

Title:

Taxonomic filtering accompanies functional expansion during long-term soil restoration

Short title:

Functional expansion via soil restoration

Authors and affiliations:

Tim Goodall^{1,*}, Susheel Bhanu Busi^{1,*}, Briony Jones², Amy Thorpe¹, Robert I. Griffiths³, John Redhead¹, Lucy Hulmes¹, Sarah Hulmes¹, Lucy Ridding¹, Jodey Peyton¹, Gloria Pereira⁴, Hyun Soon Gweon⁵, Daniel S. Read¹, Richard Pywell¹

¹UK Centre for Ecology & Hydrology, Wallingford, United Kingdom

²UK Centre for Ecology & Hydrology, Bangor, United Kingdom

³School of Environment, Bangor University, Bangor, United Kingdom

© The Author(s) 2026. Published by Oxford University Press on behalf of the International Society for Microbial Ecology. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

⁴UK Centre for Ecology & Hydrology, Lancaster, United Kingdom

⁵School of Biological Sciences, University of Reading, Reading, United Kingdom

*Corresponding authors:

Tim Goodall, tingoo@ceh.ac.uk & Susheel Bhanu Busi, susbus@ceh.ac.uk, UK Centre for Ecology & Hydrology, Wallingford, OX10 8BB, United Kingdom

Abstract

The restoration of species-rich calcareous grasslands is a critical conservation objective, yet the recovery of the invisible below-ground microbiome remains poorly quantified compared to above-ground vegetation. Using a unique 143-year land-use chronosequence on Salisbury Plain, UK, we investigated the trajectory of ecosystem reassembly across arable, regenerating (23 and 67 years), and ancient grasslands. By integrating vegetation surveys with soil physiochemistry, microbial profiling, and shotgun metagenomics, we identified a decoupling between floral and edaphic recovery. While the diversity of vegetation recovered relatively rapidly, approaching ancient grassland levels within 23–67 years, soil properties exhibited persistent legacy effects and slow convergence. Bacterial richness decreased with restoration age; this taxonomic contraction was conversely matched by an expansion in inferred metagenomic functional potential. This was reflected in increased functional gene richness and shifts in the relative abundance of specific SEED-annotated functions toward metabolic pathways associated with complex carbon cycling and stress tolerance. These shifts were congruent with the emergence of specific, unnamed genera belonging to *Pseudomonadota* and *Actinomycetota*, and the *Bacillota* species *Pristimantibacillus*. The soil ecosystem remained distinct from the 143-year stage even after 67 years of recovery, characterised by persistent legacy phosphorus and a slow accumulation of soil organic matter. These findings suggest that passive regeneration alone may be insufficient for full soil functional recovery, and that strategies targeting microbial assembly and long-term carbon dynamics warrant further evaluation.

Key words

Calcareous grassland restoration, land-use chronosequence, microbial succession, soil legacy effects, carbon cycling, above–belowground decoupling.

Introduction

Calcareous grasslands are among the most species-rich ecosystems in Northern Europe, supporting exceptional floral diversity and providing critical ecosystem services, including carbon sequestration and hydrological regulation¹. However, the extent and quality of these habitats have declined precipitously over the last century due to agricultural expansion and intensification². This conversion to agricultural land, typically involving deep tillage and the application of inorganic fertilisers, fundamentally alters the soil's physicochemical architecture and disrupts the complex biotic interactions of the below-ground microbiome³. While the immediate impacts of agricultural intensification are well-documented⁴, the capacity for these ecosystems to recover following the cessation of intensive agricultural practice, using the approach of passive restoration⁵ through natural regeneration⁶, remains a subject of significant ecological debate⁷.

Restoration studies across terrestrial ecosystems have shown that biodiversity recovery is often incomplete relative to reference systems⁷, even where restoration increases diversity and reduces degradation effects. This pattern is particularly relevant in post-agricultural landscapes, where historical land use can leave persistent soil legacies that continue to shape community assembly long after visible vegetation recovery has begun⁸. Studies from a range of restored systems have shown that microbial communities may remain distinct from reference states for years to decades, with elevated nutrient pools, especially phosphorus, implicated as one mechanism limiting recovery⁵.

These studies suggest that post-agricultural restoration should be understood not simply as vegetation reassembly, but as a longer-term reorganisation of soil biotic and abiotic states.

A central uncertainty in restoration ecology is the degree of hysteresis present in soil systems. Above-ground vegetation communities may begin to visibly reassemble once disturbance pressures are removed⁹, soil ecosystems often exhibit a legacy effect^{10,11}. Anthropogenic inputs, particularly elevated levels of phosphorus and nitrogen, can persist for decades¹², potentially locking the ecosystem into an alternative trajectory, inhibiting the return of characteristic grassland biodiversity¹³. It is unclear whether the recovery of soil microbial diversity and function tracks linearly with vegetation recovery, or if it operates on a decoupled, much slower trajectory.

Understanding these temporal dynamics is critical, as microorganisms mediate key biogeochemical cycles that underpin the long-term stability of the restored ecosystem. The need to distinguish above-ground from below-ground recovery is reinforced by a wider literature showing that plant–soil interactions¹⁴ and soil legacies operate across multiple spatial and temporal scales, from rapid rhizosphere responses to long-term shifts in organic matter quality, nutrient availability, and microbial community structure¹⁰. These dynamics are especially relevant in restoration studies based on chronosequences, which provide rare access to long timescales but require caution when inferring mechanism from spatial substitution alone¹⁵. Chronosequence approaches remain highly informative for identifying broad successional patterns in soils and microbial communities, if interpretation remains focused on trajectories and associations rather than direct proof of process.

Our study investigates these dynamics using a unique 143-year chronosequence located on Salisbury Plain in the United Kingdom¹¹. This landscape serves as a rare ‘space-for-time’ substitution experiment, where historical land-use records allow us to identify areas of calcareous grassland with distinct cessation dates of agricultural practices. By analysing soils from active arable land (0 years) and comparing them to grasslands regenerating for 23 years (recent), 67 years (middle), and over 143 years (old/ancient), we can quantify the rate and direction of edaphic functional change and

functional expansion. Here, we use the term functional expansion to refer to an increase in inferred metagenomic functional potential across the chronosequence, reflected by increased functional gene richness and shifts in the relative abundance structure of SEED-annotated functions.

Specifically, we aimed to address three critical knowledge gaps: 1) Does the invisible below-ground ecosystem recover at the same rate as the visible above-ground flora? 2) Do soil nutrient stoichiometries and microbial community structures converge over time, or do they continue to evolve over generational-scale timeframes? 3) Which specific microbial taxa and metabolic functions act as indicators of successful restoration? By integrating vegetation surveys with detailed edaphic analysis, microbial community profiling, and metagenomic sequencing, our research provides a comprehensive view of ecosystem recovery at a landscape scale.

Materials and Methods

Study site and land history

The Salisbury Plain site falls under the management of the UK Ministry of Defence and represents an area of ca. 38,000 ha. As a result of its use as a military training area for over a century, much of the landscape has remained inaccessible for agriculture beyond extensive livestock grazing and thus has not been subject to the agricultural intensification which destroyed calcareous grasslands across much of Western Europe over the twentieth century. Salisbury Plain thus contains Western Europe's largest remaining fragment of lowland calcareous grassland and forms a unique landscape for exploring spatiotemporal patterns and processes in this habitat. Approximately 37% of the current military training area consists of high-quality, species rich calcareous grassland communities, with the remainder being a mix of mesotrophic grassland, agriculturally improved pastures, and arable land. The land history classifications determined previously¹¹ from a time series of historic maps, were used to determine sites suitable for analysis in this study. In total 7 arable, 6 recent, 5 middle, and 7 ancient sites were chosen.

The minimum ages of unimproved grassland are as follows: a) Arable, 000 years denotes areas lost to agricultural improvement and currently managed for arable crops, b) Recent, 023 years; grassland from 1985 to 1996, c) Middle, 067 years; grassland from 1930 to 1967, and d) Ancient, 143 years; grassland from 1840 to 1880. Because land history records do not extend beyond ~1840, the Ancient category represents a minimum age, and earlier site histories and potential heterogeneity cannot be fully resolved.

Soil samples were collected in January 2013. Per site one sample was derived from the homogenisation of five sub-samples, collected along a 100 m linear transect using 5 cm virgin-plastic corers inserted into the soil to a depth of 15 cm (unless encountering an impenetrable chalk horizon). From each transect, soil from below the organic horizon was homogenised and stored in clean plastic bags and transfer to the laboratory for subsequent processing and storage at -20°C .

Vegetation survey

Vegetation surveys were conducted in summer 2013¹¹. Briefly, for each discrete classification (arable, recent, middle, and ancient), the cover of vascular plant species was recorded from five square quadrats (each 2 m x 2 m) placed at 20 m intervals along the same 100 m transect used for soil sampling. Cover of all vascular plant species was recorded using the DAFOR scale¹⁶. DAFOR categories were converted to ordinal abundance scores (Dominant = 5 to Rare = 1; Present = 0.1) following common practice in ecological field studies, allowing calculation of relative abundance for Shannon diversity. Although semi-quantitative, this approach preserves species rank abundance relationships.

Soil physicochemical properties

Subsamples of soil were sent to the commercial soil analysis provider NRM-Cawood Scientific, Bracknell, UK and used to determine soil pH, total, and organic carbon (% dry soil), phosphorus (mg / kg dry soil), potassium (mg / kg dry soil), magnesium (mg / kg dry soil), and soil texture percentages (clay <0.002 mm, silt 0.002-0.05 mm, sand 0.05-2.00 mm). Analysis of soil moisture (%), organic

matter by loss on ignition (LOI % dry soil), total, and organic nitrogen (% dry soil) were undertaken at the UKCEH Centralised Chemistry laboratories (Lancaster) using established protocols¹⁷. C:N ratios were calculated from organic carbon and nitrogen values.

PLFA analysis

Phospholipids were extracted from 1.5 g soil fresh weight according to a previously published methods¹⁸. The following fungal and bacterial markers were summed for each sample and the fungal to bacteria ratio calculated. Fungal markers - C18:1 ω 9c, C18:2 ω 6c, C18:2 ω 9t. Bacterial markers (Gram+ branched, Gram- monoenoics, cyclopropyl, actinomycetes) – “i-C15:0”, “a-C15:0”, “i-C:16:0”, “i-C17:0”, “a-C17:0”, “i-C18:0”, “a-C18:0”, “C16:1 ω 7c”, “C16:1 ω 9c”, “C17:1 ω 10c”, “C18:1 ω 11c”, “C18:1 ω 11t”, “cy-C17:0*”, “9,10-cy-C19:0”, “11,12-cy-C19:0”, “10Me-C16:0”, “10Me-C17:0*”, “10Me-C18:0*”.

DNA extraction

Each sample was defrosted and DNA was extracted from 0.2 g of sample using the Mobio PowerSoil DNA extraction kit (Mobio, US) following manufacturer’s instructions. Briefly, samples were lysed at 25 Hz for 20 minutes using a TissueLyser II (Qiagen, Germany). Purified DNA was eluted in 100 μ L of elution buffer. DNA purity was assessed using a NanoDrop 8000 spectrophotometer (Thermo Fisher, UK), and concentration was measured using the Qubit dsDNA kit (Thermo Fisher, UK). DNA was archived at -20 °C at UKCEH, Wallingford prior to amplicon and metagenomics analysis.

Metagenomic sequencing and data processing

Library preparation and sequencing were conducted by Mr DNA (Texas, US). Samples underwent 2x150 bp shotgun metagenomic sequencing on a HiSeq System (Illumina), targeting a depth of at least 4 Gb raw data per sample. The process generated a mean number of 9.8e+06 raw reads per sample.

sequencing data using SingleM to identify and extract fragments of 14 universal single-copy marker genes (e.g., rplB, rpsC) from the raw reads. These sequences were clustered into Operational Taxonomic Units (OTUs) and taxonomically assigned using the SingleM default database. Taxonomic levels were parsed from the resulting strings and best-name identifiers were generated by selecting the most specific classified level (from species to phylum) for each OTU to ensure informative labelling of uncultured or novel taxa. The community data was normalised by mean coverage implemented within singleM (estimated genome equivalents) and integrated with the soil chronosequence metadata. The fraction of metagenomic reads derived from prokaryotic cells (Bacteria and Archaea) was estimated using *singleM prokaryotic_fraction*.

Invertebrate isolation

Samples for extraction of soil invertebrates were collected in October 2015 (two years after soil sample collection and vegetation surveys). Consistent with previously published method²⁸, four soil plugs were cut per site for Tullgren extraction, a design considered appropriate for site-level comparative analysis, more intensive sampling would be required for scarce or larger-bodied taxa. Briefly, four 10 cm diameter cores were taken at each site by placing a plastic ring on the surface and using a knife to cut the surface vegetation and subsurface roots. The turf and first 10cm of soil were then lifted out as a plug and placed in clean plastic bags. Within 24 hours, these plugs were positioned in Tullgren funnels and all invertebrates leaving the sample were captured and stored in 100% ethanol. The invertebrate catch were subsequently removed from the storage ethanol, dried, and placed into Mobio PowerSoil DNA extraction kit (Mobio, US) for DNA extraction following manufacturer's instructions.

Amplicon sequencing

Amplicon libraries were constructed according using a dual indexing strategy²⁹ with each primer consisting of the appropriate Illumina adapter, 8-nt index sequence, a 10-nt pad sequence, a 2-nt linker and the amplicon specific primer. For the V3-V4 region of 16S rRNA gene (bacteria)

CCTACGGGAGGCAGCAG, and GGACTACHVGGGTWTCTAAT²⁹, ITS2 (fungi) GTGARTCATCGAATCTTTG, and TCCTCCGCTTATTGATATGC³⁰, and COI (invertebrates, predominantly Arthropoda, Mollusca and Nematoda) GGWACWGGWTGAACWGTWTAYCCYCC, and TAIACYTCIGGRTGICCRAARAAYCA³¹. The pooled libraries were sequenced separately on a MiSeq System (Illumina) with V3-600 cycle flow cells and demultiplexed at UKCEH. Sequence tables for 16S and COI amplicons were generated according to a previously published method³², the ITS fungal amplicon sequence table was produced using PIPITS³³. After initial quality checks using R package *microeco*³⁴ sample reads were rarefied to the minimum sample read number per amplicon of 2385, 4792, and 8184 for COI, 16S, and ITS reads respectively.

Statistical analysis

Statistical analyses assessed the recovery of soil biotic properties across the restoration chronosequence. R package *vegan*³⁵ produced alpha diversity metrics, including Shannon index, Species richness, and Pielou's evenness, to quantify structure. The co-variance of measured edaphic variables were assessed with linear models. Variance inflation factors (Supplementary table 2) indicated moderate to high multicollinearity among edaphic predictors, ranging from 2.7 to 48.5. The highest values were associated with soil organic matter (loss on ignition, VIF = 48.5) and organic nitrogen (VIF = 43.4), reflecting their strong correlation. Elevated VIF values were also observed for clay (13.6), phosphorus (10.6), and magnesium (10.0); pH (5.2), moisture (5.2), and potassium (4.2) showed moderate collinearity. In contrast, the C:N ratio exhibited relatively low multicollinearity (VIF = 2.7). The following variables were retained due to their distinct ecological relevance: C:N, organic nitrogen, magnesium, phosphorus, potassium, pH, moisture, and clay. Functional analyses were based on SEED-annotated metagenomic gene profiles and therefore reflect inferred functional potential rather than realised microbial activity. Within this framework, functional gene richness and functional composition were treated as the primary descriptors of functional expansion and correlations with LOI were used as an independent analysis linking functional patterns to soil

development. Simultaneously, functional compositional differences between restoration ages were visualised using Principal Component Analysis (PCA) with vectors of edaphic drivers added. Functional annotations across hierarchy levels (SEED levels 1–4) were reshaped into long format and summarised by restoration age. For each function and time point, mean relative abundance was calculated across samples. To identify functional genes exhibiting consistent directional change over restoration, mean relative abundances were compared across the four time points (000, 023, 067, and 143 years). Functional genes were classified as emergent if they were absent at baseline (mean abundance = 0 at 000 years) and exhibited monotonic increases thereafter, and as declining if they were present at baseline and exhibited monotonic decreases across successive time points, with a net decrease from 000 to 143 years. For each gene function, the magnitude of change was quantified as the difference in mean relative abundance between 143 and 000 years. For visualisation, genetic functions were grouped by SEED hierarchy level and direction of change (emergent or declining), and the top functions within each group were selected based on the absolute magnitude of change. Mean relative abundances at each time point were displayed as a heatmap, with restoration age on the x-axis and functions on the y-axis, ordered by magnitude of change within each ontology level. Tile colour represents mean relative abundance, after \log_{10} -transformation, to enhance contrast among low-abundance functions. This approach emphasises both the temporal dynamics and ontological structure of functional shifts during ecosystem restoration.

We defined two distinct successional trajectories to characterise microbial colonisation. Pioneer or early emergent taxa were those with zero detected coverage at the initial arable stage (000y) that reached a mean coverage threshold of >0.05 by the early restoration stage (023y). And, monotonically increasing taxa demonstrated a consistent, non-decreasing trend in mean coverage across all four time points to 143y, representing stable, long-term colonisers of the developing soil ecosystem.

Direct correlations between genetic functions and soil organic matter (LOI), a key soil health metric, were assessed using Spearman's rank correlation coefficient. To identify microbial functions that tightly track soil development, we performed a correlation-based analysis between SEED functional gene abundances and soil organic matter (LOI %). Given the asymptotic accumulation of LOI observed across the chronosequence stages, Spearman's rank-order correlation (ρ) was utilised to capture monotonic relationships while accounting for sample-level variance ($n=25$). Prior to correlation analysis, functional groups were pre-filtered to include only those exhibiting a consistent emergent or declining trajectory across the successional stages. Multiple testing was accounted for using the Benjamini-Hochberg False Discovery Rate (FDR) adjustment, with significance set at $P < 0.05$. To visualise the transition of functional drivers across different scales of metabolic complexity, the top 8 strongest correlates were selected for each hierarchical SEED level (Levels 1 through 4). Results were visualised using a lollipop plot, where the distance from the zero-intercept and colour intensity denote the strength and direction of the successional coupling.

Results

Responses across the chronosequence

To understand how the passive restoration of these soils might influence various soil and vegetative properties, we assessed several factors over the course of the recovery period. Analysis of the rates of change (Figure 1) showed that the magnitude of change varied across the chronosequence, with the largest shifts occurring early in restoration. The highest rates of change occurred in the initial, 000-to-023-year, interval, driven by very large increases in vegetation richness and decreases in soil phosphorus. However, the system did not stabilise after this initial burst. Between 067 and 143 years, significant increases continued to occur (Supplementary figures 1, 2, and table 3) particularly in the variables of soil organic matter (LOI; 20 to 23%), organic nitrogen (0.8 to 1.0%), magnesium (112 to 143 mg / kg), and the eukaryotic : bacterial gene ratio (0.48 to 0.57) . This indicates that even after

nearly seven decades of recovery, the soil ecosystem remained distinct from 143-year ancient grassland stage.

Vegetative and edaphic recovery is decoupled

The restoration trajectories revealed a decoupling between plant community reassembly and soil chemical recovery (Figure 1). Soil chemical variables exhibited a slower, continuous recovery, where soil organic matter (LOI) and nitrogen (N) concentration increased monotonically from the arable baseline to the 143-year ancient grasslands (Figure 1). This increase in organic matter and nitrogen likely reflects the progressive accumulation of increasingly complex organic matter pools during succession, in which nitrogen is contained within stabilised or recalcitrant forms. In contrast, legacies of agricultural fertilisation persisted with insignificantly elevated levels of phosphorous (P) and potassium (K) in 023 and 067 year restoration sites compared to ancient grasslands (Supplementary figure 2 and table 3).

Vegetation richness and diversity (Shannon index) responded more rapidly than soil chemical properties after agricultural cessation (Figure 1), increasing significantly to 023 years (~300%), a further increase in richness occurs 023 to 067 years and then gradual, but non-significant, increases occur between 067-to-143-year stage (Supplementary figure 1).

Microbial variance over time

The recovery of the soil microbiome followed a trajectory distinct from that of the vegetative community (Figure 1), characterised by an inverse relationship between taxonomic diversity and genetic functional potential. Contrary to the trajectory observed in plant communities, soil bacterial diversity decreased with restoration age. Both bacterial richness and diversity were significantly higher in arable soils (000 years) (Supplementary figure 1). The cessation of agricultural practices resulted in a sharp decline in bacterial richness during the initial restoration phase, with richness remaining lower and relatively stable across the older grasslands. Despite the reduction in taxonomic

bacterial richness, the inferred metagenomic functional potential of the soil microbiome expanded with ecosystem age. Functional gene richness was lowest in the arable sites but increased significantly within the first 23 years of restoration (Supplementary figure 1). This pattern represents the primary sense in which we refer to functional expansion across the chronosequence. Unlike bacterial richness, which dropped and plateaued, functional genetic richness remained elevated in the older grassland soils, suggesting that the communities become genetically more complex over time despite being comprised of fewer bacterial taxa. Concurrently, the composition of the soil community exhibited a monotonic shift toward non-prokaryotic dominance across the chronosequence. The SingleM derived eukaryotic to bacterial gene ratio (Figure 1), utilised as a metric for broad structural change, recorded lowest ratios in the arable soils and significant increases with time (*singlem_fb_ratio*, Supplementary figure 1). This indicates a continuous proliferation of fungal and other non-prokaryotic organisms, with the highest ratios observed in the ancient (143 year) samples.

Microbial community structure and emergent taxa

Despite the reduction in taxonomic richness, the inferred genetic functional potential of the microbiome expanded over time. Soil genetic richness showed a significant positive response during the 000-to-023-year interval (Figure 1, Supplementary figure 1), this shift was congruent with the early establishment of emergent taxa that were absent in arable soils (Figure 2a). By year 023, the microbial community exhibited specific pioneer genera and species with the top five taxa belonging to *Pseudomonadota*, *Actinomycetota*, including a genus of the episymbiont *Nanobdellota*. Over the course of the chronosequence taxa that demonstrate monotonic increase (Figure 2b) include specific genera of *Pseudomonadota*, *Actinomycetota*, and *Pristimantibacillus* sp. The prominence of *Actinomycetota* and *Bacillota* suggests a functional reorientation toward the decomposition of complex organic matter and resource-conservative life strategies.

Distinct functional states and metabolic shifts

To understand how the recovery of the soils influenced microbial community functions, we used an ordination method which indicated that soil functional composition did not converge on a single recovered state but instead differed across chronosequence stages. This revealed clear clustering based on restoration age, with the first two principal components explaining 18.6% of the total variance (Figure 3). The first two principal components indicated separation between the arable and early restoration (000 and 023 year) sites from the ancient grasslands (143 year). This functional separation was associated with opposing environmental vectors, where arable and early restoration sites were associated with legacies of agricultural improvement (K, P, and clay content), while ancient sites were congruent with vectors associated with ecosystem stability and organic matter sequestration (LOI, Moisture, and C:N ratio). pH, a known microbial master variable did not emerge as a significant driver of functional composition (in-laid table, Figure 3 and Supplementary figure 2), highlighting soil organic matter accrual as a prominent edaphic correlate of functional composition in this system. The 67-year sites occupied an intermediate ordination space, further indicating that functional composition maturation is not achieved after six decades of recovery.

Emergence and loss of specialised functions and pathways

We identified a suite of functional genes that exhibited monotonic increases or decreases in abundance with restoration age (Figure 4). Restoration of soil led to notable decreases in the abundance of genes associated with membrane transport (ABC-type multidrug transport system), ribosomes (SSU ribosomal proteins), archaeal protein recycling (archaeal protease subunits), DNA repair (glycosylase), and phage proteins (phage tail proteins). Overall, the recovery of soil condition leads to a reduction in genes associated with growth, maintenance, and disease, along with broader reductions in the subsystems associated with metabolism of proteins, RNA, and DNA. Whilst these functions decreased with restoration age, monotonic increases in genes and pathways along the chronosequence were also observed. Subsystems for photosynthesis and motility were more abundant in older soils, consistent with the establishment of mobile, autotrophic or mixotrophic

Discussion

The results of this calcareous grassland study challenge the assumption that ecosystem recovery is a synchronous process where below-ground health mirrors above-ground biodiversity. Instead, we observe a complex, multi-phasic recovery where soil properties lag significantly behind the recovery of vegetation.

Decoupling of vegetation and soil diversity recovery

The results of our study reveal a disconnect between the recovery of above-ground floristic diversity and below-ground edaphic genetic function. Because our study is based on a space-for-time chronosequence rather than repeated sampling of the same sites through time, the observed recovery patterns should be interpreted as consistent with, rather than definitive evidence of, particular ecological mechanisms. Congruent with earlier work^{11,40-42}, we found that increases in plant diversity and species richness are rapidly achievable. In our study, vegetation richness increased significantly in the first 23 years and continued to rise before stabilising from 067 years, suggesting a relatively rapid recovery of taxonomic diversity within a human lifetime, compared with the slower recovery of soil properties. However, broad diversity metrics can obscure underlying functional deficits. As observed in 2014¹¹ within this same landscape, even in well-connected landscapes, ancient grasslands show distinct vegetation communities and traits, suggesting that natural regeneration to a full ancient calcareous grassland community takes over a century. While the richness of the vegetation may approach ancient levels relatively quickly, the functional assembly of the plant community remains distinct for much longer. This delay mirrors the persistent legacy effects and slow convergence we observe in the soil ecosystem, which exhibited continued diversification and functional evolution well beyond 067 years. Thus, we observe a two-speed recovery model with a rapid biological reassembly of plant taxonomic richness, followed by a protracted, multi-generational recovery of plant functional traits, soil chemistry, and microbiome architecture. Another potential explanation for the observed disparity may be due to the increased

organic matter carrying capacity of calcareous soils within this landscape. For example, previously studies reported greater than 20% OM in similar soils⁴³. We will however need future empirical assessments to disentangle these complex interactions.

The persistence of agricultural legacies and carbon dynamics

The contrast in recovery rates is associated with the persistence of agricultural legacies^{8,44}. The elevated levels of available P and K in the younger restoration sites indicate that the eutrophication caused by historical fertilisation takes decades to attenuate. This nutrient enrichment may help explain the delayed establishment of the stress-tolerant, oligotrophic microbial specialists characteristic of ancient soils⁴⁵, and acts as a potential mechanism explaining the observation that in many studies, sites appear to stall at a stable, mesotrophic community for long periods^{8,11,46}.

Conversely, the accumulation of C and N follow a monotonic trajectory that did not saturate even after 067 years. Increasing total organic carbon and nitrogen should not be interpreted as evidence of increasingly copiotrophic conditions with more readily available nutrients for microbial growth. As succession proceeds, carbon and nitrogen is increasingly incorporated into complex organic matter pools⁴⁷, such that total carbon and nitrogen become decoupled from nutrient bioavailability. Older sites, despite higher organic carbon and nitrogen content, are associated with taxa indicative of slower-growing, resource-conservative life strategies⁴⁸. These patterns suggest that the carbon sequestration potential of calcareous grasslands is not a rapid solution but a long-term service that shows no evidence of saturation within the first 143 years of recovery and may represent a multi-generational process. Consequently, carbon credit schemes or conservation policies based on short-term restoration (<30 years) may severely underestimate the time required to restore the full soil carbon sink in these systems.

Functional redundancy at high microbial diversity

A key finding of our study is the inverse relationship between bacterial richness and ecosystem maturity. Here, as per previous studies, arable soils supported the highest bacterial richness and

community structure^{49,50}, did not change in this system, usefully decoupling it as a driver of observed genetic functional change.

Incomplete functional convergence and conservation implications

This study demonstrates that below-ground recovery remains incomplete within the timescale captured by the chronosequence. The ordination and rate-of-change analyses demonstrated that 67-year-old grasslands are genetically functionally distinct from 143-year-old ancient sites. The trajectory of edaphic change, particularly in C, moisture, and Mg, suggest that restoration is a continuous process spanning multiple human generations⁵⁶. This challenges binary definitions of restored versus degraded land⁵⁷. Therefore, conservation policy must recognise that within a temperate calcareous landscape 60-year-old regeneration sites are valuable but transitional ecosystems, requiring continued protection to allow the slow accumulation of the functional complexity and specialised metabolic depth which characterises these ancient soils.

Conclusion and future research

Looking forward, this 143-year chronosequence necessitates a paradigm shift in how we value and manage restoring calcareous grassland landscapes, moving beyond the assumption that ecosystem recovery is a rapid or synchronous process. The persistence of agricultural nutrient legacies and the lack of functional convergence even after seven decades demonstrate that current conservation frameworks, often bound by short-term funding cycles, are ill-equipped to capture the full trajectory of edaphic recovery. To address the temporal gap between rapid vegetation recovery and slower soil functional expansion, future restoration studies could examine approaches that extend beyond passive regeneration. Our results are consistent with the possibility that dispersal limitation and persistent nutrient legacies could constrain the re-assembly of specialised microbial communities. A future hypothesis to test is targeted soil inoculation whereby early restoration sites receive small amounts of donor soil from long-established grasslands to reintroduce oligotrophic taxa.

Experimental studies have demonstrated that whole-soil inoculation can help overcome dispersal

barriers and steer plant and soil community assembly toward reference states, accelerating the development of late-successional functions⁵⁸. Although not tested here, such approaches may enhance the recovery of tightly coupled nutrient cycling processes that often remain incomplete after decades of spontaneous regeneration⁵⁹.

The persistence of elevated phosphorus in younger sites suggests that nutrient legacies may act as a bottleneck for microbial successional speed. Future research should investigate whether mitigating these legacies, through targeted nutrient reduction or sequestration strategies, can accelerate microbial reassembly trajectories.

Within our study protists were not explicitly profiled, given their important roles as microbial predators and regulators of nutrient turnover, incorporating protist community dynamics would be a valuable next step in understanding whole-soil food web recovery during restoration.

Our findings highlight a potential mismatch between rapid vegetation recovery and slower soil functional maturation. Specifically, our findings suggest that future monitoring frameworks could consider incorporating below-ground indicators alongside above-ground metrics. For example, metagenomic assessment of genes associated with stress tolerance and carbon cycling could provide complementary indicators of functional development. More broadly, the continued increase in soil organic matter across this calcareous chronosequence suggests that carbon accrual operates over multi-decadal to centennial timescales. Collectively, this reinforces the importance of aligning restoration expectations with the protracted temporal dynamics of soil ecosystem recovery.

Study funding, conflicts of interest statement, and Data and code availability

UK Research and Innovation:

BBX011089/1 (Earlham Institute's Strategic Programme Grant: Decoding Biodiversity)

NE/S005137/1 (LOCKED UP: The role of biotic and abiotic interactions in the stabilisation and persistence of soil organic carbon)

The authors declare that there exist no conflicts of interest.

Sequence data is available at NCBI-SRA under BioProject ID: PRJNA1424699. Scripts and data tables are available at <https://doi.org/10.5281/zenodo.18632861>

References

- 1 Newton, A. C. et al. Structure, composition and dynamics of a calcareous grassland metacommunity over a 70-year interval. *J. Ecol.* 100, 196–209 (2012). <https://doi.org/10.1111/j.1365-2745.2011.01923.x>
- 2 Ridding, L. E. et al. Long-term change in calcareous grassland vegetation and drivers over three time periods between 1970 and 2016. *Plant Ecol.* 221, 377–394 (2020). <https://doi.org/10.1007/s11258-020-01016-1>
- 3 Mo, Y. et al. Agricultural practices influence soil microbiome assembly and interactions at different depths identified by machine learning. *Commun. Biol.* 7, 1349 (2024). <https://doi.org/10.1038/s42003-024-07059-8>
- 4 Firbank, L. G., Petit, S., Smart, S. et al. Assessing the impacts of agricultural intensification on biodiversity: a British perspective. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 777–787 (2008). <https://doi.org/10.1098/rstb.2007.2183>
- 5 Prach, K., Šebelíková, L., Řehouňková, K. & del Moral, R. Possibilities and limitations of passive restoration of heavily disturbed sites. *Landsc. Res.* 45, 247–253 (2020). <https://doi.org/10.1080/01426397.2019.1593335>
- 6 Ye, Y. et al. Natural regeneration enhances long-term soil carbon storage in various fractions more effectively than active restoration: a meta-analysis. *Glob. Change Biol.* 31, e70255 (2025). <https://doi.org/10.1111/gcb.70255>
- 7 Atkinson, J. et al. Terrestrial ecosystem restoration increases biodiversity and reduces its variability, but not to reference levels: a global meta-analysis. *Ecol. Lett.* 25, 1725–1737 (2022). <https://doi.org/10.1111/ele.14025>
- 8 Peddle, S. D. et al. Agricultural land-use legacies affect soil bacterial communities following restoration in a global biodiversity hotspot. *Biol. Conserv.* 290, 110437 (2024). <https://doi.org/10.1016/j.biocon.2023.110437>
- 9 Li, K., Veen, G. F. C., Ten Hooven, F. C. et al. Soil legacy effects of plants and drought on aboveground insects in native and range-expanding plant communities. *Ecol. Lett.* 26, 37–52 (2023). <https://doi.org/10.1111/ele.14129>
- 10 Nieland, M. A. & Keiser, A. D. Using ecological restoration to disentangle the mechanisms and longevity of soil functional legacies. *Funct. Ecol.* 39, 2193–2204 (2025). <https://doi.org/10.1111/1365-2435.70123>
- 11 Redhead, J. W. et al. The natural regeneration of calcareous grassland at a landscape scale: 150 years of plant community re-assembly on Salisbury Plain, UK. *Appl. Veg. Sci.* 17, 408–418 (2014). <https://doi.org/10.1111/avsc.12076>

- 12 Menge, D. N., Hedin, L. O. & Pacala, S. W. Nitrogen and phosphorus limitation over long-term ecosystem development in terrestrial ecosystems. *PLoS One* 7, e42045 (2012). <https://doi.org/10.1371/journal.pone.0042045>
- 13 Wang, Y. et al. Grassland degradation affects the response of soil bacterial and plant but not fungal diversity to nitrogen addition. *J. Ecol.* 111, 2269–2280 (2023). <https://doi.org/10.1111/1365-2745.14179>
- 14 Frouz, J. Plant–soil feedback across spatiotemporal scales from immediate effects to legacy. *Soil Biol. Biochem.* 189, 109289 (2024). <https://doi.org/10.1016/j.soilbio.2023.109289>
- 15 Walker, L. R., Wardle, D. A., Bardgett, R. D. & Clarkson, B. D. The use of chronosequences in studies of ecological succession and soil development. *J. Ecol.* 98, 725–736 (2010). <https://doi.org/10.1111/j.1365-2745.2010.01664.x>
- 16 Kershaw, K. A. & Looney, J. H. H. *Quantitative and Dynamic Plant Ecology.* (Hodder Arnold, 1985).
- 17 Reynolds, B. et al. Countryside survey: national “soil change” 1978–2007 for topsoils in Great Britain—acidity, carbon, and total nitrogen status. *Vadose Zone J.* 12, vzj2012.0114 (2013). <https://doi.org/10.2136/vzj2012.0114>
- 18 Whitaker, J. et al. Microbial carbon mineralization in tropical lowland and montane forest soils of Peru. *Front. Microbiol.* 5, 720 (2014). <https://doi.org/10.3389/fmicb.2014.00720>
- 19 Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 28, 2520–2522 (2012). <https://doi.org/10.1093/bioinformatics/bts480>
- 20 Andrews, S. *FastQC: a quality control tool for high throughput sequence data.* (Babraham Bioinformatics, 2010).
- 21 Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048 (2016). <https://doi.org/10.1093/bioinformatics/btw354>
- 22 Woodcroft, B. J. et al. Comprehensive taxonomic identification of microbial species in metagenomic data using SingleM and Sandpiper. *Nat. Biotechnol.* 34, 610–618 (2025). <https://doi.org/10.1038/s41587-025-02738-1>
- 23 Li, D., Liu, C.-M., Luo, R. et al. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676 (2015). <https://doi.org/10.1093/bioinformatics/btv033>
- 24 Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119 (2010). <https://doi.org/10.1186/1471-2105-11-119>
- 25 Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I. et al. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829 (2021). <https://doi.org/10.1093/molbev/msab293>
- 26 Overbeek, R. et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, D206–D214 (2014). <https://doi.org/10.1093/nar/gkt1226>

- 27 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930 (2013). <https://doi.org/10.1093/bioinformatics/btt656>
- 28 Arnott, A., Riddell, G., Emmerson, M. et al. Upland grassland habitats and agri-environment schemes change soil microarthropod abundance. *J. Appl. Ecol.* 58, 2256–2265 (2021). <https://doi.org/10.1111/1365-2664.13933>
- 29 Kozich, J. J., Westcott, S. L., Baxter, N. T. et al. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120 (2013). <https://doi.org/10.1128/aem.01043-13>
- 30 Ihrmark, K. et al. New primers to amplify the fungal ITS2 region—evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol. Ecol.* 82, 666–677 (2012). <https://doi.org/10.1111/j.1574-6941.2012.01437.x>
- 31 Leray, M. et al. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front. Zool.* 10, 34 (2013). <https://doi.org/10.1186/1742-9994-10-34>
- 32 Jones, B. et al. Beyond taxonomic identification: integration of ecological responses to a soil bacterial 16S rRNA gene database. *Front. Microbiol.* 12, 682886 (2021). <https://doi.org/10.3389/fmicb.2021.682886>
- 33 Gweon, H. S. et al. PIPITS: an automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. *Methods Ecol. Evol.* 6, 973–980 (2015). <https://doi.org/10.1111/2041-210X.12399>
- 34 Liu, C., Cui, Y., Li, X. & Yao, M. microeco: an R package for data mining in microbial community ecology. *FEMS Microbiol. Ecol.* 97, fiae255 (2020). <https://doi.org/10.1093/femsec/fiae255>
- 35 Oksanen, J. et al. Vegan: Community Ecology Package. R Package Version 2.2-1 2, 1–2 (2015).
- 36 Ramoneda, J. et al. Ecological relevance of flagellar motility in soil bacterial communities. *ISME J.* 18, wrae067 (2024). <https://doi.org/10.1093/ismejo/wrae067>
- 37 Durán, G. A., Sacristán, D., Farrús, E. & Vadell, J. Towards defining soil quality of Mediterranean calcareous agricultural soils: reference values and potential core indicator set. *Int. Soil Water Conserv. Res.* 12, 145–155 (2024). <https://doi.org/10.1016/j.iswcr.2023.06.001>
- 38 Abshiba, K. et al. A comprehensive review of soil quality indicators under diverse land uses in rainfed areas. *Discover Soil* 2, 91 (2025). <https://doi.org/10.1007/s44378-025-00109-5>
- 39 Bahram, M. et al. Structure and function of the global topsoil microbiome. *Nature* 560, 233–237 (2018). <https://doi.org/10.1038/s41586-018-0386-6>
- 40 Zobel, M., Suurkask, M., Rosén, E. & Pärtel, M. The dynamics of species richness in an experimentally restored calcareous grassland. *J. Veg. Sci.* 7, 203–210 (1996). <https://doi.org/10.2307/3236320>
- 41 Pywell, R. F. et al. Plant traits as predictors of performance in ecological restoration. *J. Appl. Ecol.* 40, 65–77 (2003). <https://doi.org/10.1046/j.1365-2664.2003.00762.x>

- 42 Fagan, K. C., Pywell, R. F., Bullock, J. M. & Marrs, R. H. Do restored calcareous grasslands on former arable fields resemble ancient targets? The effect of time, methods and environment on outcomes. *J. Appl. Ecol.* 45, 1293–1303 (2008). <https://doi.org/10.1111/j.1365-2664.2008.01492.x>
- 43 Armbruster, M. et al. Bacterial and archaeal taxa are reliable indicators of soil restoration across distributed calcareous grasslands. *Eur. J. Soil Sci.* 72, 2430–2444 (2021). <https://doi.org/10.1111/ejss.12977>
- 44 Fraterrigo, J. M. Land-use legacies and the role of persistence traits in species recovery. *Appl. Veg. Sci.* 19, 555–556 (2016). <https://doi.org/10.1111/avsc.12276>
- 45 Bledsoe, R. B., Goodwillie, C. & Peralta, A. L. Long-term nutrient enrichment of an oligotroph-dominated wetland increases bacterial diversity in bulk soils and plant rhizospheres. *mSphere* 5, e00035-20 (2020). <https://doi.org/10.1128/mSphere.00035-20>
- 46 Yang, Z. et al. Losses of low-germinating, slow-growing species prevent grassland composition recovery from nutrient amendment. *Glob. Change Biol.* 30, e17264 (2024). <https://doi.org/10.1111/gcb.17264>
- 47 Schmidt, M. W. I. et al. Persistence of soil organic matter as an ecosystem property. *Nature* 478, 49–56 (2011). <https://doi.org/10.1038/nature10386>
- 48 Fierer, N. et al. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J.* 6, 1007–1017 (2012). <https://doi.org/10.1038/ismej.2011.159>
- 49 Griffiths, R. I. et al. The bacterial biogeography of British soils. *Environ. Microbiol.* 13, 1642–1654 (2011). <https://doi.org/10.1111/j.1462-2920.2011.02480.x>
- 50 Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. U. S. A.* 103, 626–631 (2006). <https://doi.org/10.1073/pnas.0507535103>
- 51 Jangid, K., Whitman, W. B., Condon, L. M. et al. Soil bacterial community succession during long-term ecosystem development. *Mol. Ecol.* 22, 3415–3424 (2013). <https://doi.org/10.1111/mec.12325>
- 52 Le Bagousse-Pinguet, Y. et al. Phylogenetic, functional, and taxonomic richness have both positive and negative effects on ecosystem multifunctionality. *Proc. Natl. Acad. Sci. U. S. A.* 116, 8419–8424 (2019). <https://doi.org/10.1073/pnas.1815727116>
- 53 van der Putten, W. H. et al. Plant–soil feedbacks: the past, the present and future challenges. *J. Ecol.* 101, 265–276 (2013). <https://doi.org/10.1111/1365-2745.12054>
- 54 Fierer, N., Bradford, M. A. & Jackson, R. B. Toward an ecological classification of soil bacteria. *Ecology* 88, 1354–1364 (2007). <https://doi.org/10.1890/05-1839>
- 55 Ho, A., Di Lonardo, D. P. & Bodelier, P. L. Revisiting life strategy concepts in environmental microbial ecology. *FEMS Microbiol. Ecol.* 93, fix006 (2017). <https://doi.org/10.1093/femsec/fix006>
- 56 Mori, A. S. & Isbell, F. Untangling the threads of conservation: a closer look at restoration and preservation. *J. Appl. Ecol.* 61, 215–222 (2024). <https://doi.org/10.1111/1365-2664.14552>
- 57 du Toit, J. T. & Pettorelli, N. The differences between rewilding and restoring an ecologically degraded landscape. *J. Appl. Ecol.* 56, 2467–2471 (2019). <https://doi.org/10.1111/1365-2664.13487>

58 Wubs, E. R., van der Putten, W. H., Bosch, M. & Bezemer, T. M. Soil inoculation steers restoration of terrestrial ecosystems. *Nat. Plants* 2, 16107 (2016). <https://doi.org/10.1038/nplants.2016.107>

59 Koziol, L. & Bever, J. D. The missing link in grassland restoration: arbuscular mycorrhizal fungi inoculation increases plant diversity and accelerates succession. *J. Appl. Ecol.* 54, 1301–1309 (2017). <https://doi.org/10.1111/1365-2664.12843>

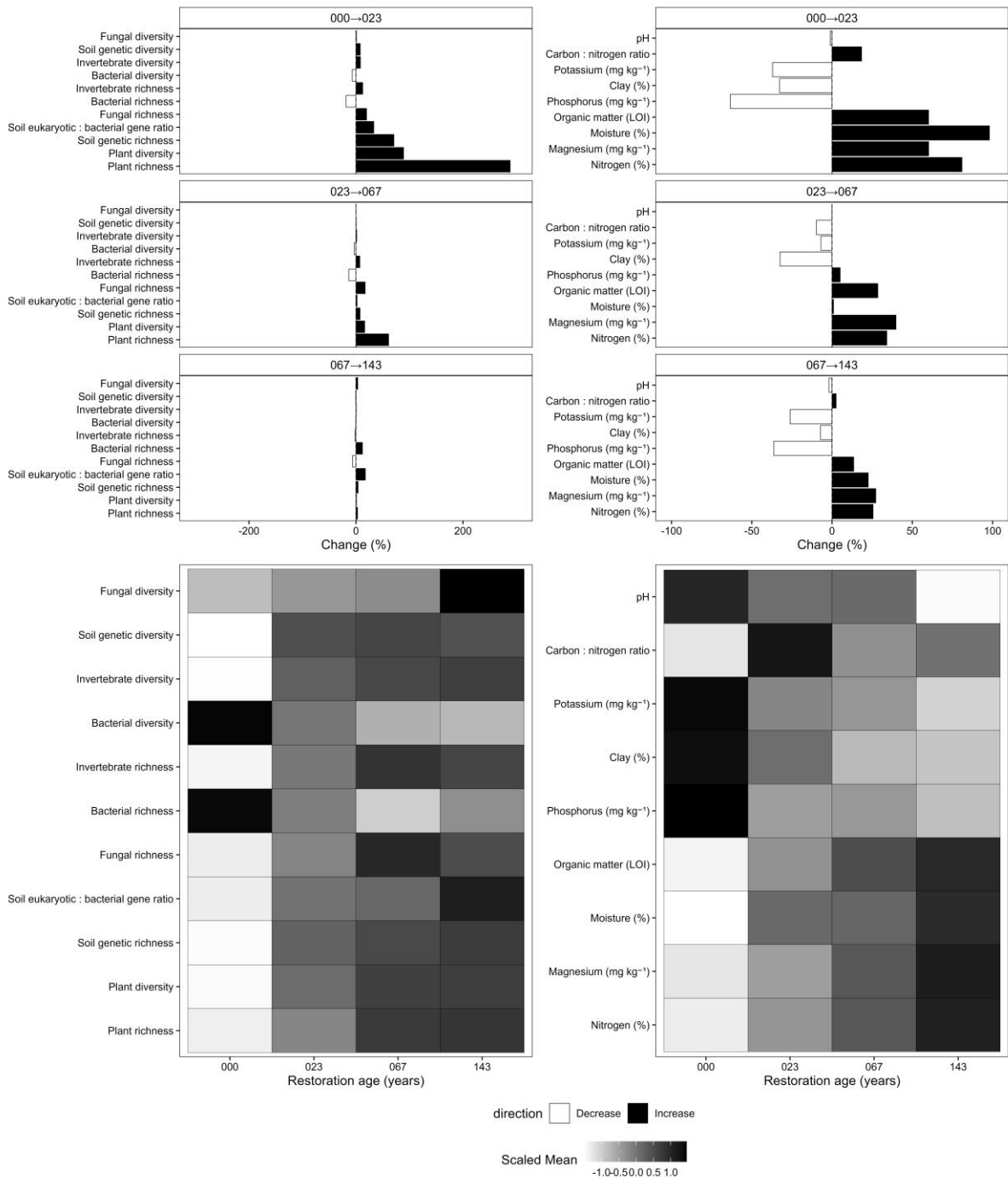


Figure 1. System response dynamics across a restoration chronosequence. Response speed of biological (left) and edaphic (right) variables shown as percent change between successive restoration intervals (000–023, 023–067, and 067–143 years), and standardised trajectories of the same variables across the full chronosequence. Bars indicate the magnitude and direction of change within each interval, while heatmaps illustrate relative temporal trajectories across restoration ages. Variables are ordered by the absolute magnitude of percent change across the chronosequence to facilitate comparison of response strength and temporal dynamics.

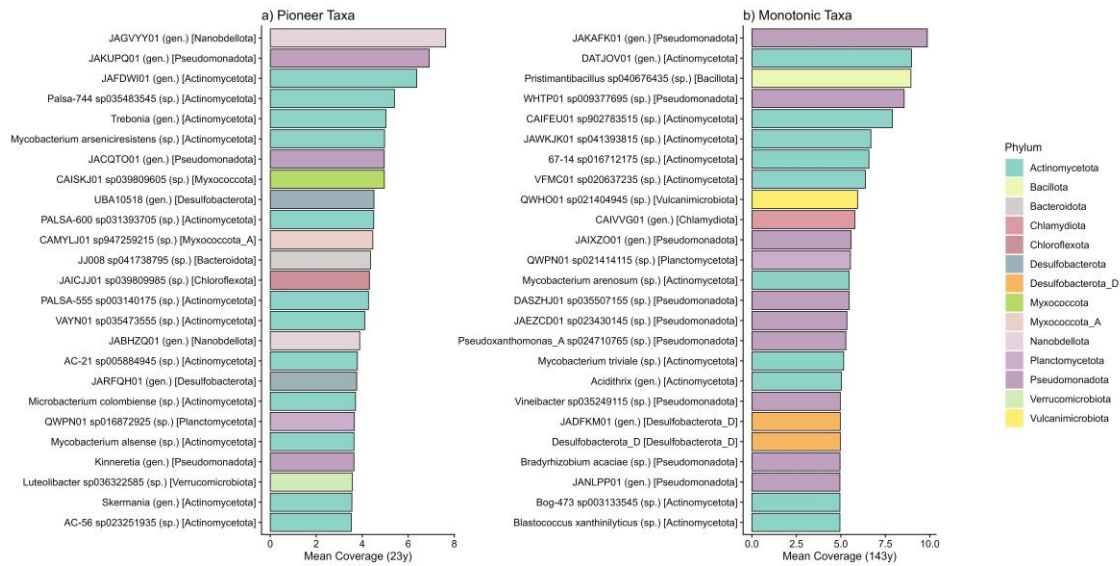


Figure 2. Successional dynamics of microbial taxa across the soil chronosequence as determined by SingleM marker gene analysis. Taxa are ranked by mean coverage (estimated genome equivalents) and labelled with the most specific classified taxonomic level, with the associated phylum provided in brackets. (a) Pioneer Taxa: The top 25 taxa that were undetected at the 0-year (arable) stage but became established by year 23 of restoration. (b) Monotonically Increasing Taxa: The top 25 taxa that exhibited a continuous increase in abundance from the arable stage through to the 143-year established grassland stage. Bars are coloured by phylum to illustrate the broad phylogenetic shifts accompanying the transition from agricultural land to established natural grassland. Taxa labels include taxonomic rank indicators: (sp.) species, (gen.) genus.

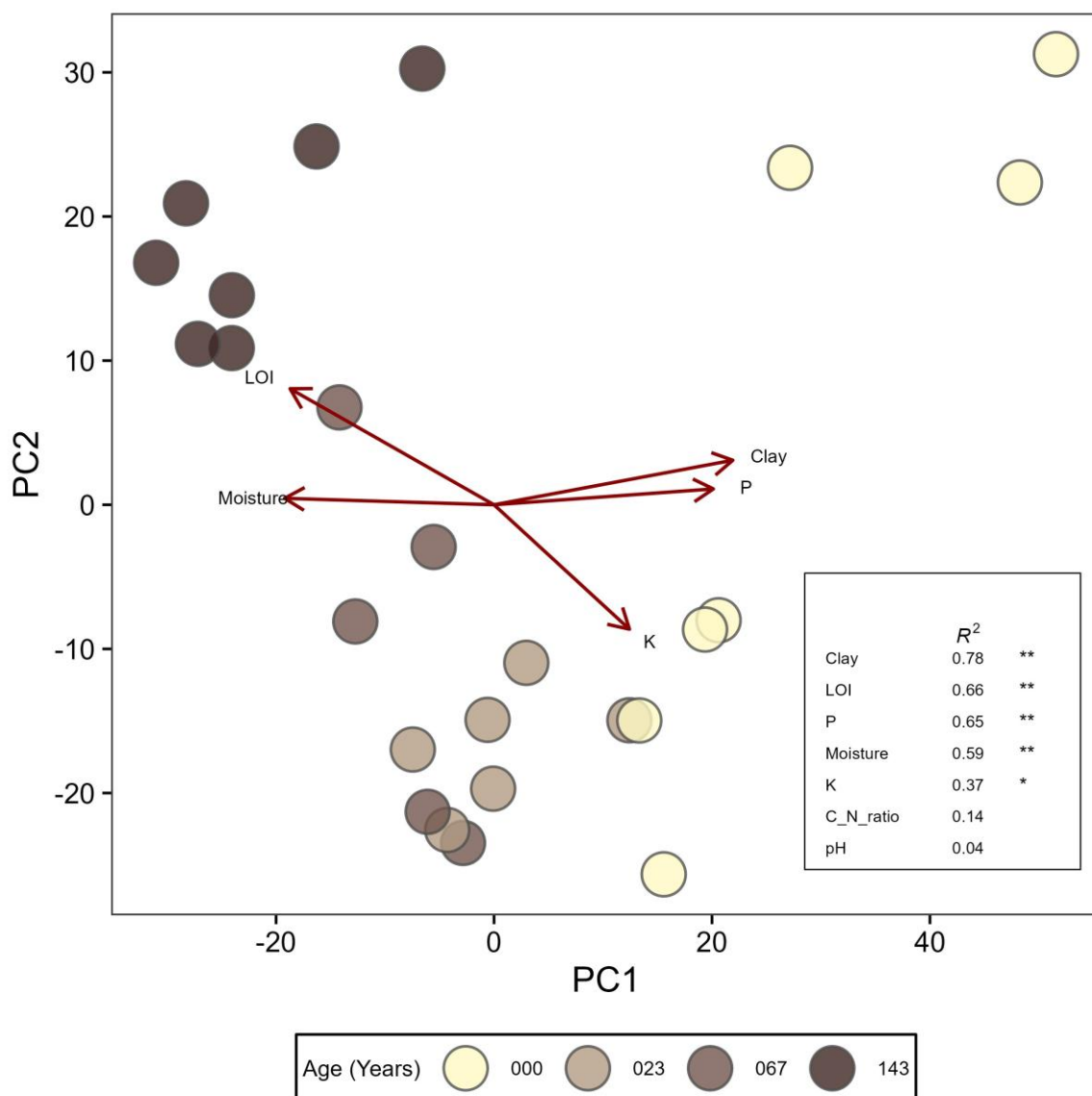


Figure 3. PCA of microbial functional groups with significant environmental drivers. Points represent functional composition, coloured by chronosequence age. Vectors represent significant environmental variables ($P < 0.05$) based on envfit analysis. Inset table shows goodness-of-fit (R^2) and significance codes for all edaphic parameters.

UNCORRECTED

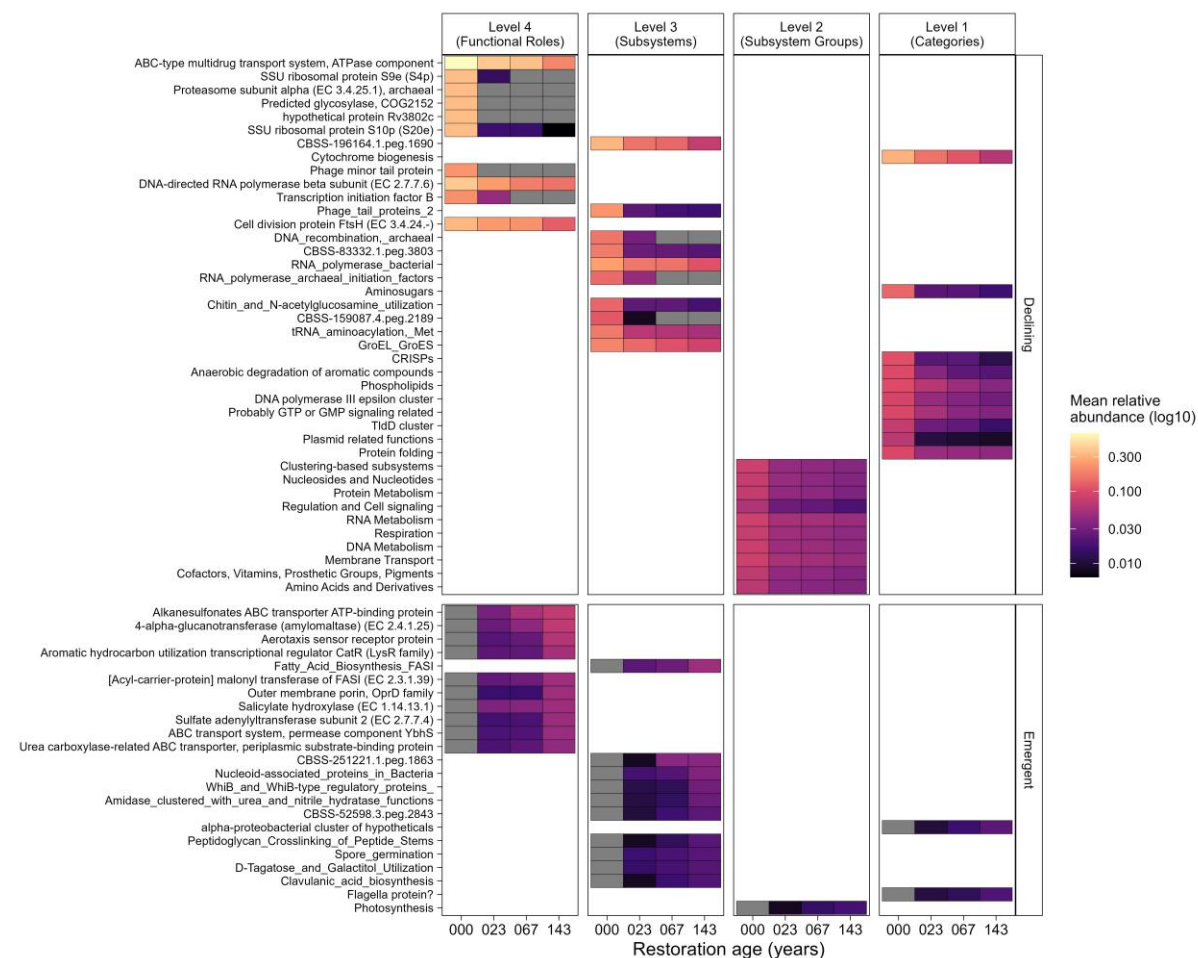


Figure 4. Ontology-resolved temporal dynamics of microbial functional potential across restoration. Heatmap showing mean relative abundances of SEED-annotated microbial functions across restoration age (000, 023, 067, and 143 years). Functions are grouped by SEED hierarchy level (columns) and by direction of change over time (rows), classified as emergent (net increase from 000 to 143 years) or declining (net decrease). Within each hierarchy level and direction, functions are ranked by the magnitude of absolute change in mean relative abundance between 000 and 143 years, and the top functions are displayed. Tile colour indicates mean relative abundance at each time point, shown on a log₁₀-transformed colour scale to enhance contrast among low-abundance functions. This representation highlights both the timing and ontology of functional shifts during ecosystem restoration.

UNCORRECTED

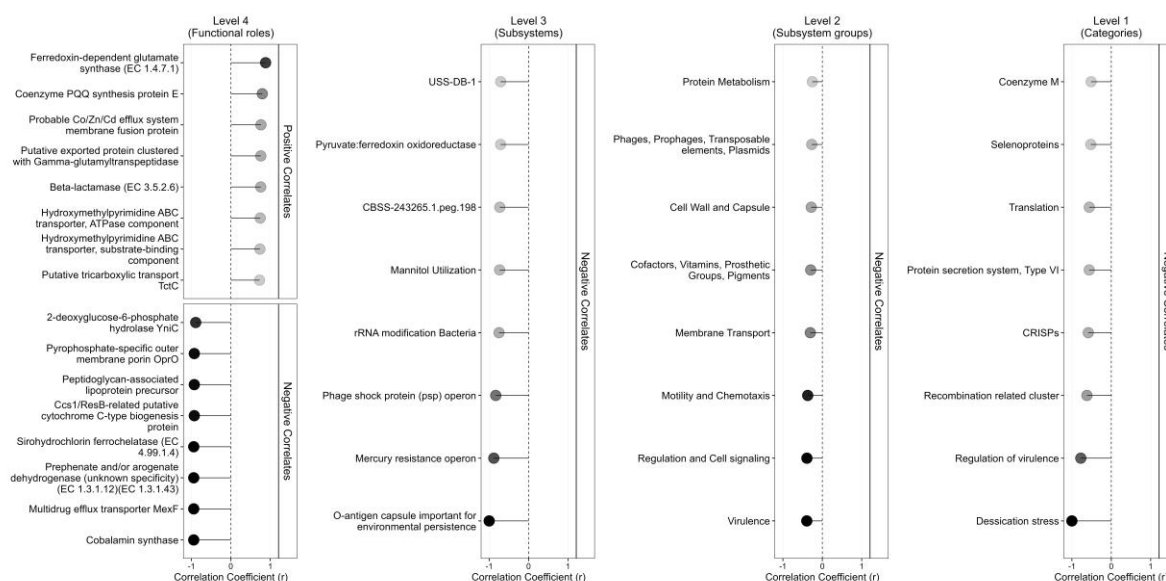


Figure 5. Key functional microbial correlates of soil organic matter (LOI) accumulation. Correlates were identified by Spearman's rank correlation (ρ) between the relative abundance of SEED functional groups and soil organic matter content (Loss on Ignition, LOI %) across all soil samples ($n=25$). The lollipop plot displays the top 8 significantly correlated functions ($P < 0.05$) per hierarchical SEED level (Levels 1–4). The position and opacity of the points represent the correlation coefficient with LOI.

UNCORRECTED