42 and a lower C/N ratio of 10.85. The detailed field experiment set up and soil characteristics were reported in²⁵. Briefly, the two main plots (5 × 2 m) were prepared by manual weeding 5 month before the experiment to prevent plants' influences on microbial communities. After weeding, each plot was divided into six 0.7 × 0.7 m sub-plots and the site was left for 5 months to stabilize, with manual weeding every week to remove regrowth seedlings. In September 2011, all sub-plots had their first 10 cm excavated, mixed and returned in place. In half of them, wheat straw residues were amended into the mixed soil (250 g per sub-plots, corresponding to 5 t dry matter ha⁻¹), while the remaining plots, serving as control, were put back in place without amendment. Four sampling dates were analysed: at the date of wheat straw amendment (T0), then 3, 51 and 125 days after amendment (respectively T3, T51, T125). It corresponded to a total of 48 soil samples (3 replicates × 2 treatments × 2 plots × 4 dates). These sampling dates have been selected from the global sampling survey, as the community changes peaked after 3 days, and resilience of the community structure occurred more rapidly for bacteria (51 days) than for fungi (125 days)²⁵.

DNA extraction and sequencing

For metagenomics, the DNA extraction was performed using a single procedure standardized by the GenoSol platform, based on the standardized 'ISO-11063: Soil quality Method to directly extract DNA from soil protocol'^{27,53}. Generally, soil DNA extraction is impaired by the complex matrix characteristics which lower target organisms' accessibility, and/or by differences in target organisms' cell wall composition and structure which reduce cell lysis efficiency, as observed for bacterial strains. This is particularly the case for microeukaryotes with fortified cell walls or very resilient cell membranes⁵⁴, and for viral particles with specific capsid structure. However, the DNA extraction procedure used in this study combines a mechanical, thermal and chemical lysis step, giving a higher access to the DNA of underestimated microbial groups²⁷. Library preparation and sequencing were performed by the GeT-PlaGe platform (Auzeville, France). As the DNA concentration were low, libraries were prepared using the TruSeq Nano DNA, which included a whole metagenome amplification step. Libraries were then sequenced with Illumina NovaSeq 2 × 150 bp technology.

For amplicon sequencing, the dataset originally from Tardy and collaborators²⁵ was reused. Briefly, a 16 S rRNA gene fragment (targeting Bacteria and Archaea) and a 18 S rRNA gene fragment (targeting Fungi) were amplified by PCR using primers F479 & R888 and primers FR1 & FF390 respectively, following the PCR conditions described previously²⁵. PCR products were then purified and pyrosequenced with a GS FLX Titanium (Roche 454 Sequencing System).

Bioinformatic analyses

After sequencing, an average of 60.6 million read pairs were obtained for each library, representing a total of 2.9 billion read pairs, all 48 samples combined (2.14 TB for the entire dataset, with a mean GC content of around 63%). The quality control of the metagenomics shotgun paired-end reads was performed with Fastqc v0.11.9 and MultiQC v1.7. As the quality was good (PhredScore over 35 throughout the reads, no N content, no adapters, etc.), paired-end reads were used directly for taxonomic assignment, with no cleaning step. Taxonomic assignment was performed on the paired-end reads with Kaiju v1.7.0, default parameters⁵⁵, against the nr_euk database (v 2018). Of all reads: 0.98% were classified as Archaea, 68.01% as Bacteria, 0.72% as Eukaryota, 0.03% as Viruses, 30.26% were unclassified. Counts for a total of 57 361 different taxa were retrieved after the assignment. These metagenomic taxonomic counts were used for downstream analysis.

The amplicon dataset was initially treated with a pipeline called GnS-PIPE and developed by the GenoSol platform (INRA, Dijon, France)²⁵. More precisely, the sequences were cleaned, clustered and taxonomically assigned with dedicated parameters²⁵. The amplicon sequencing data was then manually corrected to account for few changes in taxonomic classification since the initial study (e.g., some taxa originally classified as Fungi *incertae sedis* were moved to other *phyla*) (modification files available at⁵⁶).

Statistical analyses

Statistical analyses were conducted using R v4.3.1 in the RStudio software v2023.9.1.494. Plots were generated with ggplot2 v3.4.3 unless stated otherwise. The RMarkdown script used is available at³⁸. Taxonomically assigned counts were used as a measure of taxa abundances. They were converted as phyloseq (phyloseq v1.44.0) objects for easier analysis. Rare taxa detected in less than 10% of samples (representing 21.33% of the taxonomic dataset), which might be attributed to spurious assignment and might constitute noise in the data, were excluded from further analysis. 45 127 taxa were kept for downstream analyses.

Three normalisation methods were tested to remove the biases induced by differences in library size and sampling fractions⁵⁷. First, the traditional rarefaction, where filtered taxonomic counts were randomly subsampled at the depth of 95% of the smallest library for each large taxonomic group. Common limitations of this approach evoked in the literature are a loss of valid information and an introduction of an additional sampling bias⁵⁷. Second, the Centred Log-Ratio (CLR) transformation was performed on the filtered taxonomic

counts (microbiomeMarker v1.6.0). Developed to deal with compositional data such as microbiome data⁵⁸ the CLR transformation centres the data and eliminate library size biases. Negative values in the CLR taxonomic counts matrix were set as zero, considering that these very low abundance counts were negligible. Lastly, the ANCOM-BC normalisation procedure was tested on the filtered counts data (ANCOMBC v2.2.2). Also based on log-ratio this method has a bias correction procedure. The results of these three different normalisation techniques can be observed in a RMarkdown report³⁸. On a computation time stand-point, with our 45 127 taxa dataset, both rarefaction and ANCOM-BC ran slowly (1118.75 s and 5389.42 s respectively, running time measured with package tictoc v1.2.1). In contrast the CLR transform ran far quicker (0.38 s). The normalised datasets obtained after each of the three procedures were then used to compute Principal Coordinates Analyses (PCoA) using Euclidean distance (equivalent to an Aitchison distance for our log-ratio transformed datasets⁵⁹) with phyloseq v1.44.0. Both the CLR and ANCOM-BC methods allowed for a clear separation of samples according to their land use histories for every taxonomic group, while rarefaction was less resolute, notably for Fungi and Viruses. Total explained variability tended to be higher for the CLR data than the ANCOM-BC data. Based on these results, the CLR normalisation procedure was chosen, it performed well in term of computation time for our taxa-rich dataset, allowed for cropland and grassland communities separation, was more straightforward and better established in the literature than ANCOM-BC. Permutational Multivariate Analysis of Variance (PERMANOVA, 999 permutations) was used to estimate the significance of observed differences between groups (vegan v2.6-4).

To estimate the potential influence of technical or environmental variables on soil microbial communities, climatic data collected on site and DNA libraries properties were screened to select relevant and non-colinear variables with corrplot v0.92 and PerformanceAnalytics v2.0.4. Kept quantitative variables were scaled and the best variance partitioning model of the whole CLR normalised community data was computed, using phyloseq and vegan. This model (microbiome_data ~ wheat_straw_amendment + land_use + soil_temperature) was then used to compute RDA for each microbial group taxonomic composition.

Cooccurrence networks for each land use history X treatment combination were built using the statnet suite v2019.6, including ergm v4.6.0. For each category, all the time points were grouped together, amounting to 12 samples. Repeated network generation was performed, with each one of the 12 samples left out in turn. The metrics of the 12 repetitions per category were taken into account for the analysis.

Differential analyses were performed with DESeq2 v1.40.2 on the raw, unfiltered taxonomic counts data, agglomerated at the *genus* level (with phyloseq v1.44.0), to detect potential DAGs (padj threshold: < 0.01, log2FoldChange threshold: >0.25). For each land use and time point, the amended samples were tested against the corresponding control samples, to remove potential seasonal effects (temperature and precipitations variations notably). The resulting DAGs were then filtered according to their prevalence in their respective domain (only those representing at least 1% of their domain were kept for further exploration). A CLR transformation was performed on the data agglomerated at the genus level, with again negative values set to zero and a rare *genus* (detected in less than 10% of samples after the transformation and negative removal) filter was applied. A biclustering of the samples and the filtered DAGs was plotted using the CLR transformed data agglomerated at the genus level, and Pearson distance (corresponding to 1 – Pearson correlation) with ComplexHeatmap v2.16.0. The DAGs that did not pass the CLR filters were not included (28 out of 379, leaving 351 *genera*). The resulting clusters of DAGs were then separately analysed by PCA and pairwise PERMANOVA (pairwiseAdonis v0.4.1), to detect on which grouping variable the samples were separated. Lastly, the raw abundances of each DAGs were plotted to better grasp their dynamics in the experiment.

Amplicon sequencing data (1.18 MB for the entire dataset) was reused from²⁵, with as normalisation procedure, a rarefaction by random sampling to a level of 4000 reads for 16 S rRNA gene data (0.05% of reads assigned to Archaea, 92.17% to Bacteria and 7.77% unclassified) and 8000 reads for 18 S rRNA gene (39.36% of reads assigned to Fungi, 60.64% unclassified). No other normalisation procedure was performed. PCoA of the communities separated by domain were performed, as well as differential analyses with DESeq2 as described for the metagenomic data. Raw abundance profiles for *genera* of interest were plotted. The complete workflow can be viewed here (Supplementary Fig. 6).

Data availability

The raw read datasets were deposited at DDBJ/ENA/GenBank under the BioProject accession number PRJ-NA1159878⁶⁰. The taxonomic counts as well as the metadata and the statistical analyses performed for the current study are available in a Recherche Data Gouv repository (<https://doi.org/10.57745/OMJUXC>, <https://doi.org/10.57745/HXN1XT>, <https://doi.org/10.57745/RXJXHW>).

Received: 13 November 2024; Accepted: 15 April 2025

Published online: 27 April 2025

References

- Burki, F., Sandin, M. M. & Jany, M. Diversity and ecology of protists revealed by metabarcoding. *Curr. Biol.* **31**, R1267–R1280 (2021).
- Dequiedt, S. et al. Biogeographical patterns of soil molecular microbial biomass as influenced by soil characteristics and management: biogeography of soil microbial biomass. *Glob Ecol. Biogeogr.* **20**, 641–652 (2011).
- Constancias, F. et al. Mapping and determinism of soil microbial community distribution across an agricultural landscape. *MicrobiologyOpen* **4**, 505–517 (2015).
- Christel, A., Maron, P. A. & Ranjard, L. Impact of farming systems on soil ecological quality: a meta-analysis. *Environ. Chem. Lett.* **19**, 4603–4625 (2021).
- Amundson, R. et al. Soil and human security in the 21st century. *Science* **348**, 1261071 (2015).

6. Maron, P. A. et al. High microbial diversity promotes soil ecosystem functioning. *Appl. Environ. Microbiol.* **84**, e02738–e02717 (2018).
7. Strickland, M. S., Lauber, C., Fierer, N. & Bradford, M. A. Testing the functional significance of microbial community composition. *Ecology* **90**, 441–451 (2009).
8. Cleveland, C. C. et al. Litter quality versus soil microbial community controls over decomposition: a quantitative analysis. *Oecologia* **174**, 283–294 (2014).
9. Bhogal, A., Nicholson, F. A. & Chambers, B. J. Organic carbon additions: effects on soil bio-physical and physico-chemical properties. *Eur. J. Soil. Sci.* **60**, 276–286 (2009).
10. Liu, C., Lu, M., Cui, J., Li, B. & Fang, C. Effects of straw carbon input on carbon dynamics in agricultural soils: a meta-analysis. *Glob Change Biol.* **20**, 1366–1381 (2014).
11. Pascault, N. et al. Stimulation of different functional groups of bacteria by various plant residues as a driver of soil priming effect. *Ecosystems* **16**, 810–822 (2013).
12. Wang, H. et al. Quality of fresh organic matter affects priming of soil organic matter and substrate utilization patterns of microbes. *Sci. Rep.* **5**, 10102 (2015).
13. Sanaullah, M. et al. Decomposition and stabilization of root litter in top- and subsoil horizons: what is the difference? *Plant. Soil.* **338**, 127–141 (2011).
14. Thakur, M. P. & Geisen, S. Trophic regulations of the soil Microbiome. *Trends Microbiol.* **27**, 771–780 (2019).
15. Kramer, S. et al. Resource partitioning between bacteria, fungi, and protists in the detritusphere of an agricultural soil. *Front. Microbiol.* **7**, 1524 (2016).
16. Geisen, S. The future of (soil) Microbiome studies: current limitations, integration, and perspectives. *mSystems* **6**, e00613–e00621 (2021).
17. Dai, Z. et al. Metagenomic insights into soil microbial communities involved in carbon cycling along an elevation climosequences. *Environ. Microbiol.* **23**, 4631–4645 (2021).
18. Hu, M. et al. Dynamics of soil microbial communities involved in carbon cycling along three successional forests in Southern China. *Front. Microbiol.* **14**, 1326057 (2024).
19. Hartman, W. H., Ye, R., Horwath, W. R. & Tringe, S. G. A genomic perspective on stoichiometric regulation of soil carbon cycling. *ISME J.* **11**, 2652–2665 (2017).
20. Becker, B. & Pushkareva, E. Metagenomics provides a deeper assessment of the diversity of bacterial communities in Polar soils than metabarcoding. *Genes* **14**, 812 (2023).
21. Rieder, J., Kapopoulou, A., Bank, C. & Adrian-Kalchauer, I. Metagenomics and metabarcoding experimental choices and their impact on microbial community characterization in freshwater recirculating aquaculture systems. *Environ. Microbiome.* **18**, 8 (2023).
22. Armbricht, L. et al. Paleo-diatom composition from Santa Barbara basin deep-sea sediments: a comparison of 18S-V9 and diat-rbcL metabarcoding vs shotgun metagenomics. *ISME Commun.* **1**, 66 (2021).
23. Bell, K. L. et al. Comparing whole-genome shotgun sequencing and DNA metabarcoding approaches for species identification and quantification of pollen species mixtures. *Ecol. Evol.* **11**, 16082–16098 (2021).
24. Serite, C. P. et al. eDNA metabarcoding vs metagenomics: an assessment of dietary competition in two estuarine pipefishes. *Front. Mar. Sci.* **10**, 1116741 (2023).
25. Tardy, V. et al. Land use history shifts in situ fungal and bacterial successions following wheat straw input into the soil. *PLOS ONE.* **10**, e0130672 (2015).
26. Jangid, K. et al. Land-use history has a stronger impact on soil microbial community composition than aboveground vegetation and soil properties. *Soil. Biol. Biochem.* **43**, 2184–2193 (2011).
27. Terrat, S. et al. Meta-barcoded evaluation of the ISO standard 11063 DNA extraction procedure to characterize soil bacterial and fungal community diversity and composition. *Microb. Biotechnol.* **8**, 131–142 (2015).
28. Roux, S. A viral ecogenomics framework to uncover the secrets of Nature's microbe whisperers. *mSystems* **4**, e00111–e00119 (2019).
29. Karimi, B. et al. Biogeography of soil bacterial networks along a gradient of cropping intensity. *Sci. Rep.* **9**, 3812 (2019).
30. Christel, A. et al. Differential responses of soil microbial biomass, diversity and interactions to land use intensity at a territorial scale. *Sci. Total Environ.* **906**, 167454 (2024).
31. Shi, S. et al. The interconnected rhizosphere: high network complexity dominates rhizosphere assemblages. *Ecol. Lett.* **19**, 926–936 (2016).
32. Banerjee, S. et al. Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil. *Soil. Biol. Biochem.* **97**, 188–198 (2016).
33. Ho, A., Lonardo, D. P. D. & Bodelier, P. L. E. Revisiting life strategy concepts in environmental microbial ecology. *FEMS Microbiol. Ecol.* fix006 (2017). <https://doi.org/10.1093/femsec/fix006>
34. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
35. Trubl, G. et al. Soil viruses are underexplored players in ecosystem carbon processing. *mSystems* **3**, (2018).
36. Van Goethem, M. W., Swenson, T. L., Trubl, G., Roux, S. & Northen, T. R. Characteristics of wetting-induced bacteriophage blooms in biological soil crust. *mBio* **10**, e02287–e02219 (2019).
37. Oliverio, A. M. et al. The global-scale distributions of soil protists and their contributions to belowground systems. *Sci. Adv.* **6**, eaax8787 (2020).
38. Jarrige, D. et al. Analysis report and associated Rmarkdown for metagenomics reveals contrasted responses of microbial communities to wheat straw amendment in cropland and grassland soils. (2024). <https://doi.org/10.57745/HXN1XT>
39. Su, Y. et al. Linking soil microbial community dynamics to straw-carbon distribution in soil organic carbon. *Sci. Rep.* **10**, 5526 (2020).
40. Martins, S. J. et al. Predators of soil bacteria in plant and human health. *Phytobiomes J.* **6**, 184–200 (2022).
41. Makkar, N. S. & Casida, L. E. *Cupriavidus necator* gen. Nov., Sp. nov.; a nonobligate bacterial predator of bacteria in soil. *Int. J. Syst. Bacteriol.* **37**, 323–326 (1987).
42. Hungate, B. A. et al. The functional significance of bacterial predators. *mBio* **12**, e00466–e00421 (2021).
43. Casida, L. E. *Ensifer adhaerens* gen. Nov., Sp. nov.: a bacterial predator of bacteria in soil? *Int. J. Syst. Bacteriol.* **32**, 339–345 (1982).
44. Nogales, B., Guerrero, R. & Esteve, I. A heterotrophic bacterium inhibits growth of several species of the genus *Chlorobium*. *Arch. Microbiol.* **167**, 396–399 (1997).
45. Geisen, S., Fiore-Donno, A. M., Walochnik, J. & Bonkowski, M. *Acanthamoeba* everywhere: high diversity of *Acanthamoeba* in soils. *Parasitol. Res.* **113**, 3151–3158 (2014).
46. Weekers, P. H. H., Bodelier, P. L. E., Wijen, J. P. H. & Vogels, G. D. Effects of grazing by the free-living soil amoebae *Acanthamoeba castellanii*, *Acanthamoeba polyphaga*, and *Hartmannella vermiformis* on various bacteria. *Appl. Environ. Microbiol.* **59**, 2317–2319 (1993).
47. Bonkowski, M. & Clarholm, M. Stimulation of plant growth through interactions of Bacteria and protozoa: testing the auxiliary microbial loop hypothesis. *Acta Protozool.* **51**, 237–247 (2012).
48. Kropinski, A. M., Kuhn, J. H. & Adriaenssens, E. M. To create one (1) new genus, Amigovirus, including one (1) new species in the family Siphoviridae. (2016). <https://doi.org/10.13140/RG.2.2.29767.29608>

49. Philippe, N. et al. Pandoraviruses: amoeba viruses with genomes up to 2.5 mb reaching that of parasitic eukaryotes. *Science* **341**, 281–286 (2013).
50. Benny, G. L. & Smith, M. E. Taxonomic notes on eight species of obligate mycoparasites in the genus *Syncephalis* isolated from soil and Dung. *Mycologia* **112**, 552–569 (2020).
51. Murase, J. et al. Incorporation of plant residue-derived carbon into the microeukaryotic community in a rice field soil revealed by DNA stable-isotope probing. *FEMS Microbiol. Ecol.* **79**, 371–379 (2012).
52. Korytowski, D. A. & Smith, H. Permanence and stability of a kill the winner model in marine ecology. *Bull. Math. Biol.* **79**, 995–1004 (2017).
53. Plassart, P. et al. Evaluation of the ISO standard 11063 DNA extraction procedure for assessing soil microbial abundance and community structure. *PLoS ONE* **7**, e44279 (2012).
54. Santos, S. S. et al. Soil DNA extraction procedure influences protist 18S rRNA gene community profiling outcome. *Protist* **168**, 283–293 (2017).
55. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, 11257 (2016).
56. Jarrige, D. et al. Data for metagenomics reveals contrasted responses of microbial communities to wheat straw amendment in cropland and grassland soils. (2024). <https://doi.org/10.57745/OMJUXC>
57. Lin, H. & Peddada, S. D. Analysis of microbial compositions: a review of normalization and differential abundance analysis. *Npj Biofilms Microbiomes* **6**, 60 (2020).
58. Aitchison, J., Barcelo-Vidal, C. & Martín-Fernández, J. A. Pawlowsky-Glahn, V. Logratio analysis and compositional distance. *Math. Geol.* <https://doi.org/10.1023/A:1007529726302> (2000).
59. Martino, C. et al. A novel sparse compositional technique reveals microbial perturbations. *mSystems* **4**, e00016–19 (2019).
60. Jarrige, D. et al. FONCTIOMIC bioproject: Raw sequence files for metagenomics reveals contrasted responses of microbial communities to wheat straw amendment in cropland and grassland soils. *NCBI*. <https://www.ncbi.nlm.nih.gov/bioproject/1159878> (2024).

Acknowledgements

We would like to thank Samuel Mondy and Marine Martin of the GenoSol platform (Dijon, France) for their soil conservatory expertise and DNA extraction. We would also like to thank Alphonse Thiauw for his participation in the initial statistical analyses during his master internship. This work was supported by the Agence Nationale de Recherche (ANR) with the ANR Systerra project DIMIMOS (ANR-08-STRA-06) and the μ Flu4AMR project (ANR-20-PAMR-0004). It was also supported by INRAE MEM calls (Microbial Ecosystems & Metaomics, call 2018) that funded the FONCTIOMIC (“Linking dynamics of microbial metagenomes with soil organic matter decomposition in agricultural soil”) project.

Author contributions

Conceptualisation: PAM, ST. Methodology: DJ, OR, VL. Software: OR, DJ. Validation: PAM, ST, OR, DJ. Formal Analysis: DJ. Investigation: VT. Resources: AC, VL. Data Curation: DJ, OR, ST. Writing - Original Draft: DJ. Writing - Review & Editing: PAM, ST, VT, AC, OR, DJ, VL, AC. Visualisation: DJ. Supervision: PAM, ST. Project administration: PAM, ST. Funding acquisition: PAM, ST.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-98903-2>.

Correspondence and requests for materials should be addressed to S.T.

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

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025