Arable soil nitrogen dynamics reflect organic inputs via the extended composite phenotype

Nature Food

Neal, Andrew L.; Barrat, Harry A.; Bacq-Lebreuil, Aurélie; Qin, Yuwei; Zhang, Xiaoxian et al

https://doi.org/10.1038/s43016-022-00671-z

nature food

Article

https://doi.org/10.1038/s43016-022-00671-z

Arable soil nitrogen dynamics reflect organic inputs via the extended composite phenotype

Received: 4 February 2022

Accepted: 14 November 2022

Published online: 23 December 2022

Check for updates

Andrew L. Neal[1] ✉, Harry A. Barrat[1,11], Aurélie Bacq-Lebreuil[2,12], Yuwei Qin[3], Xiaoxian Zhang[4], Taro Takahashi[1,5], Valentina Rubio[6,7], David Hughes[8], Ian M. Clark[4], Laura M. Cárdenas[1], Laura-Jayne Gardiner[9], Ritesh Krishna[9], Margaret L. Glendining[8], Karl Ritz[2], Sacha J. Mooney[2] & John W. Crawford[10]

Achieving food security requires resilient agricultural systems with improved nutrient-use efficiency, optimized water and nutrient storage in soils, and reduced gaseous emissions. Success relies on understanding coupled nitrogen and carbon metabolism in soils, their associated influences on soil structure and the processes controlling nitrogen transformations at scales relevant to microbial activity. Here we show that the influence of organic matter on arable soil nitrogen transformations can be decoded by integrating metagenomic data with soil structural parameters. Our approach provides a mechanistic explanation of why organic matter is effective in reducing nitrous oxide losses while supporting system resilience. The relationship between organic carbon, soil-connected porosity and flow rates at scales relevant to microbes suggests that important increases in nutrient-use efficiency could be achieved at lower organic carbon stocks than currently envisaged.

Soil is the most important terrestrial sink and source of greenhouse gases, particularly the long-lived gases carbon dioxide and nitrous oxide (N_2O) and shorter-lived methane[1]. As N_2O constitutes a loss of an important plant nutrient from the system, nitrogen dynamics in soil are particularly important in linking climate regulation and food security[2]. Improving nutrient-use efficiency such that greenhouse gas emissions resulting from fertilizer production and use falls by 40% per unit of food produced is predicted to be more effective at reducing emissions from the food system than adopting higher-yielding crops or a 50% reduction in food waste[3]. Nutrient-use efficiency and carbon sequestration are coupled closely in soils[4] and co-regulated by microbial metabolism. Microbial metabolism is in turn governed by oxygen availability and the co-location of nitrogen in, and together with, organic matter (OM). The distribution of oxygen is determined by the physical soil pore structure, soil moisture content and the distribution of potential microbial respiration[5]. Therefore, an understanding of how management influences the partitioning of carbon and nitrogen between storage in soil and emissions requires an understanding of the dynamics of both the physical and biotic states of the soil.

[1]Net Zero and Resilient Farming, Rothamsted Research, North Wyke, UK. [2]School of Biosciences, The University of Nottingham, Sutton Bonington, UK. [3]Department of Environmental Sciences, Wageningen University, Wageningen, The Netherlands. [4]Sustainable Soils and Crops, Rothamsted Research, Harpenden, UK. [5]Bristol Veterinary School, University of Bristol, Langford, UK. [6]Programa de Producción y Sustentabilidad Ambiental, Instituto Nacional de Investigación Agropecuaria (INIA), Estación Experimental INIA La Estanzuela, Colonia, Uruguay. [7]School of Integrative Plant Science, Cornell University, Ithaca, NY, USA. [8]Intelligent Data Ecosystems, Rothamsted Research, Harpenden, UK. [9]IBM Research Europe - Daresbury, The Hartree Centre, Warrington, UK. [10]Adam Smith Business School, University of Glasgow, Glasgow, UK. [11]Present address: The Carbon Trust, London, UK. [12]Present address: Genesis, Lisors, France. ✉e-mail: andy.neal@rothamsted.ac.uk

https://doi.org/10.1038/s43016-022-00671-z

Table 1 | Fertility management associated with arable soils of the Broadbalk Winter Wheat Experiment and associated unmanaged soils used in various aspects of this study

Treatment	Annual fertility management	Designation
Arable, composted FYM	35 Mg ha^{-1} since 1843	FYM
Arable, 240 kgN ha^{-1} inorganic fertilizer[a]	240 kgN ha^{-1} since 1985 (144 kgN ha^{-1} between 1852 and 1984)	^{240}NPK
Arable, 192 kgN ha^{-1}, inorganic fertilizer[a]	192 kgN ha^{-1} since 1968	^{192}NPK
Arable, 144 kgN ha^{-1} inorganic fertilizer[a]	144 kgN ha^{-1} since 1852	^{144}NPK
Arable, inorganic fertilizer[a], no phosphorus	No triple superphosphate additions since 1968 (96 kgN ha^{-1} between 1968 and 2001; 192 kgN ha^{-1} since 2001)	^{192}NK
Arable, inorganic fertilizer[a], no nitrogen	No ammonium nitrate additions since 1852	PK
Arable, unfertilized	No fertilizer additions since 1852	Nil
Grassland	None, grassland since 1838	Grassland
Woodland	None, woodland since 1882	Woodland

[a]Unless stated otherwise, inorganic fertilizer included inorganic nitrogen as ammonium nitrate, 90 kgK ha^{-1} as K_2SO_4, 35 kgP ha^{-1} as triple superphosphate ($Ca(H_2PO_4)_2{\cdot}H_2O$) and 12 kgMg ha^{-1} as kieserite ($MgSO_4{\cdot}H_2O$).

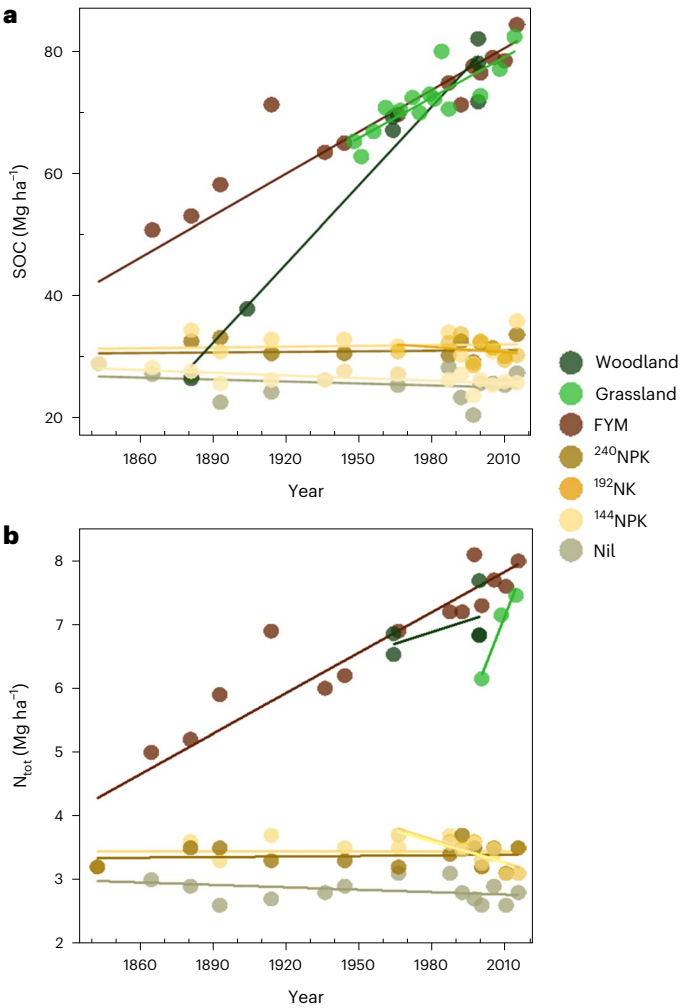

Fig. 1 | SOC and soil N_{tot} stock dynamics in Broadbalk soils. a,**b** Trends in SOC (**a**) and soil N_{tot} (**b**) stocks in unmanaged woodland ($n = 7$) and grassland ($n = 13$) soils; arable soils of the Broadbalk Winter Wheat Experiment receiving FYM ($n = 14$) annually, inorganic fertilizer at rates of 240 kgN ha^{-1} yr^{-1} as ammonium nitrate combined with phosphorus and potassium (^{240}NPK, $n = 6$), 144 kgN ha^{-1} yr^{-1} combined with phosphorus and potassium (^{144}NPK, $n = 12$) or 192 kgN ha^{-1} yr^{-1} with potassium but no phosphorus (^{192}NK, $n = 7$); and soil which has received no fertilization (Nil, $n = 14$). Soil management is summarized in Table 1.

Important soil functions, particularly transport capacity, metabolic efficiency and the sustained delivery of nutrients to plant roots, are dependent on the connectivity and heterogeneity of pore space, emphasizing the importance of a pore-centric view of soil architecture[6]. Soils exhibit spontaneous emergence of multi-scale self-organization driven by endogenous feedback between pore space architecture and microbial metagenetic—rather than taxonomic—states, reflecting an extended composite phenotype[7]. Microbiological activity influences soil pore architecture at scales below approximately 80 µm. As these are the scales that affect the relative balance and distribution of air and water in soil, changes within this range also influence the nature of microbial activity, especially the balance between aerobic and anaerobic metabolism. The resulting process-form state is a consequence of the self-organization of the integrated biophysical system of soil[7,8]. The emergent state is affected by incorporation of OM into soil which drives a tightly coupled feedback involving changes to the pore architecture associated with altered gene assemblages; hence, the extended phenotype[7–10] of the soil system is an irreducible composite reflecting physical and biotic feedbacks.

The soil extended composite phenotype incorporates microbial gene assemblages, soil structural parameters, hierarchical flux processes of gases and water, and overall system function including the likelihood of anoxia. The extended composite phenotype of soils containing low stocks of organic carbon is characterized by greater proportions of anoxic pore space and increased abundance of genes in the metagenome associated with dissimilatory respiration of nitrate and nitrite[7]. These factors effectively link the fates of nitrogen and carbon in soil.

In this Article, we provide a further test of the theory of soil as an extended composite phenotype to provide a mechanistic explanation of the links between carbon and nitrogen metabolism and show how both physical and genetic factors must be accounted for. We hypothesize that arable soils receiving high organic inputs would have a highly connected process-form state resulting in reduced losses of stored nitrogen as N_2O through anaerobic metabolism. Conversely, soils receiving low organic inputs would result in a poorly connected state, reduced oxygen flux and higher gaseous losses of nitrogen. Furthermore, using this framework, we seek to explain the link between carbon dynamics and non-equilibrium nitrogen-use efficiency (NUE), calculated as the balance of nitrogen inputs, off-takes and accumulation in soil. We hypothesize that offsetting losses of N_2O from soil would be associated with higher stocks of soil nitrogen, along with carbon and water storage—key factors in conveying resilience in rainfall-limited production systems. To test our hypotheses, we studied arable soils subjected to consistent management for over 160 years, comparing soils that had received a range of continuous organic or inorganic fertilization over this period.

Results and discussion

Co-metabolism of carbon and nitrogen

Since experiment establishment (1843), soils subject to different management or fertilization (Table 1) have developed distinct soil organic carbon (SOC) and total nitrogen (N_{tot}) stocks (Supplementary Fig. 1). Woodland and proximal grassland soils, together with arable soil receiving composted farmyard manure (FYM), were the only soils to show net SOC and N_{tot} accumulation (Fig. 1). Grassland and FYM-amended

https://doi.org/10.1038/s43016-022-00671-z

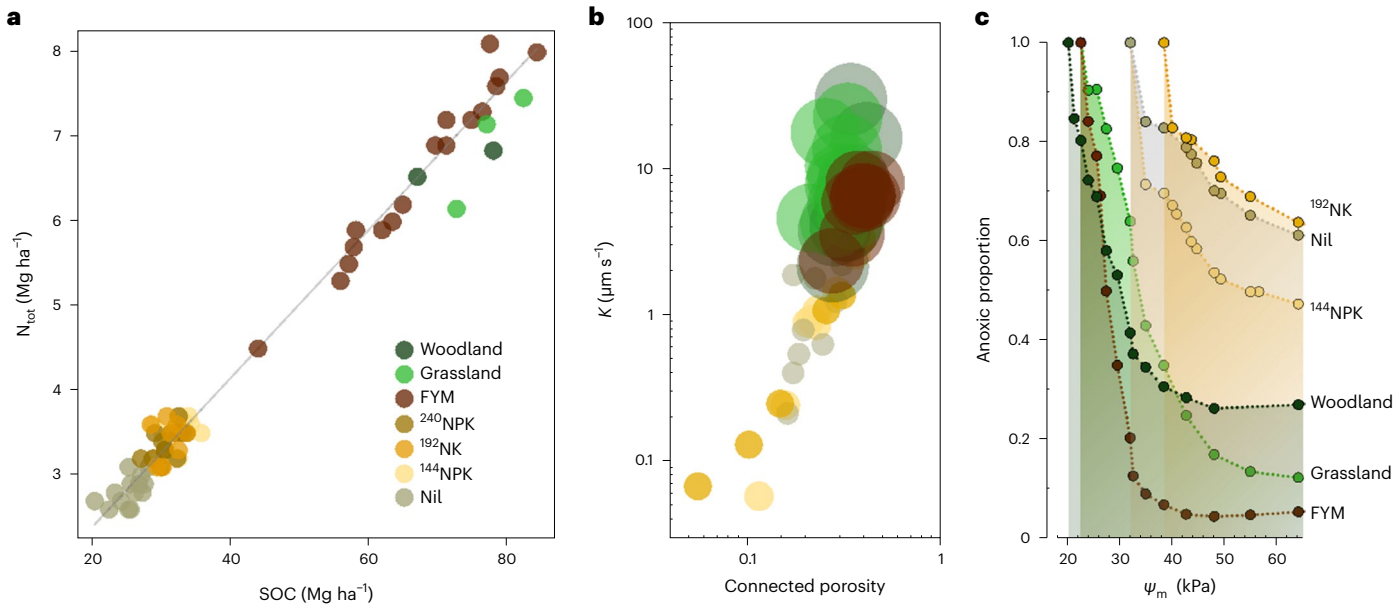

Fig. 2 | Contrasting long-term soil management results in quantitatively different process-form states. a, Geometric mean functional relationship between SOC and N_{tot} spanning the years 1843–2015 measured in FYM-amended, inorganically fertilized (^{240}NPK, ^{144}NPK and ^{192}NK) and unfertilized (Nil) arable soils of the Broadbalk Winter Wheat Experiment, and unmanaged (woodland and grassland) soils. Slope = 0.088 (bootstrapped 95% confidence interval, 0.085–0.091), $t = 60.7$, $P = 6.5 \times 10^{-63}$; $\rho = 0.990$, $t = 60.1$, $P = 1.3 \times 10^{-62}$, $n = 73$. **b**, FYM-amended, ^{144}NPK, ^{192}NK and PK arable soils, and woodland and grassland unmanaged soils are described by a combination of the connectivity of pore space, established from X-ray CT, and simulated saturated hydraulic conductivity (K, in units of μm s^{-1})—a measure of capacity, representing the maximum potential movement of resources through pore networks to organisms. Data point size is proportional to SOC stocks (Mg ha^{-1}) in each soil, shown in Supplementary Fig 1a. **c**, Process-form states control anoxia within the same soils. Low-SOC, low-connected porosity soils (inorganically fertilized and unfertilized arable soils ^{144}NPK, ^{192}NK and Nil) contain large volumes of anoxic microsites. Across a range of matric potential (ψ_m), the predicted volume of anoxic sites is consistently larger in these soils than in high-SOC, high-connected porosity soil (FYM-amended arable soil and grassland and woodland soils). Anoxic pore space was modelled under conditions of relatively low microbial respiration, $k' = 1 \times 10^{-4}$. The key shown in **a** relates to all three panels.

arable soils both contained over 70 Mg ha^{-1} SOC and over 6 Mg ha^{-1} N_{tot}, substantially more than other treatments. There was a positive geometric mean functional relationship between SOC and N_{tot} (Fig. 2), corresponding to a C:N ratio of 11.4. Despite widely varying quantities and qualities of carbon and nitrogen inputs, this indicates coupling via the same metabolic pathways and mechanisms of storage in soil. Metabolic and storage constraints in soil are both linked to microenvironmental conditions, via a strong association between SOC inputs and soil process-form relationships[7]. Therefore, we quantified the micrometre-scale structure of a subset of soils to understand the biophysical feedbacks in the co-metabolism of carbon and nitrogen.

Links between nitrogen storage and physical processes

Parameters relating to soil architecture were determined in grassland, woodland, FYM and inorganically fertilized arable soils receiving 144 kgN ha^{-1} yr^{-1} as ammonium nitrate, phosphorus and potassium (^{144}NPK), 192 kgN ha^{-1} yr^{-1} and potassium but no phosphorus (^{192}NK), and soil receiving no nitrogen fertilization but phosphorus and potassium (PK). Significant treatment-dependent differences were observed for parameters generated directly from X-ray computed tomography (CT) of pore networks in each soil (total [P_t] and connected [P_c] porosity) and those derived from simulation of the permeability (k), normalized effective oxygen diffusion coefficient (D_e') and hydraulic conductivity (K) within the pore networks (Table 2). Collectively, these measures describe the dynamical state of soil pore space and the maximum potential rate at which resources can move through the networks, that is, the capacity for flux. Grassland, woodland and FYM-amended arable soils were typified as having more extensive and more connected pore networks than inorganically fertilized arable soils. We observed a power-law relationship between P_c and K: increased SOC was associated with increases in both parameters (Fig. 2b). Regions of this relationship correspond to the process-form states of the various soils. This infers that—in the case of the soils studied here—for P_c between 0.05 and 0.4, relatively small changes in geometry, characterized by P_c, result in substantial increases in K, characterizing flow rate. This power law therefore has profound implications for soil management strategies.

The effect of soil management was most evident on permeability (k) based on treatment effect size (ω^2; Table 2). We simulated the anoxic proportion of each soil across a range of matric potentials (ψ_m). Addition of FYM to arable soils resulted in a matric potential–anoxic space profile distinct from inorganically fertilized arable soils (Fig. 2c). Inorganically fertilized ^{192}NK and ^{144}NPK soils, as well as soil that had received no fertilization (referred to as Nil, Table 1) all presented large proportions of anoxic space under relatively dry conditions (ψ_m = 50–65 kPa) and were predicted to be completely anoxic at ψ_m between 38.4 kPa and 32.0 kPa. In contrast, the process-form state of FYM-amended soil exhibited aspects of woodland and grassland soils. Under relatively dry conditions, FYM soil had low proportions of anoxic space more typical of grassland soil. At increased moisture content, both soils were completely anoxic at 22.6 kPa, woodland soil being so at 20.2 kPa.

As with an affiliated ley-arable experiment[7], organic-carbon-rich soil developed a distinct process-form state, typified by a greater proportion of connected pores, and increased hydraulic conductivity (Fig. 2b). Modelling predicts this to be a more oxygenated pore network (Fig. 2c) by virtue of its greater capacity to transport O_2. Annual addition of FYM (35 Mg ha^{-1}) to the soil has resulted in SOC and N_{tot} contents and process-form states resembling unmanaged woodland and grassland, despite regular physical disturbance by inversion tillage. This combined evidence is consistent with our hypothesis regarding

Table 2 | Soil pore topology-related parameters

	Total porosity (P_t) (%)Welch ANOVA; $F_{5,20.8}$=14.1, P=5×10^{-6} ω^2=0.467	Connected porosity (P_c) (%)Welch ANOVA; $F_{5,20.7}$=14.2, P=4×10^{-6} ω^2=0.459	Permeability (k) (μm²) Welch ANOVA; $F_{5,21}$=36.5, P=1×10^{-9} ω^2=0.775	Diffusion coefficient (D_e') ANOVA; $F_{5,47}$=19.5, P=2×10^{-10} ω^2=0.636	Hydraulic conductivity (K) (μm s^{-1})Welch ANOVA; $F_{5,21}$=15.8, P=2×10^{-6} ω^2=0.361
Woodland	33.1 (1.3)	32.8 (1.4)	0.396 (0.131)	0.278 (0.020)	8.86 (2.94)
Grassland	31.0 (1.2)	30.8 (1.2)	1.13 (0.089)	0.254 (0.015)	9.61 (1.61)
FYM	37.7 (1.6)	37.5 (1.6)	0.265 (0.028)	0.335 (0.023)	5.92 (0.62)
^{144}NPK	23.6 (2.4)	22.4 (2.7)	0.045 (0.012)	0.135 (0.016)	1.01 (0.27)
^{192}NK	24.9 (4.5)	23.4 (5.1)	0.065 (0.027)	0.151 (0.032)	1.45 (0.60)
PK	22.6 (1.7)	21.6 (1.8)	0.048 (0.011)	0.129 (0.014)	1.07 (0.24)

The topology-related parameters are derived directly from binary images generated from X-ray CT of aggregates from FYM-amended and inorganically fertilized (^{144}NPK, ^{192}NK and PK) arable soils of the Broadbalk Winter Wheat Experiment and unmanaged woodland and grassland (total and connected porosity) and estimates of permeability (k), effective diffusion coefficient of oxygen (D_e'), which is normalized by dividing the diffusion coefficient of oxygen in water without pore constraints, and hydraulic conductivity (K) of the soil pore networks derived from lattice Boltzmann simulation (and where $K=gk/v$, where g is the gravitational constant, k the permeability and v the viscosity of water). The mean associated with each parameter is provided together with the standard error of the mean in parentheses. Results of ANOVA are shown together with estimates of treatment effect size (ω^2) for each parameter, calculated as $(SS_{bg}-df_{bg}\times MS_{wg})/(SS_{total}+MS_{wg})$, where SS indicates the sum of squares, df the degree of freedom and MS the mean square, and the subscripts bg and wg indicate between-group and within-group, respectively. The data summarizes the results of nine replicate measurements of total and connected porosity and simulations of k, D_e' and K for each soil.

the influence of organic carbon in soils in creating a more highly connected and oxygenated structure.

Consequences of soil process-form state for potential biological function

Nitrogen transformation-associated metagenomes were determined in grassland, woodland, FYM, ^{144}NPK, ^{192}NK and PK soils from shotgun metagenomic approaches. There was a significant treatment effect on gene assemblages (permutational multivariate analysis of variance, 99,999 permutations: pseudo-$F_{5,12}$ = 22.5, P_{perm} = 1 × 10^{-5}). Post hoc pairwise comparisons indicated no significant difference in gene assemblages between ^{192}NK and PK soils. All other comparisons were significantly different.

Hierarchical clustering of soils based on relative gene abundance showed that gene assemblages of unmanaged woodland and grassland soils were distinct from those of arable soils (Fig. 3). Genes separated into two broad clusters based on their distribution between unmanaged and arable soils. The first included genes associated with amino acid metabolism and other nitrogen-assimilation pathways, including nitrate assimilation (Kyoto Encyclopedia of Genes and Genomes (KEGG) module M00615) and assimilatory nitrate reduction (M00531). Several genes associated with dissimilatory nitrate reduction (M00530) and denitrification (M00529) were also associated with this cluster. The relative abundance of these genes was greater in woodland and grassland soils—which were most alike—than in arable soils (Fig. 3). More subtly, genes associated with nitrate assimilation were relatively more abundant in FYM-amended than in inorganically fertilized arable soils.

The second gene cluster was generally more abundant in arable soils. This comprised genes associated with several modules, including nitrification (M00528), denitrification (M00529), dissimilatory nitrate reduction (M00530) and complete nitrification (M00804), (compare with Supplementary Information). This pattern of limited genotypic differences between arable soils is broadly consistent with known responses of microbial communities to nitrogen fertilization[11–14]. This is evidence for direct influence of process-form states upon nitrogen-associated gene assemblages in soil, particularly the association of nitrogen assimilatory pathways with soils having the greatest SOC and N_{tot}, connected porosity and hydrodynamic conductivity. The greater relative abundance of genes associated with nitrogen assimilation in FYM-amended soil is consistent with this soil having a similar process-form state (Fig. 2b) and oxygen status (Fig. 2c) to woodland and grassland soils. Soils having low SOC (and N_{tot}) and thus low connected porosity and hydrodynamic conductivity were associated more with dissimilatory pathways, utilizing oxidized forms of nitrogen as alternative electron acceptors for respiration. Within arable plots particularly, there was a subtle shift in genes associated with denitrification and dissimilatory nitrate reduction: NO-forming and cytochrome *c*-type nitrate reductases (*nirK* and *napAB*, respectively) were relatively more abundant in FYM-amended than in inorganically fertilized soils. These genes are typically associated with more oxygenated environments than the functionally equivalent but structurally dissimilar genes *nirS* (*c*-type cytochrome) and the nitrite oxidoreductase *nxrAB*[15–18] and were most abundant in grassland and woodland soils, respectively. In addition, FYM had the highest relative abundance of genes of the high-affinity nitrate-binding transporter (*nrtABC*) of all arable soils. These genes were otherwise most abundant in grassland soils. Nitrite reductases typical of more reducing environments (*nirS* and *nxrAB*) were indicative of arable soils in general, particularly ^{144}NPK and ^{192}NK treatments. In FYM-amended soil, the more oxygen-rich process-form state (Fig. 2c) influences the gene assemblage such that it shares similarities to those of grassland and woodland soils.

Assessment of gene abundance associated with nitrogen transformations by quantitative PCR provides evidence of the pattern and rate of nitrogen cycling processes in soil[19]. However, the relative abundance of different genes discussed above (and in Supplementary Information) generated from metagenomics cannot in themselves be used as the basis for estimates of process rates[20]. We assumed that relative gene abundance reflects the likelihood of a process with which they are associated when environmental conditions allow. To test this assumption, we measured N_2O emissions in the field, relating them to relative gene abundance and predictions of anoxic pore space (Fig. 2c).

N_2O emissions from soil

As lower levels of organic carbon are associated with less extensively developed soil pore networks[21,22] (Table 2) and the resultant low diffusivity restricts oxygen permeability[7,23–26] (Fig. 2c), elevated N_2O emissions are often associated with poor soil architecture. We placed static chambers in the field on FYM, ^{240}NPK, ^{192}NK, ^{144}NPK and PK soils between April and November 2019 to compare soil N_2O emissions. For each measurement date, the greatest emissions were typically measured from ^{240}NPK soils and the least from PK soils (Supplementary Fig. 2). We detected no significant influence of soil temperature (analysis of covariance; $F_{1,134}$ = 0.901, P = 0.344) or potential soil moisture deficit (analysis of covariance; $F_{1,134}$ = 0.272, P = 0.603) upon N_2O emissions. Comparing soils directly, there was a significant influence of arable soil management (analysis of variance ; $F_{4,116}$ = 8.7, P = 3 × 10^{-6}) on soil mean daily N_2O emission (Fig. 4). FYM-amended soil emitted significantly less N_2O than ^{240}NPK soil (t = 3.1, P = 0.005), despite receiving similar annual

https://doi.org/10.1038/s43016-022-00671-z

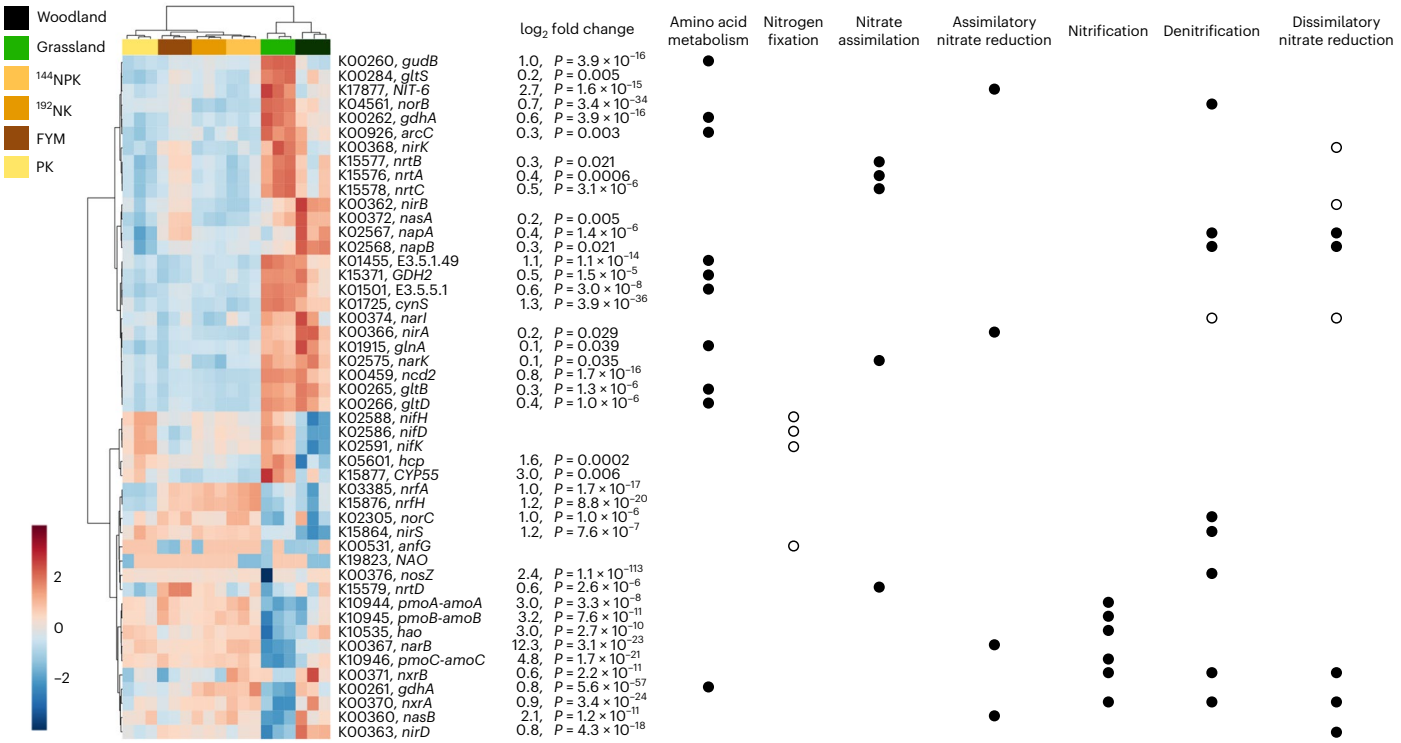

Fig. 3 | Nitrogen metabolism-associated gene assemblages in triplicate pseudo-replicate soil metagenomes from FYM-amended, inorganically fertilized (^{144}NPK, ^{192}NK and PK) arable soils of the Broadbalk Winter Wheat Experiment and unmanaged (woodland and grassland) soils, determined from shotgun metagenomics. Left: heat map representation and bi-hierarchical clustering of gene orthologues (vertical clustering) and soil management (horizontal clustering) based on centred-log ratio scaled orthologue relative abundance. Genes are identified by their KEGG orthologue (K) number and gene nomenclature. The largest $\log_2$ fold difference between soils is shown together with a measure of significance of the difference determined using a two-tailed Wald's test and Benjamini–Hochberg false discovery rate correction (*P*). Right: association of individual genes with nitrogen metabolism-related functional units of gene sets associated with metabolic pathways (modules). Closed circles indicate genes for which a significant difference in abundance between soils was identified; open circles indicate genes for which no significant difference was observed. The KEGG ontology does not discriminate between the genes *amoABC* associated with ammonia oxidation and *pmoABC* associated with methane oxidation.

nitrogen inputs and having more than double the nitrogen stocks of inorganically fertilized soils (Supplementary Fig. 1).

Non-equilibrium nitrogen use efficiency of the soil–plant system

We hypothesized that differences in extended composite phenotype of FYM-amended and inorganically fertilized arable soils would be reflected in historical nitrogen fluxes and stocks through the soil–plant systems. Traditionally, NUE is expressed as the ratio between the amount of nitrogen introduced to and harvested from a system within a single season[27–29]. This implicitly assumes a system is operating under a long-term equilibrium, where soil nitrogen stocks remain temporally constant. For the inorganically fertilized plots, this condition is largely satisfied due to the long trial history meaning annual production of soil organic nitrogen virtually equals annual nitrogen mineralization[30,31]. However, when this condition is not satisfied—FYM-amended soils continue to accumulate nitrogen (Fig. 1b)—applied nitrogen stored in soil beyond a single cropping cycle is incorrectly considered 'wasted'. Although nitrogen that accumulates in soil is not recovered by the immediate crop, it may contribute to system resilience by supporting production of subsequent crops. To account for this, we calculated non-equilibrium NUE for treatments, apportioning nitrogen introduced to each system between that taken up by the crop, stored in soil and 'lost' from the soil–plant system. For lost nitrogen, gaseous losses resulting from denitrification may be up to double those via leaching[32]. As ^{240}NPK and FYM receive similar levels of nitrogen inputs (but differ in levels of readily available nitrogen), the effect of organic carbon input can be evaluated. Mean annual nitrogen fluxes through the three pools following addition to the different winter wheat crop systems (2000–2015) show progressively greater losses of nitrogen inputs with increasing fertilization (Fig. 5a). While N_{tot} associated with wheat grain and straw increases as nitrogen inputs increase, the proportion of total assimilated nitrogen reduces (Supplementary Fig. 3), indicating progressively reduced long-term NUE. Nitrogen inputs required to generate a 1 Mg grain harvest show similar trends (Fig. 5b). Set alongside inorganically fertilized systems, and particularly ^{240}NPK, the FYM system shows altered allocation of added nitrogen to each pool. Reduced total (Fig. 5a) and proportional (Supplementary Fig. 3) allocation of nitrogen to lost pools and increased nitrogen stocks (compare with Figs. 1 and 5a) suggest greater NUE. This increased efficiency is tempered by reduced nitrogen allocation to crops and grain yields (Fig. 5). The 2000–2015 average grain yield from FYM-amended soil was 1.1 Mg ha^{-1} less than from ^{240}NPK fertilized soil (Tukey-Kramer Studentized Q = 4.2; $P = 0.035$). This is consistent with lower area-scaled (but greater yield-scaled) N_2O emissions from soils receiving organic amendments rather than inorganic fertilizer[33]. Measured and simulated micropore-scale behaviour of soil extended composite phenotypes reflect our field-scale description of non-equilibrium NUE, particularly reductions in the absolute and proportional size of the lost nitrogen pool. Our data suggest that the application of FYM to soil is directly related to the reduction of nitrogen losses as we observe a 1.5-fold increase in soil P_t and a 4-fold increase in k in FYM-amended over inorganically fertilized soils.

The extended composite phenotype and system resilience

Consistent with our hypothesis that nitrogen balance in soils would be altered under different emergent process-form states, we observe that the reduced area-scaled N_2O emissions in FYM-amended soils associated

https://doi.org/10.1038/s43016-022-00671-z

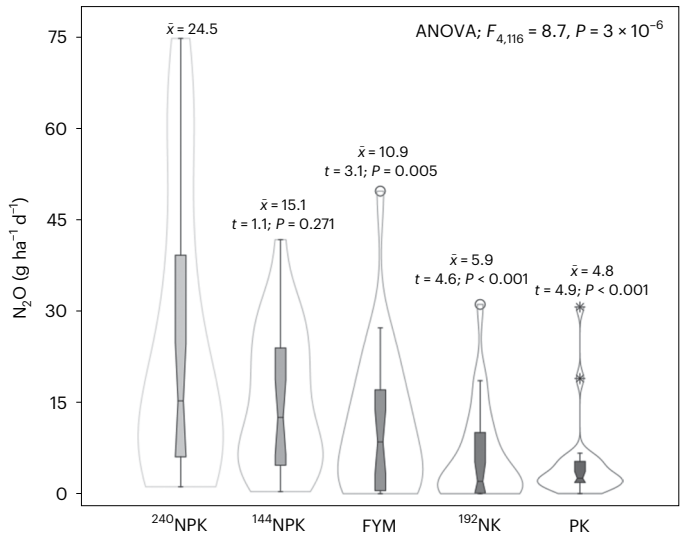

Fig. 4 | Kernel density estimation of the distribution of N_2O emissions from FYM-amended (n = 25) and inorganically fertilized (^{240}NPK, n = 30; ^{144}NPK, n = 24; ^{192}NK, n = 25; and PK, n = 23) soils of the Broadbalk Winter Wheat Experiment measured between April and November 2019. Box plots represent the interpolated 25% and 75% quartiles, and the median is represented as a horizontal line. Notches represent approximate 95% confidence intervals of the medians. Whiskers indicate the extent of the distribution of data 1.5 times the interquartile range. Measurements of N_2O emissions falling outside 1.5 times the interquartile range (outliers, determined based on generalized extreme Studentized deviate tests) are represented by open circles; those falling outside 3 times the interquartile range are represented by an asterisk. The results of a one-factor analysis of variance is shown, together with the adjusted mean ($\bar{x}$) for each treatment; t (two-tailed) and probabilities (P) indicate the results of Holm–Šidák a priori contrasts of FYM, ^{144}NPK, ^{192}NK and PK mean N_2O emissions to ^{240}NPK emissions.

with high connected porosity and predicted greater oxygen availability are associated with greater N_{tot} stocks over the long term. By combining physical structure with metagenetic and phenotypic description, we invoke a more fundamental and mechanistic conceptualization of the role of OM in soil function. We demonstrate that the nitrogen-associated metagenome responds to nitrogen fertilization markedly, but that the form of nitrogen input exerts relatively little influence on redox-sensitive *nirK-napAB* and *nirS-nxrAB* gene abundances. Given the magnitude of the changes we observe in the process-form state, it is likely that OM controls N_2O emissions by influencing gene expression (phenotype) rather than altering gene abundance (genotype). In effect, reduced N_2O losses are a co-benefit of increased OM inputs. This regulation results in nitrogen accumulation in soil conferring a degree of resilience, consistent with previously observed reductions in the need for inorganic nitrogen application to achieve a target yield in organic carbon-rich soils[34]. Evidence for higher SOC stocks being associated with higher crop yields is equivocal[35,36]. However, the resilience of rainfed crop yields to drought does appear to be related to higher SOC stocks[37,38], although in the soils studied here such effects are observed only where nitrogen availability limits crop yield[39].

Conclusions

Our adoption of a nonlinear and systems-level approach, founded on quantification of soil pore architecture and interactions with the metagenome, shows that the behaviour of the soil system cannot be understood by studying the behaviour of the physical and biotic components in isolation. The form of nitrogen fertilization in arable systems does little to influence nitrogen-metabolism-associated gene assemblages, but control over expression is exerted by the emergent process-form state which is dependent upon OM inputs. While the applications of FYM used in our experimental system are impractical[40], the power-law relationship observed in Fig. 2 implies that similar levels of efficiency and resilience could be achieved at lower SOC input and stocks than those used here. Our data suggest that optimally efficient systems can achieve both increased SOC stocks and reduced N_2O emissions.

Methods

Field experiment and soil sampling

Soil used to generate physical data relating to soil structure and biological data derived from metagenomics was sampled from arable soils which had received inorganic fertilizer inputs ranging from no inputs to different combinations of inorganic nitrogen (ammonium nitrate), phosphorus and potassium, comparing rates of nitrogen inputs from 0 kg ha^{-1} yr^{-1} to 240 kg ha^{-1} yr^{-1} from contrasting treatments of the Broadbalk Long-Term field experiment established in 1843 (51° 48′ 35″ N, 00° 22′ 30″ W). The experiment tests the long-term consequences of different fertilizer and manure applications on the yield of winter wheat. Detailed description of the treatments compared in this study are provided in Supplementary Information: comprehensive details regarding the experiment history are provided by ref. 41. We also studied two soils which had not been subject to management for over a century: part of the original Broadbalk experiment which since 1882 has been allowed to revert naturally to a woodland (referred to as the woodland treatment) and plots of the Highfield Ley-Arable experiment managed as a mown sward[41] (referred to as the grassland treatment). Details of the management associated with the different soils compared in his work are provided in Table 1. Treatments are not replicated across the Broadbalk experiment. Therefore, samples associated with metagenomics and X-ray CT (both described below) are by necessity pseudo-replicates. Assessment of the differences between treatments as a consequence are relative to the pooled within-plot variability in both univariate analysis of variance and permutational multivariate analysis of variance tests, rather than some form of between-plot variability, and we cannot statistically separate spatial from 'treatment' effects. Our approach therefore tests the hypothesis that there are differences in measured and simulated soil parameters, and assemblages and relative abundance of nitrogen cycle-associated genes in the different plots at the Broadbalk site. We cannot conclude, based on statistical inference, that any differences observed result from different fertility management directly. However, it is reasonable to interpret our observations of between-plot differences in soil structure and relative gene abundance within the context of the broader data relating to SOC and N_{tot} stocks, held by the *e*-RA, a permanent managed database for secure storage of data from Rothamsted's Long-term Experiments, and N_2O emissions—all of which are temporally replicated—to suggest explanations for the patterns we observe.

Soil structure and hydrodynamic behaviour

X-ray CT and image analysis. We generated X-ray CT images from grassland, woodland, FYM, ^{144}NPK, ^{192}NK and Nil treatments at 1.5 µm resolution and scales relevant to microbes (10^0–10^2 µm), requiring imaging of 0.7-mm- to 2.0-mm-diameter soil aggregates. Aggregates (n = 9) were selected at random from soil collected from each plot. Each was scanned using a Phoenix Nanotom system (GE Measurement and Control Solution) operated at 90 kV, a current of 65 µA and a voxel resolution of 1.5 µm. Initial image analysis was performed using ImageJ version 1.51j8. Images were threshold-adjusted using a bi-level bin approach[42] using QuantIm version 4.01. Porosity was calculated directly from threshold-adjusted binary images, and pore connectivity was determined according to ref. 42.

Pore network permeability, hydraulic conductivity and oxygen diffusion in soil. The permeability (k) and hydraulic conductivity (K) of soil pore networks and the effective diffusion coefficient (D_e) for oxygen within the pores was calculated for each soil structure imaged

https://doi.org/10.1038/s43016-022-00671-z

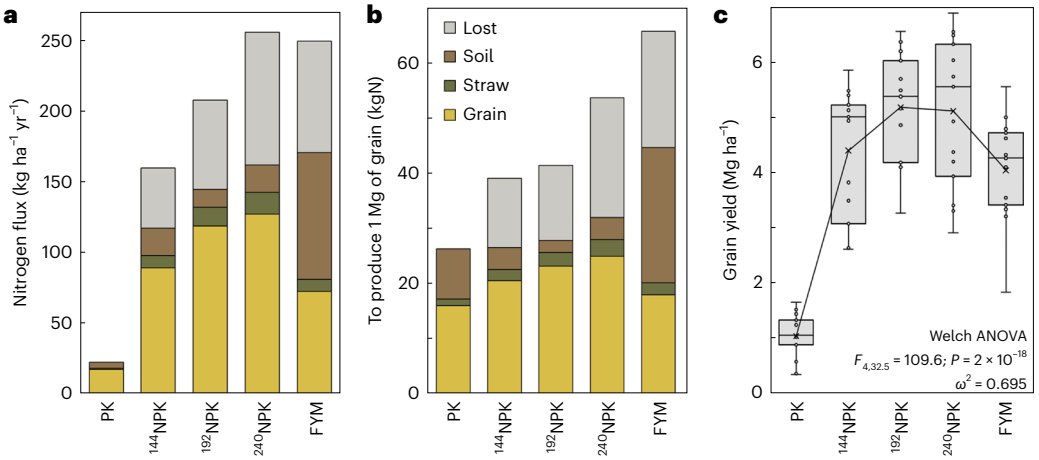

Fig. 5 | Long term non-equilibrium nitrogen use efficiency of FYM-amended and inorganically fertilized soils. Non-equilibrium nitrogen use efficiency of FYM-amended (n = 11) and inorganically fertilized (^{240}NPK, n = 12; ^{192}NPK, n = 10; ^{144}NPK, n = 10; and PK, n = 9) soils of the Broadbalk Winter Wheat Experiment between 2000 and 2015 indicates divergent allocation of **a**, absolute nitrogen inputs to arable systems and **b**, nitrogen inputs supporting 1 Mg of grain production, between grain, straw, soil and lost pools. Small increases in nitrogen soil stocks observed for inorganic plots could be due to measurement error, as these systems are assumed to have reached steady states. **c**, Grain yields for inorganically fertilized and FYM-amended soils for the same period. Box plots show interpolated 25% and 75% quartiles, and the median is represented as a horizontal line. Whiskers indicate minimum and maximum values. Each treatment mean is represented by an X, and the line connects treatment means. Results of analysis of variance are shown together with estimates of treatment effect size (ω^2) on yield, calculated as $(SS_{bg} - df_{bg} \times MS_{wg})/(SS_{total} + MS_{wg})$, where SS indicates the sum of squares, df the degree of freedom and MS the mean square, and the subscripts bg and wg indicate between-group and within-group, respectively. Two-tailed, post hoc Tukey–Kramer pairwise comparisons adopting the step-down, recursive reject Copenhaver–Holland multiple comparison procedure indicate that the mean PK grain yield was significantly lower than the means of all other treatments (smallest difference, Tukey-Kramer Studentized Q = 11.7, P = 2 × 10^{-12}). FYM mean grain yield was significantly lower than the mean yields of both the ^{192}NPK and ^{240}NPK treatments (smallest difference, Q = 4.2, P = 0.033). There was no significant difference between the mean grain yields of FYM and ^{144}NPK treatments.

by X-ray CT using lattice Boltzmann simulation[43,44]. Permeability and hydraulic conductivity were calculated as described by ref. 25. For oxygen diffusion, hierarchical soil structures revealed by X-ray CT images indicate that gaseous oxygen in the atmosphere moves into soil primarily through its inter-aggregate pores and is then dissolved in water before moving into aggregates, largely by molecular diffusion. As gaseous oxygen diffuses up to 10^3-fold more quickly than oxygen dissolved in water, microbial community activity is constrained mainly by oxygen diffusion within aggregates. The ability of aggregates to conduct dissolved oxygen and other soluble substrates depends on the intra-aggregate pore geometry, and we quantified it with effective diffusion coefficients calculated directly by mimicking solute movement through the pore geometry using numerical simulations. The movement of solutes, including oxygen, within the pore geometry is assumed to be diffusion dominated. To account for the effect of temperature on diffusion, we calculated a normalized effective diffusion coefficient D_e' dividing the effective diffusion coefficient by the molecular diffusion coefficient of oxygen, D, in water at the same temperature, that is, $D_e' = D_e/D$ (refs. 25,26). Detailed description of pore-scale simulation is provided in Supplementary Information.

Modelling anoxia within soils. Having established D_e' for oxygen in the soil pore networks of each soil, we simulated oxygen consumption by microbes under various levels of water saturation[7]. The first step of simulation was to determine water distributions in the pore networks under different matric potentials (ψ_m). Once water distributions were determined for a given ψ_m, we simulated oxygen dissolution at air–water interfaces and then diffusion towards solid–water interfaces where it was reduced by microbial respiration. Microbial consumption was assumed to occur in water-filled voxels adjacent to the water–solid wall and described by a Monod kinetic equation. Oxygen diffusion and reduction was simulated to steady state. As the development of anaerobic volume is a balance between oxygen diffusion within the pore network and microbial consumption, we simulated two scenarios: fast ($k' = 1 \times 10^{-2}$) and slow ($k' = 1 \times 10^{-4}$) microbial consumption of oxygen to determine whether the relative anaerobicity of soils under the same ψ_m was a consequence of their structures and not microbial respiration rate. Once each system had reached a steady state, we considered sites where the concentration of dimensionless dissolved oxygen was less than 20% to be anaerobic[45]. We repeated the procedure to achieve different water distributions calculated by varying ψ_m and then calculated the proportional change in the volumetric anaerobic sites with the ψ_m for both the fast and slow microbial reactions. Detailed description of these steps is provided in Supplementary Information.

DNA extraction, sequencing and quality control. Three pseudo-replicate samples of soils from the grassland, woodland, FYM, ^{144}NPK, ^{192}NK and PK treatments were sampled in October 2015. Soil community DNA was extracted from a minimum of 2 g of thawed soil using MoBio PowerSoil DNA isolation kits (Mo Bio Laboratories). About 10 µg of high-quality DNA was provided for sequencing for each of the samples. Shotgun metagenomic sequencing of DNA was performed using 150-base paired-end chemistry on an Illumina HiSeq 2500 sequencing platform by Beijing Novogene Bioinformatics Technology. The generated raw sequences were limited to a minimum quality score of 25 and a minimum read length of 70 bases using Trimmomatic[46].

Bioinformatic analysis of metagenome sequences. To assess the general abundance of genes in metagenomes, we mapped individual metagenomic sequences to the RefSeq non-redundant protein database held at the National Center for Biotechnology Information (downloaded on 22 August 2018) using DIAMOND version 0.8.27 (ref. 47) in BLASTX mode imposing a bitscore cut-off of 55. For each sequence, only the match with the highest bitscore was considered. Sequences not matching the non-redundant database were considered currently unclassified. MEGAN Ultimate version 6.10.2[48] was used to associate metagenome sequences with KEGG[49] functional orthologues (ko) and modules (M). From all the reads binned to at

least one KEGG orthologous group, we selected those associated with nitrogen metabolism (ko00910), amino acid metabolism (ko09105) and metabolism of other amino acids (ko09106) for detailed study of distribution differences between soils. N_2O is a product of autotrophic and heterotrophic nitrification pathways as well as denitrification. Emission rates differ markedly, both between processes and at different soil saturations[50]. Emission rates of N_2O due to denitrification are far greater than from either nitrification pathway[50]. Not only is nitrification a minor source of N_2O, but also arable systems exhibit the lowest maximum potential nitrification-derived N_2O emissions across a broad range of ecosystems, and this declines with increasing management intensity[51]. For these reasons, we assumed that N_2O emissions were predominantly driven by the influence of oxygen on anaerobic respiration via the denitrification pathway. For each module in the KEGG ontology, genes may be associated with several sub-modules. These are not formed of exclusive sets of genes and so are likely to reflect a range of function and ecophysiologies. In addition, the ontology does not distinguish between *amoABC* associated with nitrification and *pmoABC* associated with methane oxidation. Details of individual genes included in the analysis are provided in Supplementary Information, and the distribution of genes across different modules is shown in Fig. 3.

Emissions of N_2O from soil. All nitrogen fertilizers were applied on 12 April 2019, and FYM was applied on 17 September 2018 and 23 September 2019. Measurement of gaseous emissions from ^{240}NPK, FYM and PK soils were taken on 11 dates between 11 April and 7 October 2019. Measurements from ^{144}NPK and ^{192}NK soils were taken on 8 dates between 23 April and 7 October 2019. Gas sampling was performed using in-field static chambers[52]. Three chambers (dimensions 40 cm × 40 cm × 25 cm height) were inserted to a depth of 5 cm in the soil of each treatment. Details of sampling times and frequency are provided in Supplementary Information. N_2O concentrations in gas samples collected in the field were analysed on a PerkinElmer Clarus 500 Gas Chromatograph equipped with a TurboMatrix 110 automated headspace sampler, fitted with an electron capture detector set at 300 °C for N_2O analysis. The static chamber approach used here is prone to errors derived from gas chromatograph instrumental noise and temperature and pressure changes within the chamber leading to artefactual negative fluxes of N_2O (ref. 53). To avoid these artefacts, only non-zero flux estimates were used to analyse differences in N_2O emissions between treatments: negative fluxes were removed from the analysis. We summed total N_2O emissions and tested for differences in the mean N_2O emissions between treatments for the sampling period assessed on an area basis per day.

Non-equilibrium nitrogen use efficiency of the soil–plant system. The balance of nitrogen inputs, off-takes in wheat grain and straw, and accumulation in arable soils was estimated for FYM, ^{240}NPK, ^{192}NPK, ^{144}NPK and PK Broadbalk fertilizer treatments between 2000 and 2015. Historical data relating to nitrogen inputs (as ammonium nitrate fertilizer or FYM, within seed grain and atmospheric deposition), off-takes (as harvested grain and straw) and soil nitrogen stocks were acquired from the *e*-RA managed database of data from Rothamsted's long-term experiments. The end-of-season fate of annual nitrogen additions in the form of inorganic fertilizer or FYM introduced to the Broadbalk continuous wheat plots was separated into three pools: (1) that taken up for the current season's production (that is, nitrogen content in straw and harvested grain), (2) the pool held within the system for potential use in the future (that is, change in soil nitrogen stock) and (3) that lost from the system without being utilized (for example, as N_2O emitted to the atmosphere or nitrate leached to groundwater). The agronomic performance of each plot was then evaluated by non-equilibrium nitrogen use efficiency, here defined as the ratio between nitrogen input and (1) + (2) above.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data relating to the Broadbalk and Highfield long-term experiments can be accessed via the electronic Rothamsted Archive (http://www.era.rothamsted.ac.uk/experiment/rbk1 and http://www.era.rothamsted.ac.uk/experiment/rrn1, respectively). Historical data relating to nitrogen inputs (as ammonium nitrate fertilizer or FYM, within seed grain and atmospheric deposition), off-takes (as harvested grain and straw) and soil nitrogen stocks are available at https://doi.org/10.23637/rbk1-yldS10115-01. All soil images are available upon reasonable request from the corresponding author. Sequence data associated with this research have been deposited in the European Nucleotide Archive with the accession number PRJEB43407. X-ray computed tomography was performed at the Hounsfield Facility of the University of Nottingham where the resulting images are stored. Due to the large number and size of the files, original greyscale and converted binary images are available upon reasonable request from Sacha Mooney, University of Nottingham (sacha.mooney@nottingham.ac.uk). Source data are provided with this paper.

Code availability

C++code for lattice Boltzmann simulation of the dissolution and diffusion of oxygen in 3D soil structures under different soil water contents is available at https://doi.org/10.6084/m9.figshare.21493704.

References

1. Laborde, D. et al. Agricultural subsidies and global greenhouse gas emissions. *Nat. Commun.* **12**, 2601 (2021).
2. Bouwman, A. F., Boumans, L. J. M. & Batjes, N. H. Emissions of N_2O and NO from fertilized fields: summary of available measurement data. *Glob. Biogeochem. Cycles* **16**, 1058–1070 (2002).
3. Clark, M. A. et al. Global food system emissions could preclude achieving the 1.5° and 2°C climate change targets. *Science* **370**, 705–708 (2020).
4. Guenet, B. et al. Can N_2O emissions offset the benefits from soil organic carbon storage? *Glob. Change Biol.* **27**, 237–256 (2020).
5. Rappoldt, C. & Crawford, J. W. The distribution of anoxic volume in a fractal model of soil. *Geoderma* **88**, 329–347 (1999).
6. Letey, J. The study of soil structure—science or art. *Aust. J. Soil Res.* **29**, 699–707 (1991).
7. Neal, A. L. et al. Soil as an extended composite phenotype of the microbial metagenome. *Sci. Rep.* **10**, 10649 (2020).
8. Phillips, J. D. Soils as extended composite phenotypes. *Geoderma* **149**, 142–151 (2009).
9. Dawkins, R. *The Extended Phenotype. The Gene as the Unit of Selection* (Oxford Univ. Press, 1982).
10. Dawkins, R. Extended phenotype—but not too extended. A reply to Laland, Turner and Jablonka. *Biol. Philos.* **19**, 377–396 (2004).
11. Ramirez, K. S. et al. Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems. *Ecology* **91**, 3463–3470 (2010).
12. Fierer, N. et al. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J.* **6**, 1007–1017 (2012).
13. Geisseler, D. & Scow, K. M. Long-term effects of mineral fertilizers on soil microorganisms: a review. *Soil Biol. Biochem.* **75**, 54–63 (2014).
14. Ouyang, Y. et al. Effect of nitrogen fertilization on the abundance of nitrogen cycling genes in agricultural soils: a meta-analysis of field studies. *Soil Biol. Biochem.* **127**, 71–78 (2018).
15. Knapp, C. et al. Spatial heterogeneity of denitrification genes in a highly homogeneous urban stream. *Environ. Sci. Technol.* **43**, 4273–4279 (2009).

16. Graham, D. W. et al. Correlations between in situ denitrification activity and *nir*-gene abundances in pristine and impacted prairie streams. *Environ. Pollut.* **158**, 3225–3229 (2010).
17. Tatariw, C. et al. Denitrification in a large river: consideration of geomorphic controls on microbial activity and community structure. *Ecology* **94**, 2249–2262 (2013).
18. Marchant, H. et al. Denitrifying community in coastal sediments performs aerobic and anaerobic respiration simultaneously. *ISME J.* **11**, 1799–1812 (2017).
19. Petersen, D. G. et al. Abundance of microbial genes associated with nitrogen cycling as indices of biogeochemical process rates across a vegetation gradient in Alaska. *Environ. Microbiol.* **14**, 993–1008 (2012).
20. Rocca, J. et al. Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. *ISME J.* **9**, 1693–1699 (2015).
21. Schjønning, P. et al. Modelling soil pore characteristics from measurements of air exchange: the long-term effects of fertilization and crop rotation. *Eur. J. Soil Sci.* **53**, 331–339 (2002).
22. Bacq-Labreuil, A. et al. Effects of cropping systems upon the three-dimensional architecture of soil systems are highly contingent upon texture. *Geoderma* **332**, 73–83 (2018).
23. Arah, J. R. M. & Ball, B. C. A functional model of soil porosity used to interpret measurements of gas diffusion. *Eur. J. Soil Sci.* **45**, 135–144 (1994).
24. Ball, B. C. Soil structure and greenhouse gas emissions: a synthesis of 20 years of experimentation. *Eur. J. Soil Sci.* **64**, 357–373 (2013).
25. Zhang, X. et al. Relationship between soil carbon sequestration and the ability of soil aggregates to transport dissolved oxygen. *Geoderma* **403**, 115370 (2021).
26. Zhang, X. et al. The effects of long-term fertilizations on soil hydraulic properties vary with scales. *J. Hydrol.* **593**, 125890 (2021).
27. Raun, W. R. & Johnson, G. V. Improving nitrogen use efficiency for cereal production. *Agron. J.* **91**, 357–363 (1999).
28. Erisman, J. W. et al. An integrated approach to a nitrogen use efficiency (NUE) indicator for the food production–consumption chain. *Sustainability* **10**, 925 (2018).
29. Cárdenas, L. M. et al. Nitrogen use efficiency and nitrous oxide emissions from five UK fertilised grasslands. *Sci. Total Environ.* **661**, 696–710 (2019).
30. Jenkinson, D. S. *The Nitrogen Economy of the Broadbalk Experiments. I. Nitrogen Balance in the Experiments. Rothamsted Experimental Station Report for 1976, Part 2*. 103–109 (1977). Rothamsted Research: Harpenden.
31. Powlson, D. S. et al. The nitrogen cycle in the Broadbalk Wheat Experiment: recovery and losses of ^{15}N-labelled fertilizer applied in spring and inputs of nitrogen from the atmosphere. *J. Agric. Sci.* **107**, 591–609 (1986).
32. Addiscott, T. & Powlson, D. Partitioning losses of nitrogen fertilizer between leaching and denitrification. *J. Agric. Sci.* **118**, 101–107 (1992).
33. Skinner, C. et al. Greenhouse gas fluxes from agricultural soils under organic and non-organic management—a global meta-analysis. *Sci. Total Environ.* **468–469**, 553–563 (2014).
34. Körschens, M. et al. Effect of mineral and organic fertilization on crop yield, nitrogen uptake, carbon and nitrogen balances, as well as soil organic carbon content and dynamics: results from 20 European long-term field experiments of the twenty-first century. *Arch. Agron. Soil Sci.* **59**, 1017–1040 (2013).
35. Oelofse, M. et al. Do soil organic carbon levels affect potential yields and nitrogen use efficiency? An analysis of winter wheat and spring barley field trials. *Eur. J. Agron.* **66**, 62–73 (2015).
36. Hijbeek, R. et al. Do organic inputs matter—a meta-analysis of additional yield effects for arable crops in Europe. *Plant Soil* **411**, 293–303 (2017).
37. Huang, J., Hartemink, A. E. & Kucharik, C. J. Soil-dependent response of US crop yields to climate variability and depth to groundwater. *Agric. Syst.* **190**, 103085 (2021).
38. Kane, D. A. et al. Soil organic matter protects US maize yields and lowers insurance payouts under drought. *Environ. Res. Lett.* **16**, 044018 (2021).
39. Macholdt, J. et al. The effects of cropping sequence, fertilization and straw management on the yield stability of winter wheat (1986–2017) in the Broadbalk Wheat Experiment, Rothamsted, UK. *J. Agric. Sci.* **158**, 65–79 (2020).
40. Guenet, B. et al. Can N_2O emissions offset the benefits from soil organic carbon storage? *Glob. Change Biol.* **27**, 237–256 (2020).
41. Macdonald, A. et al. *Guide to the Classical and Other Long-Term Experiments, Datasets and Sample Archive* (Rothamsted Research, 2018).
42. Vogel, H.-J., Weller, U. & Schlüter, S. Quantification of soil structure based upon Minkowski functions. *Comput. Geosci.* **36**, 1236–1245 (2010).
43. Zhang, X. et al. A multi-scale lattice Boltzmann model for simulating solute transport in 3D X-ray micro-tomography images of aggregated porous materials. *J. Hydrol.* **541**, 1020–1029 (2016).
44. Li, Z. et al. Direct methods to calculate the mass exchange between solutes inside and outside aggregates in macroscopic model for solute transport in aggregated soil. *Geoderma* **320**, 126–135 (2018).
45. Harrison, D. E. F. & Pirt, S. J. The influence of dissolved oxygen concentration on the respiration and glucose metabolism of *Klebsiella aerogenes* during growth. *J. Gen. Microbiol.* **46**, 193–211 (1967).
46. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
47. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
48. Huson, D. H. et al. MEGAN Community Edition—interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **12**, e1004957 (2016).
49. Kanehisa, M. et al. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2015).
50. Bateman, E. J. & Baggs, E. M. Contributions of nitrification and denitrification to N_2O emissions from soils at different water-filled pore space. *Biol. Fertil. Soils* **41**, 379–388 (2005).
51. Liang, D. & Robertson, G. P. Nitrification is a minor source of nitrous oxide (N_2O) in an agricultural landscape and declines with increasing management intensity. *Glob. Change Biol.* **7**, 5599–5613 (2021).
52. Chadwick, D. R. et al. Optimizing chamber methods for measuring nitrous oxide emissions from plot-based agricultural experiments. *Eur. J. Soil Sci.* **65**, 295–307 (2014).
53. Cowan, N. J. et al. Investigating uptake of N_2O in agricultural soils using a high-precision dynamic chamber method. *Atmos. Meas. Tech.* **7**, 4455–4462 (2014).

Acknowledgements

This research was supported by UK Research and Innovation's (UKRI) Biotechnology and Biological Science Research Council (BBSRC)-funded Soil to Nutrition strategic programme (BBS/E/C/000I0310 for A.L.N., X.Z., D.H., I.M.C. and J.W.C., and BBS/E/C/000I0320 for T.T. and L.M.C.). The Broadbalk Wheat Experiment is part of the Rothamsted Long-term Experiments National Capability supported by BBSRC (BBS/E/C/000J0300 for M.L.G.) and the Lawes

Agricultural Trust. H.A.B. was supported by funding from the Soils Training and Research Studentships programme provided by UKRI's BBSRC and Natural Environment Research Council. L.-J.G. and R.K. were supported by the Hartree National Centre for Digital Innovation, a collaboration between UKRI's Science and Technology Facilities Council and IBM Research Europe.

Author contributions

A.L.N., conceptualization, methodology, software, formal analysis, writing—original draft, writing—review and editing, visualization, funding acquisition; H.A.B., formal analysis, investigation; A.B.-L., formal analysis, investigation; Y.Q., investigation; X.Z., conceptualization, investigation, software, writing—review and editing; T.T., conceptualization, methodology, formal analysis, investigation; V.R., formal analysis, investigation; D.H., software, data curation; I.M.C., software, investigation, data curation; L.M.C., resources, writing—review and editing, supervision; L.-J.G., software, formal analysis, data curation; R.K., software, formal analysis, data curation; M.L.G., data curation, writing—review and editing; K.R., conceptualization, writing—review and editing, supervision; S.J.M., conceptualization, writing—review and editing, resources, supervision; J.W.C., conceptualization, writing—review and editing, funding acquisition.

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/s43016-022-00671-z.

Correspondence and requests for materials should be addressed to Andrew L. Neal.

Peer review information *Nature Food* thanks Shuli Niu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Andrew L. Neal

Last updated by author(s): Nov 7, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	C++ code for lattice Boltzman simulation of the dissolution and diffusion of oxygen in 3D soil structures under different soil water contents is available at https://doi.org/10.6084/m9.figshare.21493704. Data related to soil porosity and connected porosity were generated using ImageJ version 1.51j8 and QuantIm version 4.01.
Data analysis	Analysis of data relating to soil porosity and lattice Boltzmann simulations was conducted using PAST version 4.06b, data analysis relating to nitrous oxide emissions was conducted using SigmaPlot (Systat) version 14.5. Individual metagenomic sequences were mapped to the RefSeq non-redundant protein database held at NCBI (downloaded August 22nd, 2018) using DIAMOND version 0.8.27. MEGAN Ultimate version 6.10.252 was used to associate metagenome sequences with Kyoto Encyclopaedia of Genes and Genomes (KEGG) functional orthologs and modules. Multivariate analysis of gene orthologs was conducted using PRIMER PERMANOVA+ version 7.0.17 (PRIMER-e, Auckland, New Zealand). Analysis of ortholog relative abundance was performed using MicrobiomeAnalyst (https://www.microbiomeanalyst.ca/MicrobiomeAnalyst/home.xhtml). All figures were generated using Powerpoint 2019 (Microsoft) and SigmaPlot.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data relating to the Broadbalk and Highfield long-term experiments can be accessed via the electronic Rothamsted Archive (http://www.era.rothamsted.ac.uk/experiment/rbk1 and http://www.era.rothamsted.ac.uk/experiment/rrn1, respectively). Historical data relating to nitrogen inputs (as ammonium nitrate fertiliser or FYM, within seed grain and atmospheric deposition), off-takes (as harvested grain and straw), and soil nitrogen stocks are available from http://doi.org/10.23637/rbk1-yldS10115-01. X-ray computed tomography was performed at the Hounsfield Facility of the University of Nottingham where the resulting images are stored. Original greyscale and converted binary images are available upon reasonable request from Sacha Mooney, University of Nottingham.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	No human subjects were involved in this research.
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This was a study of the effect of long-term agricultural management, including inorganic fertilizer and farmyard manure fertility management, upon the physico-chemical and biological status of soil. The site was established in 1843 to investigate the long-term consequences of different fertiliser and manure applications on the yield of winter wheat, as well as the effect of cessation of all cultivation from a part of the site in 18821. The plots were subject to contrasting fertility management: receiving 35 Mg/ha/yr composted farmyard manure from cattle (FYM); 144 kg-N/ha/yr as ammonium nitrate together with 90 kg-K/ha/yr as K2SO4, 35 kg-P/ha/yr as triple superphosphate [Ca(H2PO4)2·H2O], and 12 kg-Mg/ha/yr as MgSO4 (144NPK); 192 kg-N/ha/yr as ammonium nitrate together with 90 kg-K/ha/yr as K2SO4 and 12 kg-Mg/ha/yr as MgSO4 (i.e. lacking phosphorus fertilization, 192NK); 90 kg-K/ha/yr as K2SO4, 35 kg-P/ha/yr as triple superphosphate, and 12 kg-Mg/ha/yr as MgSO4 (i.e. lacking nitrogen fertilization, PK); and soil which had never received any fertilization for the whole duration of the experiment (nil). These arable plots are sown continuously with winter wheat. In addition to these regularly tilled arable soils, we also studied two soils which had not been subject to management for over a century: part of the original Broadbalk experiment which since 1882 has been allowed to revert naturally to a mixed species woodland of ash (Fraxinus excelsior L.), sycamore (Acer pseudoplatanus L.) and hawthorn (Crataegus monogyna Jacq.) which was sampled at the same time as the plots described above; soil was also collected in 2015 from plots of the Highfield Ley-Arable experiment (http://www.era.rothamsted.ac.uk/experiment/rrn1) managed as a mown sward of mixed fescue (Festuca pratensis L.), Timothy grass (Phleum pratense L.) and white clover (Trifolium repens L.) since 1838. treatments are not replicated on the Broadbalk experiment, three pseudo-replicates were collected from each treatment plot for nucleic acid extraction and four replicates for X-ray computed tomography (CT) imaging. These pseudo-replicates were spaced equidistant along the 18-m long plot. Replicate plots are available on the Highfield Ley-Arable experiment and soil was sampled from each of three replicates. The top 10-cm of soil was sampled with a 3-cm diameter auger for nucleic acid extraction. For each pseudo-replicate, ten cores were pooled and mixed thoroughly whilst sieving through a 2-mm mesh. Samples were then frozen and stored at -80 °C. For X-ray computed tomography, cylindrical soil cores (6.8-cm diameter, 12.0-cm height) were collected and stored undisturbed in unplasticized polyvinyl chloride pipes at 4 °C until they were scanned. Measurement of gaseous emissions from 240NPK, FYM and PK soils were taken on fourteen dates between April 11th and October 7th 2019, and measurements from 144NPK and 192NK soils were taken on eleven dates between April 23rd and October 7th 2019. Gas sampling was performed using in-field static chambers. Three chambers (dimensions 40 cm x 40 cm x 25 cm height) were inserted to a depth of 5 cm in each treatment soil. Gas samples were taken with a syringe from the headspace at the time of chamber closure and again after 40 minutes. Samples were transferred in the field to pre-evacuated 20-

	mL vials. Once wheat plants had become too tall for a single chamber, additional chambers were stacked on top and a correction to the headspace volume was applied to calculate gas fluxes. On each sampling day, two chambers were selected to collect a third sample at 60 minutes to check the validity of the linearity assumption of gas accumulation in the chamber headspace.
Research sample	Samples represented soil from the Broadbalk winter wheat long-term experiment and permanent grassland from the Highfield Ley-Arable experiment, both maintained by Rothamsted Research.
Sampling strategy	sampling strategy was based upon previous experience of working with the experimental soils and financial constraints associated with metagenomic sequencing and X-ray computed tomography.
Data collection	Various individuals collected data relating to different aspects of this study. Bacq-Labreuil generated and analyzed the X-ray computed tomography of soil pore space under supervision of Ritz and Mooney. Lattice Boltzmann simulation derived from soil imaging was performed by Zhang. DNA was extracted from soils by Clark and Neal. Hughes, Qin, Neal, Krishna and Gardiner all contributed to bioinformatic analysis of sequence data.
Timing and spatial scale	1) soil organic carbon and total nitrogen stocks 1843 - 2015; 14 data points over the period; stocks of carbon and nitrogen in soil are known to change slowly and this sampling frequency is sufficient to capture the general trends for each soil management approach. 2) X-ray computed tomography, October 2015; three pseudo-replicate soil cores collected, nine aggregates scanned at a 1.5-micrometre resolution. This is a typical number of core replicates, and many studies employing X-ray computed tomography of soils in the literature adopt similar levels of replication. 3) soil metagenomics, October 2015; DNA was extracted from three pseudo-replicate soil samples for each treatment, this was largely dictated by financial constraints 4) nitrous oxide emissions, eleven sampling dates between 17th September 2018 and 23rd September 2019. Sampling was performed between 10:00 to 14:00 hours as this was considered representative of the daily mean. Sampling frequency was established following Cárdenas et al. (Cárdenas, L.M., et al., (2019). Nitrogen use efficiency and nitrous oxide emissions in five UK fertilised grasslands. Sci. Total Environ. 661, 696–710.) with higher frequency following fertiliser application (up to five samplings on the first two weeks), then twice per week for three weeks, followed by one sample every two weeks for five months followed by once per month until the end of the experiment.
Data exclusions	No data were excluded.
Reproducibility	Not applicable
Randomization	The field experiment used in this research incorporates no randomization.
Blinding	Blinding was not relevant to this study as it was based upon a long-term field experiment.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	The soil is classified as a Chromic luvisol. The soil texture is described as clay loam to silty clay loam over clay-with flints. The soils contain a large number of flints and are slightly calcareous. Below about 2-m depth the soil becomes chalk. The experiment is under-drained and the site is free draining. There is considerable variation in soil texture across the site, with clay contents ranging from 19% - 39%. Thirty-year mean meteorological data (1991-2020) annual rainfall 763.5 mm, annual temperature 10.2 °C, mean soil temperature 30-cm under grass 10.9 °C (http://doi.org/10.23637/OARES30YrMeans9120). Comprehensive information can be found at http://www.era.rothamsted.ac.uk/experiment/rbk1
Location	51˚ 48’ 35” North, 00˚ 22’ 30” West: Elevation, 130 Metres
Access & import/export	All sites associated with this research are managed by Rothamsted Research as part of the Long-term Experiments National Capability supported by UK Research and Innovation’s Biotechnology and Biological Science Research Council. Access was provided by Rothamsted Research and samples were processed on site.
Disturbance	Soil samples were collected with augers. This caused little disturbance to the environment, particularly to ploughed arable soils.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging