

Metagenomics reveals compartment-specific microbiomes and shared carbohydrate-active enzyme (CAZyme) repertoires across manure and soil in integrated crop-livestock systems

Tumelo Isaac Thebe

North-West University

Ben Jesuorsemwon Enagbonma

North-West University

Tshifhiwa Paris Mamphogoro

MamphogoroT@arc.agric.za

Agricultural Research Council-Animal Production

Olubukola Oluranti Babalola

North-West University

Article

Keywords: Integrated crop-livestock systems, Goat manure, Maize rhizosphere, Shotgun metagenomics, CAZymes, Environmental filtering

Posted Date: September 3rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10768653/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Integrated crop-livestock systems connect animal, soil and plant microbiomes, but the extent to which linked compartments share microbial and carbohydrate-processing features while retaining distinct environmental signatures remains unclear. Shotgun metagenomics was used to compare three composite samples each of goat manure (GM), maize rhizosphere soil (RS) and control soil (CR) from a South African mixed-farming system. GM was enriched in *Bacillota* and *Bacteroidota*, whereas *Actinomycetota* predominated in RS and CR. Genus-level Bray-Curtis ordination separated GM from the soil compartments (PERMANOVA $F = 24.88$, $p = 0.002$), although 70 genera were prevalent across all three. A five-variable RDA using harmonised pH, moisture, total N, P and K explained 98.9% of genus-level variation (adjusted $R^2 = 0.971$), but the unique physicochemical component was not significant after conditioning on compartment ($p = 0.329$), and correlation patterns were similarly compartment dependent. All six resolved CAZy classes and 594 EC-resolved features were shared, yet EC-level profiles remained strongly compartment structured (PERMANOVA $F = 151.92$, $p = 0.002$). The results indicate strong compartment-specific microbial assembly alongside conservation of broad carbohydrate-processing potential without functional equivalence.

Introduction

Integrated crop-livestock systems (ICLS) combine crop and animal production within a shared management framework in which nutrients, plant residues and animal-derived organic matter are cycled between system components. This integration can improve resource-use efficiency and soil functioning while reducing dependence on external nutrient inputs ^{1,2}. Bansal, et al. ¹, for example, reported that crop-livestock integration altered soil microbial phospholipid fatty acid profiles and increased carbon and nitrogen associated with soil aggregates. Such benefits arise from a biologically active continuum in which manure deposition, crop residues, root exudation and management history create heterogeneous microhabitats rather than a single homogeneous soil environment ^{3,4}.

Within this continuum, goat manure and the maize rhizosphere represent ecologically connected but strongly contrasting microbial habitats. Manure is a host-derived, nutrient-rich matrix shaped by diet, gastrointestinal passage and post-excretion conditions, whereas the rhizosphere is a root-influenced soil zone structured by plant carbon inputs, soil physicochemistry and microbial competition ^{5,6}. Soil properties can both constrain and be modified by microbial activity, creating feedbacks among pH, nutrient availability, moisture, mineral surfaces and community composition ⁷. Across a meta-analysis of 101 studies, Fu, et al. ⁸ found consistent rhizosphere-associated shifts in bacterial community structure and a tendency for rhizosphere pH to decrease relative to bulk soil, particularly in neutral to alkaline soils. A nearby control soil without routine goat-manure inputs therefore provides an important reference for distinguishing rhizosphere- and management-associated signals from the background soil microbiome.

Manure application may reshape agricultural microbiomes through at least two non-exclusive processes: dispersal of manure-associated microorganisms into soil and modification of the soil environment that subsequently filters resident taxa. Peng, et al.⁹ showed that manure amendments in integrated crop-livestock farms altered bacterial community composition and the relative representation of several major soil phyla. Likewise, long-term field studies indicate that management history can leave persistent microbial legacies even after amendment regimes change¹⁰, while rhizosphere selection can further reorganise communities at the root-soil interface^{6,11}. As a result, compositional differences between manure-exposed rhizosphere soil and control soil cannot be assumed to represent direct microbial transfer; they may instead reflect environmental filtering, plant selection, dispersal, or combinations of these processes.

Taxonomic differentiation also does not exclude ecological continuity across compartments. Microbial generalists and repeatedly dispersed taxa can persist across heterogeneous habitats, producing a prevalence-based core or shared microbiome even when their relative abundances differ substantially^{11,12}. Neu, et al.¹² emphasised that core microbiome definitions depend strongly on prevalence thresholds, taxonomic resolution and study design. In an ICLS, recurrent detection of the same genus in manure and rhizosphere samples can therefore identify a candidate cross-compartment interface, but strain identity, directional transfer, colonisation or functional equivalence require deeper resolution and testing.

Carbohydrate-active enzymes (CAZymes) provide a complementary functional axis for examining whether taxonomically distinct communities retain overlapping capacities for carbohydrate transformation¹³. The CAZy framework encompasses glycoside hydrolases (GH), glycosyltransferases (GT), carbohydrate-binding modules (CBM), carbohydrate esterases (CE), auxiliary activities (AA), and polysaccharide lyases (PL), with lower-level family and enzyme commission (EC) annotations providing progressively finer functional resolution^{14,15}. López-Mondéjar, et al.¹⁶ demonstrated that CAZyme complements are strongly associated with microbial phylogeny and habitat, whereas Berlemont and Martiny¹⁷ found ecosystem-specific carbohydrate-utilisation profiles across nearly 2000 metagenomes. Importantly, metagenomic annotations represent encoded potential rather than realised activity; Zhao, et al.¹⁸ showed that CAZyme-encoding gene inventories can be decoupled from transcript and protein expression. Thus, broad class-level similarity is not equivalent to functional redundancy unless activity evidence supports that inference.

Shotgun metagenomics enables taxonomic and gene-centric functional profiling within the same samples and is therefore well suited to resolving these linked questions across manure and soil habitats^{19,20}. Previous South African studies have demonstrated that maize rhizosphere communities differ from surrounding soils and that physicochemical variables can explain part of this variation²¹. Recent South African metagenomic evidence further shows that cropping history can alter both rhizosphere microbial composition and its association with soil physicochemical conditions. Enagbonma, et al.²² reported distinct sorghum rhizosphere communities under contrasting crop rotations, with Actinomycetota

predominating across treatments and several soil properties contributing to community differentiation. Adebayo, et al. ²³ similarly combined shotgun metagenomics with soil physicochemical analysis and constrained ordination to examine management-associated rhizomicrobiomes. However, comparative evidence spanning animal manure, manure-exposed maize rhizosphere soil and a non-manured soil reference within the same ICLS remains limited, particularly where taxonomic structure, hierarchical CAZy profiles and measured physicochemical gradients are evaluated together.

This study therefore compared microbial community structure, prevalence-based shared taxa and CAZy repertoires among goat manure (GM), maize rhizosphere soil (RS) and control soil (CR) from a South African integrated crop-livestock system. We hypothesised that contrasting compartment conditions would generate distinct microbial assemblages, whereas broad carbohydrate-processing categories would remain widely shared but differ quantitatively at finer CAZy resolution. We addressed three questions: (1) how do taxonomic composition, alpha and beta diversity, and prevalent shared genera differ among GM, RS and CR; (2) to what extent are CAZy repertoires shared or compositionally differentiated across class, family and EC resolutions; and (3) how do measured physicochemical properties covary with genus-level microbial community composition across GM, RS and CR, and which associations persist beyond compartment identity?

Results

Physicochemical characterization of soil and manure

The physicochemical properties differed appreciably between RS and CR (Table 1). RS had lower pH (6.45 ± 0.03 vs. 7.35 ± 0.01) and higher moisture content ($3.03 \pm 0.07\%$ vs. $2.42 \pm 0.08\%$), total N (638.79 ± 5.79 vs. 592.90 ± 3.30 mg kg⁻¹), P (137.00 ± 6.24 vs. 117.67 ± 4.73 mg kg⁻¹), NH₄-N (0.46 ± 0.06 vs. 0.26 ± 0.03 mg kg⁻¹) and NO₃-N (11.95 ± 0.38 vs. 8.44 ± 1.30 mg kg⁻¹) than CR. Mg was also higher in RS, whereas organic C, Ca, S-value and clay content were generally higher in CR; K and sand content were broadly similar. GM was mildly alkaline (pH 7.96 ± 0.13) and had a moisture content of $5.00 \pm 0.12\%$, corresponding to $95.00 \pm 0.12\%$ dry matter (Table 1). Total C averaged $36.47 \pm 1.23\%$, while total N, P and K were $2.37 \pm 0.16\%$, $0.637 \pm 0.025\%$ and $0.517 \pm 0.049\%$ dry weight, respectively; NH₃-N averaged 474.44 ± 9.35 mg kg⁻¹. Several RS and CR parameters showed complete separation among the replicates; the minimum attainable two-sided exact permutation p-value was 0.10; these differences were therefore interpreted descriptively and together with the multivariate analyses.

Table 1

(a) and (b). Physicochemical properties of maize rhizosphere soil (RS), control soil (CR), and goat manure (GM). Values are mean $\pm$ SD, Exact p-values are from two-sided permutation tests comparing RS and CR.

(a). Physicochemical properties of maize rhizosphere soil (RS) and control soil (CR).			
Physicochemical parameter	RS	CR	Exact p (RS vs CR)
pH	6.45 $\pm$ 0.03	7.35 $\pm$ 0.01	0.10
Moisture content (%)	3.03 $\pm$ 0.07	2.42 $\pm$ 0.08	0.10
Total N (mg kg ⁻¹)	638.79 $\pm$ 5.79	592.90 $\pm$ 3.30	0.10
P (mg kg ⁻¹)	137.00 $\pm$ 6.24	117.67 $\pm$ 4.73	0.10
NH ₄ -N (mg kg ⁻¹)	0.46 $\pm$ 0.06	0.26 $\pm$ 0.03	0.10
Organic C (%)	0.61 $\pm$ 0.01	0.64 $\pm$ 0.01	0.10
NO ₃ -N (mg kg ⁻¹)	11.95 $\pm$ 0.38	8.44 $\pm$ 1.30	0.10
K (% dry weight)	6.40 $\pm$ 0.70	6.43 $\pm$ 0.29	1.00
Ca (%)	69.73 $\pm$ 1.97	72.73 $\pm$ 1.58	0.20
Mg (%)	23.77 $\pm$ 1.33	20.60 $\pm$ 1.50	0.10
Na (%)	0.10 $\pm$ 0.17	0.23 $\pm$ 0.06	0.40
S-value	4.328 $\pm$ 0.077	5.006 $\pm$ 0.360	0.10
Sand (%)	79.57 $\pm$ 0.93	79.05 $\pm$ 1.08	0.50
Silt (%)	1.86 $\pm$ 0.16	1.35 $\pm$ 0.50	0.30
Clay (%)	18.57 $\pm$ 0.78	19.60 $\pm$ 0.68	0.20
(b). Physicochemical properties of goat manure (GM).			
Physicochemical parameter	GM		
pH	7.96 $\pm$ 0.13		
Moisture content (%)	5.00 $\pm$ 0.12		
Dry matter (%)	95.00 $\pm$ 0.12		
Total C (%)	36.47 $\pm$ 1.23		
Total N (% dry weight)	2.37 $\pm$ 0.16		
P (% dry weight)	0.637 $\pm$ 0.025		
K (% dry weight)	0.517 $\pm$ 0.049		
NH ₃ -N (mg kg ⁻¹)	474.44 $\pm$ 9.35		

Redundancy analysis and correlations

The harmonised five-variable RDA showed a strong association between physicochemical gradients and genus-level community composition across GM, RS and CR (Supplementary Table S1). The model explained 98.9% of total variation (adjusted $R^2 = 0.971$; pseudo-F = 54.52; exact $p = 0.00035$), with RDA1 and RDA2 accounting for 88.7% and 6.5% of total variation, respectively (Fig. 1). Total N, P and moisture aligned with GM, whereas K was oriented towards the soil compartments and pH contributed to the manure-soil gradient. Predictor collinearity was high (VIFs: pH 8.5, moisture 51.8, total N 273.7, P 390.8 and K 57.1); accordingly, individual vectors were not interpreted as independent effects. After conditioning on compartment identity, the unique physicochemical fraction was approximately 9.3% of total variation and was not significant under restricted permutation testing ($p = 0.329$).

Correlation analyses produced the same compartment-dependent pattern (Fig. 2). Across all nine samples, 96 of 100 Pearson associations between the 20 dominant genera and five harmonised variables remained significant after FDR correction (Fig. 2a). Strong relationships included *Clostridium* with total N ($r = 0.9998$), *Sphingomonas* with K ($r = 0.9991$) and *Solirubrobacter* with P ($r = -0.9984$). However, only five associations remained nominally significant when permutations were restricted within GM, RS and CR, and none survived FDR correction. In the soil-only Spearman analysis (Fig. 2b), 39 of 300 associations had nominal exact $p < 0.05$, including *Baekduia*-moisture ($p = 1.000$), *Sphingomonas*-K ($p = 0.971$) and *Conexibacter*-total N ($p = 0.943$), but none remained significant after global FDR correction.

Information on sequence data

Across the nine metagenomes, raw sequencing output averaged 16.80 ± 3.02 Gb per sample, and clean bases averaged 16.74 ± 2.99 Gb per sample. Mean Q20 and Q30 scores were $99.18 \pm 0.16\%$ and $96.69 \pm 0.63\%$, respectively. GC content averaged $60.38 \pm 9.95\%$. Assembly generated $1,136,991 \pm 191,765$ scaffolds per sample, with a mean scaffold length of 947.34 ± 132.31 bp (Supplementary Table S2).

Microbial community composition across compartments

Kingdom-level composition

Kingdom-level composition differed among compartments (Supplementary Fig. S1). Bacteria dominated all samples and accounted for $83.92 \pm 1.03\%$ in GM, $92.31 \pm 0.50\%$ in RS and $93.74 \pm 0.31\%$ in CR (Kruskal-Wallis $p = 0.027$). Archaea contributed $2.15 \pm 1.09\%$, $1.54 \pm 0.34\%$ and $1.30 \pm 0.12\%$ in GM, RS and CR, respectively. Viral assignments were more abundant in GM ($1.71 \pm 0.13\%$) than in RS and CR (approximately 0.01% each). The unresolved fraction was also higher in GM ($12.19 \pm 1.91\%$) than in RS ($6.13 \pm 0.30\%$) and CR ($4.94 \pm 0.18\%$).

Phylum-level composition

Phylum-level composition showed pronounced compartmental differentiation (Fig. 3). GM was dominated by *Bacillota* ($46.95 \pm 1.73\%$), followed by *Pseudomonadota* ($12.44 \pm 9.35\%$) and *Bacteroidota* ($11.60 \pm 10.06\%$). RS was dominated by *Actinomycetota* ($60.19 \pm 2.81\%$), followed by *Pseudomonadota* ($9.82 \pm 0.39\%$), *Chloroflexota*, and *Acidobacteriota*. CR was also dominated by *Actinomycetota* ($60.30 \pm 1.23\%$), but its secondary profile comprised *Pseudomonadota* ($11.65 \pm 0.75\%$), *Chloroflexota*, *Acidobacteriota*, and other soil-associated phyla. *Bacteroidota* differed significantly among compartments and was highest in GM ($p = 0.027$). *Actinomycetota* and *Bacillota* showed strong directional contrasts but did not reach $p < 0.05$ in the three-group tests ($p = 0.066$ and $p = 0.051$, respectively).

Genus-level composition

The top 20 resolved genera displayed clear manure-versus-soil abundance patterns (Fig. 4). *Solirubrobacter* was detected at 0.001% in GM and was more abundant in RS (4.62%) and CR (5.81%; $p = 0.027$). *Nocardioides* showed a similar pattern (GM = 0.004%, RS = 2.21%, CR = 3.46%; $p = 0.027$). *Bradyrhizobium* was highest in RS (1.82%), followed by CR (1.60%) and GM (0.12%; $p = 0.039$). In contrast, *Acinetobacter* was highest in GM (8.94%) and occurred at 0.003% in RS and 0.009% in CR ($p = 0.039$). *Desemzia* was also highest in GM (1.38%), compared with 0.000013% in RS and 0.00083% in CR ($p = 0.027$). *Sphingomonas*, *Alistipes*, *Bacteroides* and *Methanobrevibacter* showed corresponding compartment-associated trends, but their three-group comparisons were not significant ($p = 0.061$).

Taxonomic alpha diversity

Genus-level alpha diversity differed by metric (Fig. 5a-c). Observed richness was highest in GM (3475.0 ± 120.2), followed by RS (3342.7 ± 98.9) and CR (3311.0 ± 52.4), but the difference was not significant ($H = 4.62$, $p = 0.099$). Shannon diversity differed among compartments ($H = 7.20$, $p = 0.027$) and was highest in GM (2.714 ± 0.061), followed by CR (2.251 ± 0.127) and RS (1.991 ± 0.049). Simpson diversity ($1 - D$) showed the same ordering and also differed among groups ($H = 7.20$, $p = 0.027$), with values of 0.715 ± 0.032 in GM, 0.589 ± 0.032 in CR, and 0.520 ± 0.022 in RS.

Taxonomic beta diversity

Bray-Curtis PCoA showed clear genus-level separation of GM from RS and CR (Supplementary Fig. S2). PCoA1 explained 88.09% of the variation and separated the manure samples from the two soil-derived compartments. PCoA2 explained 8.46% and mainly represented within-GM variation, while RS and CR remained closely positioned. PERMANOVA confirmed a significant compartment effect on genus-level composition ($F = 24.88$, $p = 0.002$; 999 permutations). Multivariate dispersion did not differ significantly among compartments ($H = 5.42$, $p = 0.066$), indicating that the PERMANOVA result was not primarily attributable to unequal within-group dispersion.

Shared prevalent genera across compartments

Prevalence-based analysis identified extensive but uneven overlap among compartments (Fig. 6; Supplementary Table S3). At the 70% prevalence threshold, 70 genera were detected in all three compartments. An additional 71 genera were shared exclusively by GM and RS, 684 by RS and CR, and 53 by GM and CR. Thus, the total numbers detected in both members of each pair, including the three-way intersection, were 141 for GM-RS, 754 for RS-CR and 123 for GM-CR. The markedly larger RS-CR overlap was consistent with the close taxonomic similarity of the two soil-derived compartments. The genera recurrently detected in GM and RS included *Bradyrhizobium*, *Sphingomonas*, *Streptomyces*, *Nocardioides*, *Azospirillum*, *Bacillus*, *Bifidobacterium*, *Clostridium*, *Escherichia* and *Mycobacterium* (Supplementary Table S3). These results describe recurrent genus-level detection only and do not establish strain identity, transfer between compartments or confirmed plant-growth-promoting or probiotic activity.

CAZyme composition across compartments

CAZy level 1 functional profiles

CAZy level 1 profiles were dominated by the unresolved fraction, but the resolved component was consistently composed mainly of GH, GT and CBM classes (Fig. 7). GH accounted for $3.45 \pm 0.40\%$ in GM, $3.35 \pm 0.03\%$ in RS and $3.96 \pm 0.10\%$ in CR ($p = 0.099$). GT differed among compartments and was lower in GM ($2.51 \pm 0.03\%$) than in RS ($3.87 \pm 0.05\%$) and CR ($4.18 \pm 0.03\%$; $p = 0.027$). CBM showed the same directional pattern, with $0.88 \pm 0.04\%$ in GM, $1.26 \pm 0.02\%$ in RS and $1.44 \pm 0.04\%$ in CR. CE contributed $0.36 \pm 0.03\%$, $0.40 \pm 0.00\%$ and $0.43 \pm 0.01\%$ in GM, RS and CR, respectively.

CAZy level 2 functional profiles

At CAZy level 2, the dominant resolved functional families included GT2, GT4, GH13, CBM50, GT1, GH28, GH3, GH5, GH23, CBM48, GT51, GH43 and GH1 (Supplementary Fig. S3; Supplementary Table S4). GT2 was higher in RS (1.30%) and CR (1.33%) than in GM (0.75%; $p = 0.039$). GT4 showed the same directional pattern (GM = 0.47%, RS = 1.09%, CR = 1.12%; $p = 0.051$). GH13 was detected across all compartments and was highest in CR ($0.70 \pm 0.02\%$), followed by RS ($0.56 \pm 0.00\%$) and GM ($0.50 \pm 0.03\%$). In contrast, CBM50 was highest in GM (0.35%) compared with RS (0.25%) and CR (0.26%; $p = 0.027$). Several additional GH, GT and CBM families differed in relative abundance among compartments (Supplementary Fig. S3).

CAZy EC-level profiles

EC-resolved CAZy profiles were dominated by glycosyltransferase-associated annotations (Fig. 8; Supplementary Table S5). The most abundant features included hexosyltransferases (EC 2.4.1.-), UDP-Glc:1,2-diacylglycerol 3-glucosyltransferase (EC 2.4.1.157), N-acetylgalactosaminyl transferases, GDP-mannose alpha-mannosyltransferases, rhamnosyl transferases, glucuronosyltransferases and beta-glucosyltransferases. Most dominant EC features had higher relative abundance in RS and CR than in GM. Hexosyltransferases averaged $0.62 \pm 0.00\%$ in GM, $0.88 \pm 0.01\%$ in RS and $0.87 \pm 0.00\%$ in CR. UDP-

Glc:1,2-diacylglycerol 3-glucosyltransferase averaged $0.56 \pm 0.00\%$, $0.78 \pm 0.01\%$ and $0.76 \pm 0.01\%$ in GM, RS and CR, respectively. N-acetylgalactosaminyltransferases and GDP-mannose alpha-mannosyltransferases showed comparable soil-associated patterns.

Alpha diversity of CAZy functional profiles

EC-resolved CAZy alpha diversity differed among compartments (Fig. 9a-c). Observed richness was highest in GM (797.7 ± 4.0), followed by CR (787.7 ± 4.0) and RS (786.0 ± 4.4); the overall test was borderline at the predefined threshold ($H = 6.01$, $p = 0.050$). Shannon diversity differed among groups ($H = 7.20$, $p = 0.027$) and was highest in CR (4.708 ± 0.017), followed by RS (4.586 ± 0.008) and GM (4.093 ± 0.067). Simpson diversity showed the same ordering (CR = 0.920 ± 0.001 , RS = 0.909 ± 0.002 , GM = 0.831 ± 0.010 ; $H = 7.20$, $p = 0.027$).

Beta diversity of CAZy functional profiles

Bray-Curtis PCoA of EC-resolved CAZy profiles showed clear separation among compartments (Fig. 9). PCoA1 explained 95.49% of the variation and separated GM from RS and CR, whereas PCoA2 explained 2.74% and represented residual variation among the soil-derived samples. PERMANOVA confirmed a significant compartment effect ($F = 151.92$, $p = 0.002$; 999 permutations) (Fig. 10). Multivariate dispersion did not differ significantly among compartments ($H = 5.07$, $p = 0.079$). The EC-level PERMANOVA signal was therefore consistent with differences in group centroids rather than a significant dispersion effect.

Shared CAZy features across compartments

Shared-feature analysis showed broad overlap in CAZy repertoires (Supplementary Fig. S4; Supplementary Table S6). All six resolved level 1 classes were detected in every compartment. At level 2, 179 families were shared across all three compartments; an additional 181 were shared exclusively by GM and RS, 229 by RS and CR, and 180 by GM and CR. At EC resolution, 594 features occurred in all three compartments, with 599 additional GM-RS, 645 RS-CR, and 557 GM-CR pairwise-exclusive features.

Discussion

This study combined shotgun metagenomics, physicochemical profiling and community analyses to examine how microbial communities are organised across goat manure (GM), maize rhizosphere soil (RS) and control soil (CR) within an integrated crop-livestock system. The first research question concerned taxonomic composition, diversity and the extent of shared genera among compartments. The strongest result was the separation of GM from both soil-derived compartments at genus level: PCoA1 accounted for 88.09% of the variation and PERMANOVA confirmed a marked compartment effect ($F = 24.88$, $p = 0.002$), whereas dispersion did not differ significantly (Supplementary Fig. S2). This multivariate separation was consistent with the contrasting phylum profiles, in which GM was dominated

by Bacillota and Bacteroidota while Actinomycetota accounted for approximately 60% of both RS and CR communities (Supplementary Fig. S3). Comparable manure-versus-soil differentiation has been reported where animal-derived amendments introduce a distinct microbial pool but soil conditions subsequently impose strong ecological filtering^{9,24}.

The taxonomic contrasts were also evident among dominant genera (Fig. 4). *Acinetobacter* reached 8.94% in GM but was nearly absent from RS and CR, whereas *Solirubrobacter* and *Nocardioides* were strongly associated with the soil compartments; *Bradyrhizobium* was most abundant in RS (1.82%). These patterns are consistent with the ecological origins of the sampled matrices. *Bacillota*- and *Bacteroidota*-rich communities are characteristic of ruminant-associated environments in which fermentation and degradation of dietary substrates are prominent²⁵, whereas Actinomycetota and several of the soil-enriched genera are commonly associated with organic-matter turnover and nutrient transformations in terrestrial habitats^{5,26}. Peng, et al.⁹ similarly observed restructuring of major bacterial groups after manure amendment, although the direction and magnitude of individual taxon responses differed, reinforcing the context dependence imposed by soil chemistry, vegetation and management history⁷.

Alpha- and prevalence-based analyses refined this interpretation by showing that compartment differentiation was driven more by community structure than by a simple loss or gain of taxa. Observed genus richness did not differ significantly, whereas Shannon and Simpson diversity were highest in GM and lowest in RS (Fig. 5). Such a pattern is compatible with rhizosphere selection, in which root-derived carbon, nutrient gradients and competition favour a subset of taxa without necessarily reducing the total number detected^{6,8}. At the same time, 70 genera were recurrently detected across all three compartments, but the RS-CR overlap was much larger (754 genera) than GM-RS (141) or GM-CR (123) overlap (Fig. 6; Supplementary Table S3). Neu, et al.¹² emphasised that a core microbiome is operationally dependent on prevalence and abundance thresholds; accordingly, the shared genera identified here represent consistently detected taxa under the specified threshold rather than a universal core. Their recurrence may reflect ecological generalism, repeated dispersal or independent occurrence of different species within the same genus, and therefore does not establish manure-to-rhizosphere strain transfer²⁷.

Several genera shared between GM and RS, including *Bradyrhizobium*, *Sphingomonas*, *Streptomyces*, *Azospirillum* and *Bacillus*, contain strains with documented plant-beneficial and probiotic traits (Supplementary Table S7)^{28,29}. Htwe, et al.³⁰, for example, demonstrated plant-growth benefits from *Bradyrhizobium*- and *Streptomyces*-based inoculation, while Kim, et al.³¹ characterised plant-growth-promoting traits in a *Sphingomonas* rhizobacterium. Their presence in the shared set is therefore ecologically noteworthy, but genus-level metagenomic detection cannot be used to assign plant-growth-promoting or probiotic phenotypes. Such traits are strain-dependent and require culture-based, genome-resolved or phenotype-linked validation^{32,33}. The taxonomic findings thus support a restricted,

selectively beneficial, but potentially important manure-soil-plant microbial interface while maintaining a clear distinction between recurrent occurrence, transfer and demonstrated function.

The second research question asked whether carbohydrate-processing potential was conserved despite the strong taxonomic differentiation. The data supported this hypothesis at broad functional resolution. All six resolved CAZy classes occurred in every compartment, yet their proportional composition differed; notably, GT was lower in GM (2.51%) than in RS (3.87%) and CR (4.18%) (Fig. 7), while several level-2 GT, GH and CBM families also showed compartment-associated enrichment (Supplementary Fig. S3). At finer resolution, Shannon and Simpson diversity of EC-resolved CAZy profiles differed among compartments (Fig. 9), and EC-level PCoA showed even stronger separation than the taxonomic ordination, with PCoA1 explaining 95.49% of variation and PERMANOVA yielding $F = 151.92$ ($p = 0.002$) (Fig. 10). Nevertheless, 594 EC-resolved features were prevalent in all three compartments (Supplementary Fig. S4; Supplementary Table S6). López-Mondéjar, et al. ¹⁶ showed that CAZyme complements are shaped jointly by microbial phylogeny and habitat, while Berlemont and Martiny ¹⁷ demonstrated ecosystem-specific carbohydrate-utilisation profiles across environmental metagenomes. The present findings fit this framework: broad carbohydrate-processing categories were conserved, but their relative organisation remained strongly habitat-specific.

The distribution of individual CAZy families provides a biological basis for this pattern. Soil-associated enrichment of GT2, GT4 and several glycosyltransferase-related EC features (Fig. 8) is consistent with greater relative representation of glycan biosynthesis, cell-envelope and extracellular-polysaccharide-associated processes, whereas GH3, GH5, GH13, GH28, GH43, CBM48 and CBM50 indicate potential for transformations involving beta-glucosides, cellulose, hemicellulose, pectin, alpha-glucans and microbial cell-wall substrates ^{34,35} (Supplementary Tables S4-S5). Glycosyltransferases construct diverse glycoconjugates, while carbohydrate-binding modules facilitate interactions between carbohydrate-active enzymes and substrates ^{36,37}. The coexistence of these functions is plausible in a system receiving root-derived carbon, crop residues and animal-derived plant fibre ^{38,39}. The presence of these features in the GM compartment is consistent with digestive activities of livestock and ruminant gut microbiomes ^{40,41}. Zhao, et al. ¹⁸ showed that CAZyme-encoding gene inventories can be decoupled from transcript and protein expression; the shared CAZy-annotated repertoire observed here should therefore be interpreted as conservation of broad carbohydrate-processing potential. Expression- or activity-based measurements would further determine whether similar genes perform comparable biochemical functions across GM, RS and CR ^{18,42}.

The third research question examined how measured physicochemical properties covaried with microbial community composition. RS had lower pH than CR (6.45 vs. 7.35) but higher moisture, total N, P, $\text{NH}_4\text{-N}$ and $\text{NO}_3\text{-N}$, whereas GM represented a distinct chemical matrix with pH 7.96, 2.37% total N and 0.637% P (Table 1). These contrasts are ecologically plausible because pH, water availability and nutrient pools influence microbial physiological constraints and substrate availability ⁷. Fu, et al. ⁸ similarly reported consistent rhizosphere-associated shifts in bacterial structure and pH, while Peng, et al. ⁹ linked

manure-associated community differences to environmental properties, finding that soil bacterial communities are largely influenced by pH and moisture among other factors. Additionally, they note that the type of animal manure amendment contributes to community variations. Thus, physicochemical differentiation provides a credible environmental context for the taxonomic patterns, although it does not by itself establish causal effects.

The five-variable RDA captured this coordinated gradient strongly: pH, moisture, total N, P and K jointly explained 98.9% of genus-level variation (adjusted $R^2 = 0.971$; exact $p = 0.00035$), with 88.7% represented on RDA1 (Fig. 1). Enagbonma, et al.²² likewise found that several edaphic variables, including P, inorganic N forms and texture, covaried with sorghum-rhizosphere microbial composition, and their BEST analysis identified a multivariable subset explaining approximately 36% of community variation. Adebayo, et al.²³ and Akinola, et al.²¹ also reported context-dependent physicochemical associations in South African crop rhizospheres. In the present dataset, however, large VIFs for moisture, N, P and K showed that the five predictors largely represented the same compartmental contrast, and the unique physicochemical fraction was not significant after conditioning on compartment ($p = 0.329$). The RDA therefore supports a shared environmental-compartment gradient instead of the independent effects of individual variables.

The correlation analyses reinforced this interpretation (Fig. 2). In the unrestricted nine-sample Pearson analysis, 96 of 100 genus-parameter associations remained FDR-significant, including *Clostridium*-total N ($r = 0.9998$), *Sphingomonas*-K ($r = 0.9991$) and *Solirubrobacter*-P ($r = -0.9984$) (Fig. 2a). Yet none remained FDR-significant after permutations were restricted within GM, RS and CR. The soil-only Spearman analysis was similarly conservative: 39 of 300 associations were nominally significant, but none survived FDR correction (Fig. 2b). These results indicate that taxa and physicochemical properties vary along the same compartment-associated gradient, rather than demonstrating that single variables independently control individual genera. This interpretation is consistent with Enagbonma, et al.²², who found multiple environmental correlates of rhizosphere community structure rather than one universal edaphic determinant.

The combined taxonomic, CAZy and environmental results support a model in which compartment identity is the dominant organising factor, while a restricted set of taxa and broad carbohydrate-processing capacities remain shared across the animal-soil-plant continuum. The close taxonomic relationship between RS and CR, despite measurable physicochemical differences, suggests that the soil matrix provides a strong background constraint onto which rhizosphere- and management-associated filtering are superimposed (Fig. 3; Table 1). For integrated crop-livestock management, this implies that manure inputs may alter both resource conditions and microbial inocula, but the resulting rhizosphere community remains strongly filtered by the soil-plant environment^{7,9}. Larger spatially and temporally replicated studies integrating metagenome-assembled genomes, strain-resolved tracking, metatranscriptomics, metaproteomics and direct enzyme assays will be required to distinguish microbial transfer from environmental filtering and to determine whether the shared genetic potential translates into comparable ecosystem processes^{12,18}.

Conclusion

Shotgun metagenomics demonstrated strong compartment-specific microbial structuring across goat manure, maize rhizosphere soil and control soil. GM was dominated by Bacillota and Bacteroidota, whereas Actinomycetota predominated in the soils, although 70 genera were recurrently detected across all compartments. Physicochemical gradients were closely aligned with community composition, but partial RDA and restricted/FDR-corrected correlations indicated that individual environmental effects could not be separated from compartment identity. Broad carbohydrate-processing potential was conserved, with all six major CAZy classes and 594 EC-resolved features shared, yet EC-level composition remained strongly compartment-specific. These findings support environmental filtering within a connected manure–soil–plant continuum without demonstrating direct microbial transfer or functional equivalence. Larger replicated, strain-resolved and activity-based studies are required to establish transfer pathways and realised ecosystem functions.

Materials and Methods

Sampling and experimental structure

Sampling was conducted at Mafoko Farm in North West Province, South Africa (25.99° S, 25.94° E). The farm operates as an integrated crop-livestock system under a semi-arid climate and contains two plots approximately 60 m apart. Plot A had a history of continuous maize cultivation and regular goat-manure application, whereas Plot B had a crop-rotation history and did not receive routine goat-manure inputs. At sampling, *Zea mays* L. cv. PANNAR 12 was approximately 45 days after sowing at the early tasselling stage. The farm maintained a herd of Boer goats (*Capra hircus*) that grazed locally available vegetation and received supplemental feed and crop residues.

Three composite RS samples were collected from Plot A. Each composite comprised soil tightly adhering to the roots of five maize plants sampled at spatially separated points. Three CR composites were collected from Plot B, with each composite formed from four soil points. Three GM composites were obtained from freshly excreted droppings collected from three separate goat pens. Sterile nylon sheets were placed on pen floors before collection to minimise contact with underlying material. Approximately 500 g of each composite was transferred to sterile bags, transported under cold-chain conditions and stored at -20°C until DNA extraction⁴³.

Soil and manure physicochemical characterization

Physicochemical analyses were performed on separate aliquots derived from the same spatial composite units used for metagenomic sampling. For RS and CR, approximately 20 g of each composite soil sample was air-dried at room temperature, homogenised and passed through a 2 mm sieve before analysis. Soil pH was measured potentiometrically in a 1:2.5 soil-to-solution suspension. Particle-size distribution was determined by the hydrometer method and expressed as percentages of sand, silt and clay following the approach described by Hunduma and Kebede⁴⁴. Soil organic carbon was quantified

using the Walkley-Black dichromate oxidation method ⁴⁵, while total nitrogen was measured after digestion and mineral nitrogen fractions (NH₄-N and NO₃-N) were quantified colorimetrically after extraction. Available phosphorus was determined using Bray-1 extraction followed by spectrophotometric quantification, consistent with soil physicochemical workflows used in comparable rhizosphere metagenomic studies ²³.

Exchangeable K, Ca, Mg and Na were extracted with 1 M ammonium acetate at pH 7 and quantified instrumentally; K was expressed as percentage dry weight in the laboratory output, whereas soil N and P were expressed as mg kg⁻¹. The S-value was calculated from the sum of extractable base cations. For GM, pH and moisture were determined using procedures comparable with those applied to the soil matrices, and dry matter was calculated as 100 minus moisture content. Total C, total N, P, K and NH₃-N were determined using standard solid-manure analytical procedures, with N, P and K expressed on a dry-weight basis ⁴⁶. Because soil and manure differ in matrix composition and reporting basis, physicochemical results were interpreted within matrix and only directly comparable variables were used in cross-compartment constrained ordination.

DNA extraction, library preparation, and sequencing

Total community DNA was extracted using the ZymoBIOMICS DNA Miniprep Kit (Zymo Research, USA) according to the manufacturer's protocol for heterogeneous environmental matrices. DNA concentration and purity were assessed using a NanoPhotometer NP80 Touch. Libraries were prepared using an Illumina-compatible workflow. DNA was fragmented to approximately 350 bp, end-repaired, adapter-ligated, purified and PCR-amplified. Library integrity and insert-size distribution were assessed using Advanced Analytical Technologies instrumentation, and libraries were quantified by qPCR. Paired-end sequencing (2 × 150 bp) was performed on an Illumina NovaSeq X Plus platform at Novogene, Singapore.

Quality control, assembly, and gene catalogue construction

Raw reads were quality-filtered with fastp v0.23.4, a one-pass FASTQ preprocessing tool described by ⁴⁷. Read pairs were removed when either read contained adapter contamination, more than 10% ambiguous bases, or more than 50% bases with a Phred score below 5. Clean reads were screened against *Zea mays* and *Capra hircus* reference genomes using Bowtie2 v2.5.4 with the parameters --end-to-end --sensitive -I 200 -X 400, and host-matching reads were removed before assembly ⁴⁸.

High-quality non-host reads were assembled de novo using MEGAHIT v1.1.2 with the meta-large preset. Scaffolds were split at internal ambiguous bases to produce scaftigs, and open reading frames were predicted from scaftigs of at least 500 bp using MetaGeneMark v2.10; ORFs shorter than 100 nt were discarded. Redundant genes were collapsed with CD-HIT v4.6.8 at 95% sequence identity and 90% alignment coverage, following the sequence-clustering framework of Li and Godzik ⁴⁹. Reads were then mapped back to the non-redundant catalogue using Bowtie2; genes supported by two or fewer mapped

reads per sample were excluded, and the retained sequences constituted the unigene catalogue. Gene abundance was estimated from mapped-read counts normalised by gene length⁵⁰.

Taxonomic annotation and community structure analysis

Predicted proteins were aligned against the Micro_NR database using DIAMOND v2.1.9 in blastp mode with an e-value threshold of $1e^{-5}$. Hits satisfying $e\text{-value} \leq \text{minimum } e\text{-value} \times 10$ were retained, and taxonomic assignments were resolved using the Lowest Common Ancestor algorithm in MEGAN v6.8.20⁵¹. Taxonomic abundance tables were generated from kingdom to species by summing the abundances of genes assigned to each taxon. Primary structural analyses used phylum- and genus-level profiles because these ranks provided interpretable ecological resolution while limiting overinterpretation of uncertain species-level assignments.

Carbohydrate-active enzyme (CAZyme) annotation and profiling

Unigenes were aligned against the CAZy database using DIAMOND v2.1.9 in blastp mode with an e-value threshold of $1e^{-5}$, and best-hit assignments were used for functional summarisation^{52,53}. Three abundance matrices were analysed: CAZy level 1, representing broad enzyme classes; CAZy level 2, representing lower-level families; and EC-resolved annotations. Level 1 included glycoside hydrolases (GH), glycosyltransferases (GT), carbohydrate-binding modules (CBM), carbohydrate esterases (CE), auxiliary activities (AA) and polysaccharide lyases (PL). Relative abundance at each CAZy level was calculated by summing the abundances of genes assigned to each feature. The resulting matrices were analysed in parallel with the taxonomic profiles. CAZy annotations were interpreted as predicted functional potential rather than evidence of gene expression or enzyme activity^{42,54}.

Core microbiome and shared taxa identification

Shared taxonomic and CAZy features were identified using a prevalence-based framework. Because core microbiome definitions are threshold-dependent¹², an explicit detection and prevalence rule was used. A feature was considered detected when its relative abundance was at least 1×10^{-5} and compartment-prevalent when detected in at least 70% of samples within that compartment. Because each compartment contained three samples, the 70% criterion required detection in all three. Pairwise and three-way overlaps were visualised using Venn diagrams, the unresolved 'Others' category was excluded, and shared-feature results were interpreted as recurrent detection rather than evidence of direct transfer or functional equivalence.

Statistical analysis

All statistical analyses were performed in Python v3.16 using reproducible scripts. Relative-abundance summaries are reported as mean $\pm$ standard deviation for three composite samples per compartment. RS and CR physicochemical variables were compared using exact two-sided permutation tests of the difference in means; with $n = 3$ per group, all 20 unique allocations were enumerated. The alpha level was

0.05, and Benjamini-Hochberg false-discovery-rate correction was evaluated across the 15 soil comparisons. The minimum attainable two-sided exact p-value was 0.10 and these univariate soil comparisons were interpreted descriptively. Feature-wise differences among GM, RS and CR were evaluated using Kruskal-Wallis tests. Genus-level and EC-resolved CAZy alpha diversity were calculated using observed richness, Shannon diversity and Simpson diversity ($1 - D$); overall differences were assessed using Kruskal-Wallis tests, followed where appropriate by two-sided pairwise Mann-Whitney U tests with Benjamini-Hochberg adjustment ⁵⁵.

Bray-Curtis dissimilarity matrices were visualised by principal coordinate analysis (PCoA). Compartment-level differences in genus composition and EC-resolved CAZy composition were tested by PERMANOVA with 999 permutations ⁵⁶, and homogeneity of multivariate dispersion was assessed separately because PERMANOVA can be influenced by unequal within-group dispersion ⁵⁷. Top-feature visualisations were restricted to the 20 most abundant resolved genera or CAZy features; heatmaps used log10-transformed relative abundances after addition of a small pseudocount.

Relationships between physicochemical properties and genus-level microbial composition were assessed using redundancy analysis (RDA) and correlation analyses. For the GM-RS-CR RDA, pH, moisture, total N, P and K were harmonised across all nine samples; soil N and P were converted from mg kg^{-1} to percentage units to match manure measurements. The unresolved “Others” category was excluded, genera occurring at relative abundance $\geq 1 \times 10^{-5}$ in at least three samples were retained, and resolved abundances were renormalised and Hellinger-transformed ⁵⁸. The five variables were fitted jointly as RDA constraints, with significance assessed using all $9! = 362,880$ permutations. Variance inflation factors were used to assess collinearity, and partial RDA conditioning on compartment identity was evaluated with $3!^3 = 216$ within-compartment permutations. Matrix-specific physicochemical variables were displayed only as passive vectors and were not fitted as constraints ^{22,23}.

Pearson correlations were calculated between the 20 most abundant resolved genera and the five harmonised variables across all nine samples using exact two-sided permutation tests and Benjamini-Hochberg correction ⁵⁵. A within-compartment restricted-permutation analysis assessed whether associations persisted beyond the overall compartment effect. As a complementary soil-specific analysis, Spearman correlations were calculated between the 20 dominant genera and 15 physicochemical variables across RS and CR using all $6! = 720$ permutations, global false-discovery-rate correction and a $3! \times 3!$ within-compartment sensitivity test.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests

Funding

This work was supported by Department of Agriculture (South Africa) [Grant number: A148/API012403000048] awarded to T.P.M.

Author Contribution

Conceptualization, T.P.M. and O.O.B.; methodology, T.I.T., B.J.E., T.P.M., and O.O.B.; investigation, formal analysis, data curation and visualization, T.I.T.; writing—original draft preparation, T.I.T.; resource acquisition, software and validation, O.O.B.; funding acquisition, project administration, T.P.M.; supervision and writing—review and editing, B.J.E., T.P.M., and O.O.B. All authors have read and agreed to the published version of the manuscript.

Acknowledgement

The authors acknowledge the management of the participating farm for granting access for sampling.

Data Availability

The shotgun metagenomic sequence data generated and analysed during this study are available in the NCBI Sequence Read Archive under accession numbers SRS28420652, SRS28420653, SRS28420654, SRS28420659, SRS28420660, SRS28420661, SRS28420667, SRS28420668 and SRS28420669.

References

1. Bansal, S., Chakraborty, P. & Kumar, S. Crop–livestock integration enhanced soil aggregate-associated carbon and nitrogen, and phospholipid fatty acid. *Sci. Rep.* **12**, 2781–2781. 10.1038/s41598-022-06560-6 (2022).
2. Farias, G. D. et al. Integrated crop-livestock system with system fertilization approach improves food production and resource-use efficiency in agricultural lands. *Agronomy Sustainable Development* **40** (2020).
3. Carpinelli, S., Fonseca, A. F., d., Neto, P. H. W., Dias, S. H. B. & Pontes, L. d. S. Spatial and Temporal Distribution of Cattle Dung and Nutrient Cycling in Integrated Crop–Livestock Systems. *Agronomy*

- 10, 672–672. 10.3390/agronomy10050672 (2020).
4. Wendel, A. S., Bauke, S. L., Amelung, W. & Knief, C. Root-rhizosphere-soil interactions in biopores. *Plant. Soil.* **475**, 253–277. 10.1007/s11104-022-05406-4 (2022).
5. Cao, Y. et al. The multi-kingdom microbiome of the goat gastrointestinal tract. *Microbiome* **11**, 219–219. 10.1186/s40168-023-01651-6 (2023).
6. Zhou, Y. et al. Microbial communities along the soil-root continuum are determined by root anatomical boundaries, soil properties, and root exudation. *Soil. Biol. Biochem.* **171**, 108721–108721. 10.1016/j.soilbio.2022.108721 (2022).
7. Philippot, L., Chenu, C., Kappler, A., Rillig, M. C. & Fierer, N. The interplay between microbial communities and soil properties. *Nat. Rev. Microbiol.* **22**, 226–239. 10.1038/s41579-023-00980-5 (2023).
8. Fu, X. et al. Critical transition of soil microbial diversity and composition triggered by plant rhizosphere effects. *Front. Plant. Sci.* **14**, 1252821–1252821. 10.3389/fpls.2023.1252821 (2023).
9. Peng, M., Tabashsum, Z., Millner, P. D., Parveen, S. & Biswas, D. Influence of Manure Application on the Soil Bacterial Microbiome in Integrated Crop-Livestock Farms in Maryland. *Microorganisms* **9**, 2586–2586. 10.3390/microorganisms9122586 (2021).
10. Hermans, S. M. et al. From pine to pasture: land use history has long-term impacts on soil bacterial community composition and functional potential. *FEMS Microbiol. Ecol.* **96** 10.1093/femsec/fiaa041 (2020).
11. Xu, Q. et al. Microbial generalists and specialists differently contribute to the community diversity in farmland soils. *J. Adv. Res.* **40**, 17–27 (2021).
12. Neu, A. T., Allen, E. E. & Roy, K. Defining and quantifying the core microbiome: Challenges and prospects. *Proceedings of the National Academy of Sciences* **118**, (2021). 10.1073/pnas.2104429118
13. Drula, E. et al. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* **50**, D571–D577. 10.1093/nar/gkab1045 (2022).
14. Ren, C. et al. Altered microbial CAZyme families indicated dead biomass decomposition following afforestation. *Soil. Biol. Biochem.* **160**, 108362–108362. 10.1016/j.soilbio.2021.108362 (2021).
15. Yu, Z. et al. The CAZyme family regulates the changes in soil organic carbon composition during vegetation restoration in the Mu Us desert. *Geoderma* **452**, 117109–117109. 10.1016/j.geoderma.2024.117109 (2024).
16. López-Mondéjar, R., Tláskal, V., Rocha, U. N. & Baldrian, P. Global Distribution of Carbohydrate Utilization Potential in the Prokaryotic Tree of Life. *mSystems* **7**, e00829-00822, (2022). doi:10.1128/msystems.00829-22
17. Berlemont, R. & Martiny, A. C. Glycoside Hydrolases across Environmental Microbial Communities. *PLoS Comput. Biol.* **12**, e1005300. 10.1371/journal.pcbi.1005300 (2016).

18. Zhao, Z., Baltar, F. & Herndl, G. J. Decoupling between the genetic potential and the metabolic regulation and expression in microbial organic matter cleavage across microbiomes. *Microbiol. Spectr.* **12** 10.1128/spectrum.03036-23 (2024).
19. Akinola, S. A., Ayangbenro, A. S. & Babalola, O. O. The diverse functional genes of maize rhizosphere microbiota assessed using shotgun metagenomics. *J. Sci. Food Agric.* **101**, 3193–3201. 10.1002/jsfa.10948 (2020).
20. Enagbonma, B. J. & Babalola, O. O. Metagenomics Reveals the Microbiome Multifunctionalities of Environmental Importance From Termite Mound Soils. *Bioinform. Biol. insights.* **17** 10.1177/11779322231184025 (2023).
21. Akinola, S. A., Ayangbenro, A. S. & Babalola, O. O. Metagenomic Insight into the Community Structure of Maize-Rhizosphere Bacteria as Predicted by Different Environmental Factors and Their Functioning within Plant Proximity. *Microorganisms* **9**, 1419–1419. 10.3390/microorganisms9071419 (2021).
22. Enagbonma, B. J., Modise, D. M. & Babalola, O. O. Effects of Legume–Cereal Rotation on Sorghum Rhizosphere Microbial Community Structure and Nitrogen-Cycling Functions. *Microbiol. Open.* **14** 10.1002/mbo3.70085 (2025).
23. Adebayo, A. A., Enagbonma, B. J. & Babalola, O. O. Comparative metagenomics on community structure and diversity of rhizomicrobiome associated with monoculture and soybean precedent carrot. *Sci. Rep.* **15**, 28161–28161. 10.1038/s41598-025-13605-z (2025).
24. Sukhum, K. V. et al. Manure Microbial Communities and Resistance Profiles Reconfigure after Transition to Manure Pits and Differ from Those in Fertilized Field Soil. *mBio* **12** (2021).
25. Li, F., Hitch, T. C. A., Chen, Y., Creevey, C. J. & Guan, L. L. Comparative metagenomic and metatranscriptomic analyses reveal the breed effect on the rumen microbiome and its associations with feed efficiency in beef cattle. *Microbiome* **7** 10.1186/s40168-019-0618-5 (2019).
26. Lu, W. et al. Metagenomic sequencing reveals the taxonomic and functional characteristics of rumen microorganisms in Dongliu buffalo. *Sci. Rep.* **15**, 18398–18398. 10.1038/s41598-025-03059-8 (2025).
27. Xu, Q. et al. Microbial generalists and specialists differently contribute to the community diversity in farmland soils. *J. Adv. Res.* **40**, 17–27. <https://doi.org/10.1016/j.jare.2021.12.003> (2022).
28. Mingmongkolchai, S. & Panbangred, W. Bacillusprobiotics: an alternative to antibiotics for livestock production. *J. Appl. Microbiol.* **124**, 1334–1346. 10.1111/jam.13690 (2018).
29. Elshaghabee, F. M. F., Rokana, N., Gulhane, R. D., Sharma, C. & Panwar, H. Bacillus As Potential Probiotics: Status, Concerns, and Future Perspectives. *Front. Microbiol.* **8** 10.3389/fmicb.2017.01490 (2017).
30. Htwe, A. Z., Moh, S. M., Soe, K. M., Moe, K. & Yamakawa, T. Effects of Biofertilizer Produced From Bradyrhizobium and Streptomyces Griseoflavus on Plant Growth, Nodulation, Nitrogen Fixation, Nutrient Uptake, and Seed Yield of Mung Bean, Cowpea, and Soybean. *Agronomy* **9**, 77. 10.3390/agronomy9020077 (2019).

31. Kim, Y. J. et al. Genomic Characterization of a Newly Isolated Rhizobacteria *Sphingomonas panacis* Reveals Plant Growth Promoting Effect to Rice. *Biotechnol. Bioprocess Eng.* **24**, 119–125. 10.1007/s12257-018-0386-2 (2019).
32. Alayande, K. A., Aiyegoro, O. A. & Ateba, C. N. Probiotics in Animal Husbandry: Applicability and Associated Risk Factors. *Sustainability* **12**, 1087–1087 (2020).
33. Arsène, M. M. J. et al. The use of probiotics in animal feeding for safe production and as potential alternatives to antibiotics. *Vet. World.* **14**, 319–328. 10.14202/vetworld.2021.319-328 (2021).
34. Jalal, R. S. & Alshehrei, F. Rhizospheric glycosyltransferase repertoires as a resource for enabling sustainable bioprocessing and green biocatalyst discovery. *Sci. Rep.* 10.1038/s41598-026-42974-2 (2026).
35. Alabbosh, K. F. & Jalal, R. S. Rhizosphere-derived Glycoside Hydrolases: A Biotechnological Treasure Trove from Arid Plant Ecosystems. *BioResources* **21**, (2026). 10.15376/biores.21.1.706-749
36. Sidar, A. et al. Carbohydrate Binding Modules: Diversity of Domain Architecture in Amylases and Cellulases From Filamentous Microorganisms. *Front. Bioeng. Biotechnol.* **8**, 871–871. 10.3389/fbioe.2020.00871 (2020).
37. Zabolina, O. A., Zhang, N. & Weerts, R. Polysaccharide Biosynthesis: Glycosyltransferases and Their Complexes. *Front. Plant. Sci.* **12**, 625307–625307. 10.3389/fpls.2021.625307 (2021).
38. Cortés-Tolalpa, L., Wang, Y., Salles, J. F. & Elsas, J. D. v. Comparative Genome Analysis of the Lignocellulose Degrading Bacteria *Citrobacter freundii* so4 and *Sphingobacterium multivorum* w15. *Front. Microbiol.* **11**, 248–248. 10.3389/fmicb.2020.00248 (2020).
39. Li, N., Zhang, R., Zhou, J., Huang, Z. & Structures Biochemical Characteristics, and Functions of β -Xylosidases. *J. Agric. Food Chem.* **71**, 7961–7976 (2023).
40. Qin, X., Bao, R., Huang, W. & Li, Q. Facilitating the enzymatic hydrolysis of polysaccharides by carbohydrate active enzymes and enhanced humification process with microbial consortium revealed by metagenomics analysis during cow manure-straw composting. *J. Environ. Chem. Eng.* **13**, 115428. <https://doi.org/10.1016/j.jece.2025.115428> (2025).
41. Li, W., Zhang, Y., Mao, W., Wang, C. & Yin, S. Functional potential differences between Firmicutes and Proteobacteria in response to manure amendment in a reclaimed soil. *Can. J. Microbiol.* **66**, 689–697. 10.1139/cjm-2020-0143%M (2020).
42. Santos-Pereira, C. et al. Functional and sequence-based metagenomics to uncover carbohydrate-degrading enzymes from composting samples. *Appl. Microbiol. Biotechnol.* **107**, 5379–5401. 10.1007/s00253-023-12627-9 (2023).
43. Jorge, Á. F., Lóczy, D., Thiele-Bruhn, S. & Raúl, Z. B. *Handbook of plant and soil analysis for agricultural systems* (Universidad Politécnica de Cartagena, 2019).
44. Hunduma, S. & Kebede, G. Measurement of soil particle size distribution using hydrometer analysis. *American-Eurasian J. Agricultural Environ. Sci.* **20**, 243–254 (2020).
45. Walkley, A., Black, I. A., AN EXAMINATION OF THE DEGTJAREFF METHOD FOR DETERMINING SOIL ORGANIC MATTER & AND A PROPOSED MODIFICATION OF THE CHROMIC ACID TITRATION

- METHOD. *Soil Sci.* **37**, 29–38, doi:10.1097/00010694-193401000-00003 (1934).
46. Wilson, M. *Recommended Methods of Manure Analysis*. (2022).
47. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**, (2023). 10.1002/imt2.107
48. Pereira-Marques, J. et al. Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis. *Front. Microbiol.* **10**, 1277–1277. 10.3389/fmicb.2019.01277 (2019).
49. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158> (2006).
50. Nayfach, S. et al. Automated and Accurate Estimation of Gene Family Abundance from Shotgun Metagenomes. *PLoS Comp. Biol.* **11** 10.1371/journal.pcbi.1004573 (2015).
51. Huson, D. H. et al. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comp. Biol.* **12** 10.1371/journal.pcbi.1004957 (2016).
52. Ma, J. et al. Metagenomic Analysis Identifies Sex-Related Cecal Microbial Gene Functions and Bacterial Taxa in the Quail. *Front. Veterinary Sci.* **8**, 693755–693755. 10.3389/fvets.2021.693755 (2021).
53. Zheng, X. et al. Integrated Metagenomic and Metabolomics Profiling Reveals Key Gut Microbiota and Metabolites Associated with Weaning Stress in Piglets. *Genes* **15**, 970–970. 10.3390/genes15080970 (2024).
54. Treiber, M. L., Taft, D. H., Korf, I., Mills, D. A. & Lemay, D. G. Pre- and post-sequencing recommendations for functional annotation of human fecal metagenomes. *BMC Bioinform.* **21**, 74–74. 10.1186/s12859-020-3416-y (2020).
55. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Royal Stat. Soc. Ser. B (Statistical Methodology)*. **57**, 289–300. 10.1111/j.2517-6161.1995.tb02031.x (1995).
56. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x> (2001).
57. Anderson, M. J. Distance-Based Tests for Homogeneity of Multivariate Dispersions. *Biometrics* **62**, 245–253. 10.1111/j.1541-0420.2005.00440.x (2006).
58. Legendre, P. & Gallagher, E. D. Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**, 271–280. 10.1007/s004420100716 (2001).

Figures

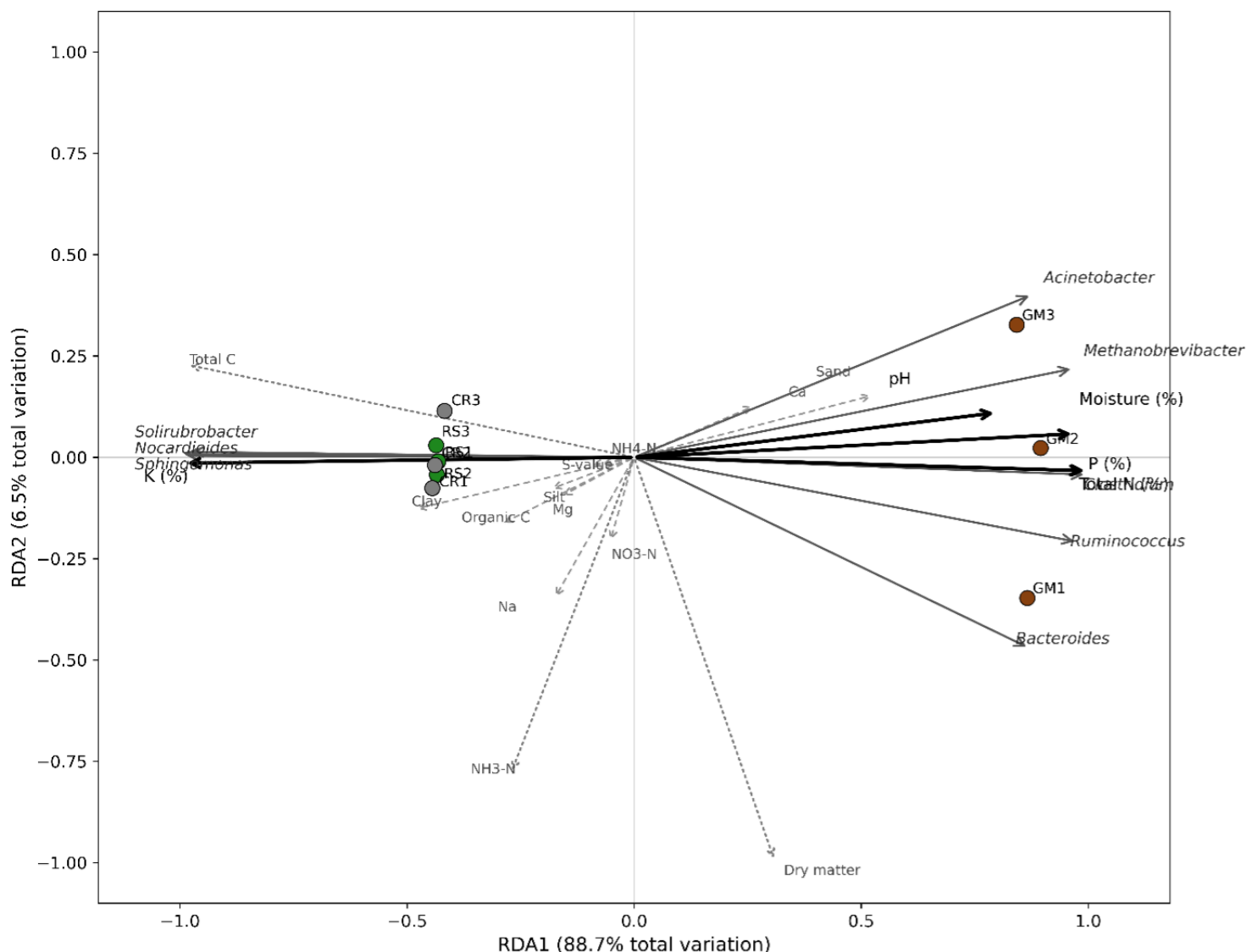


Figure 1

Redundancy analysis of genus-level microbial community composition across goat manure (GM), maize rhizosphere soil (RS) and control soil (CR). The ordination was based on Hellinger-transformed resolved genus-level relative abundances, with pH, moisture, total N, P and K fitted jointly as harmonised active constraints. Soil-specific physicochemical variables (dashed vectors) and manure-specific variables (dotted vectors) were passively projected and were not included as active constraints; representative genus vectors are shown for ecological context. The full model explained 98.9% of total community variation (adjusted $R^2 = 0.971$; pseudo-F = 54.52; exact permutation p = 0.00035), with RDA1 and RDA2 accounting for 88.7% and 6.5% of total variation.

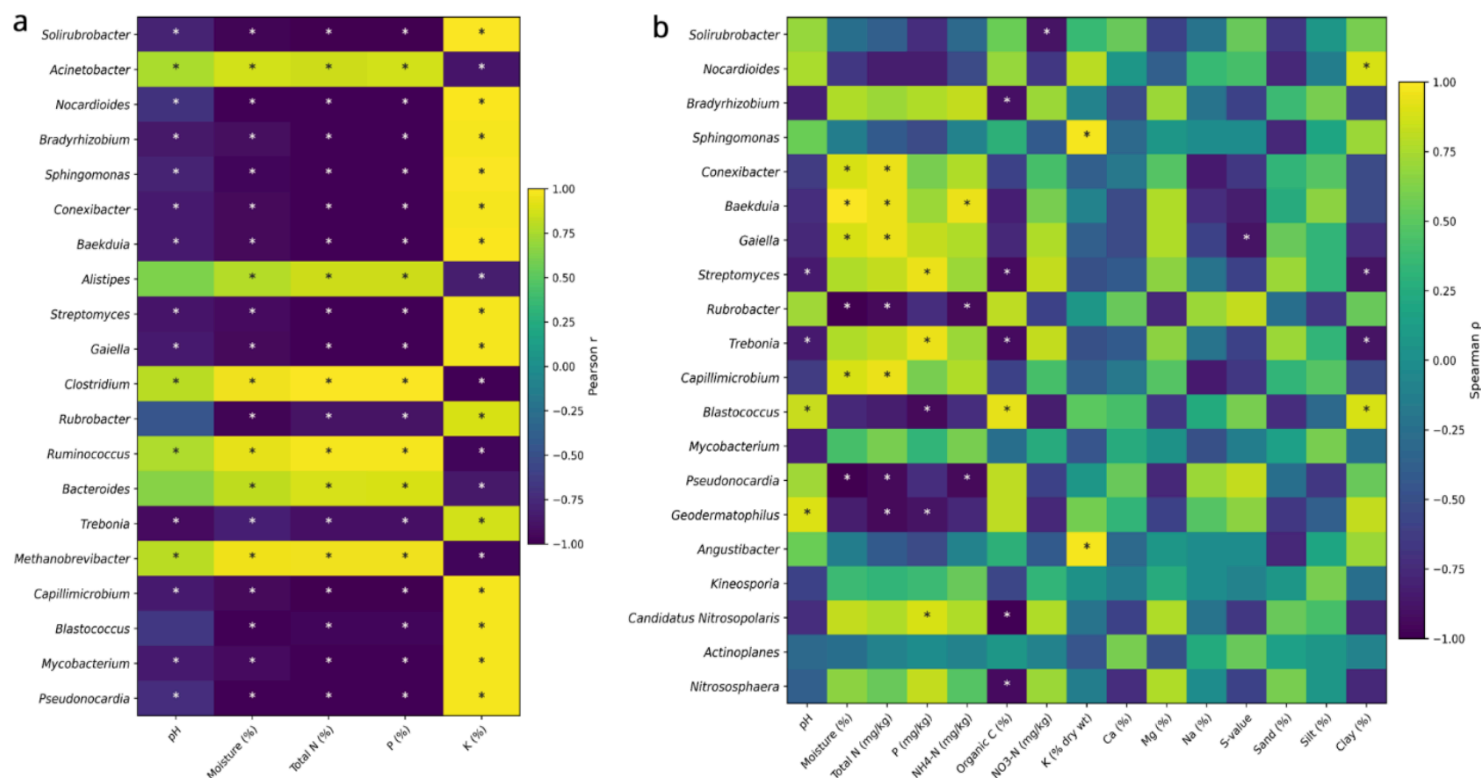


Figure 2

Taxon-environment correlations across the manure-soil system. (a) Pearson correlations between the 20 most abundant resolved genera and five harmonised physicochemical variables across all nine GM, RS and CR composites. Cell values are Pearson's r ; asterisks denote associations significant after Benjamini-Hochberg correction in the unrestricted analysis. Although 96 of 100 associations were FDR-significant, none remained FDR-significant when permutations were restricted within compartments. (b) Spearman correlations between the 20 dominant genera and 15 soil physicochemical variables across the six RS and CR composites. Cell values are Spearman's ρ ; asterisks denote nominal exact $p < 0.05$.

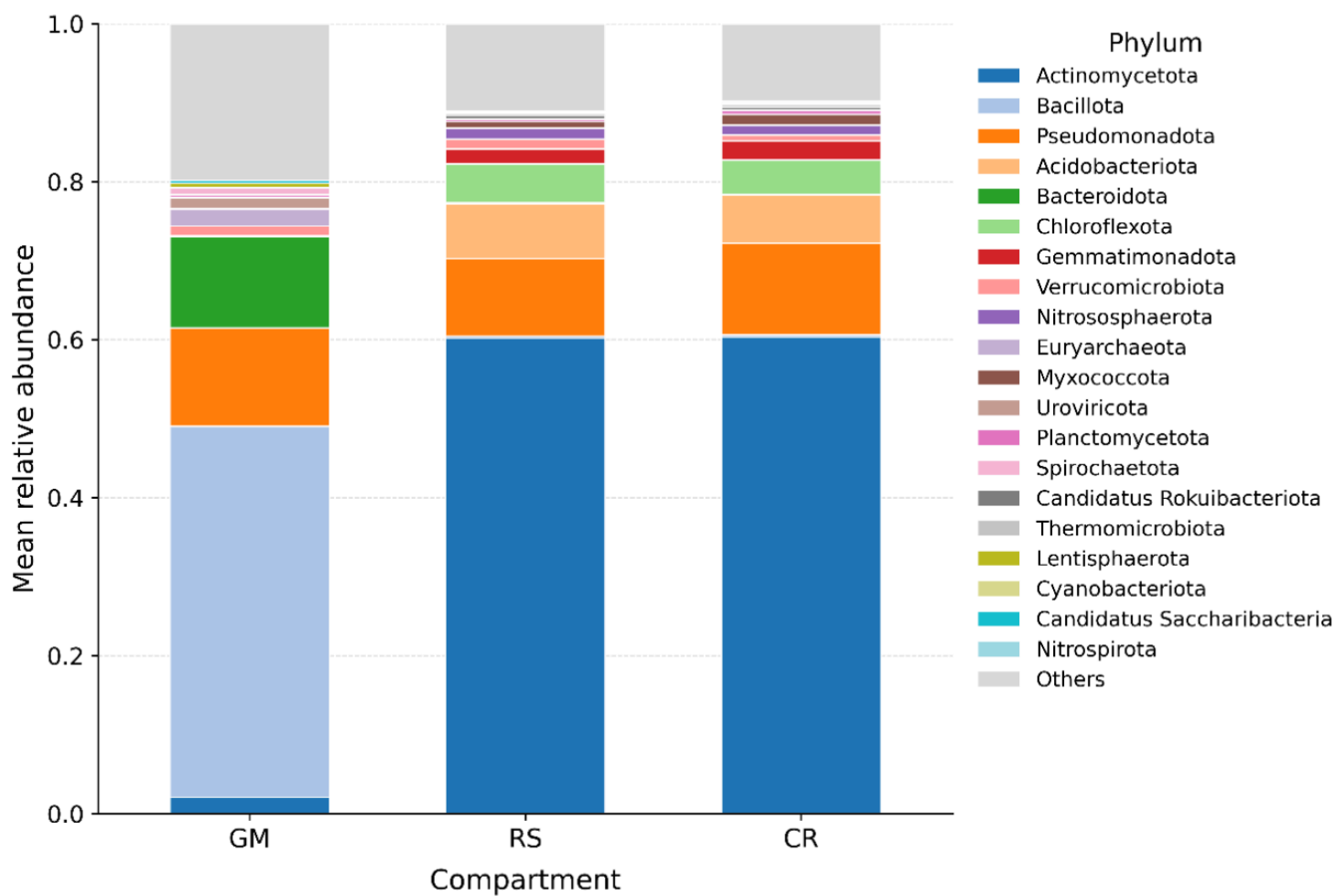


Figure 3

Mean phylum-level composition of microbial communities in goat manure (GM), maize rhizosphere soil (RS) and control soil (CR). The 20 most abundant resolved phyla are shown.

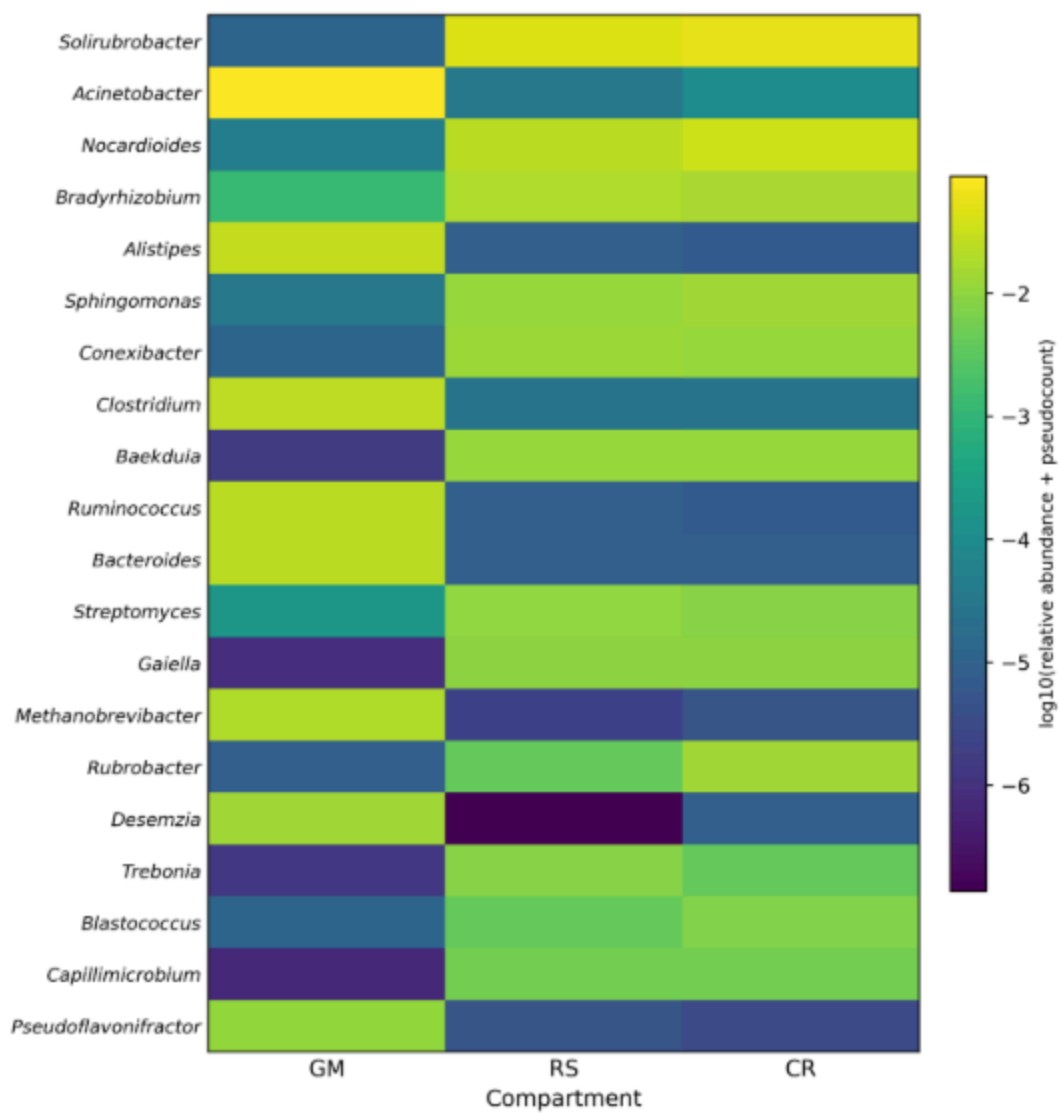


Figure 4

Heatmap of the 20 most abundant resolved genera across goat manure (GM), maize rhizosphere soil (RS) and control soil (CR). Values represent log10-transformed mean relative abundance after addition of a pseudocount.

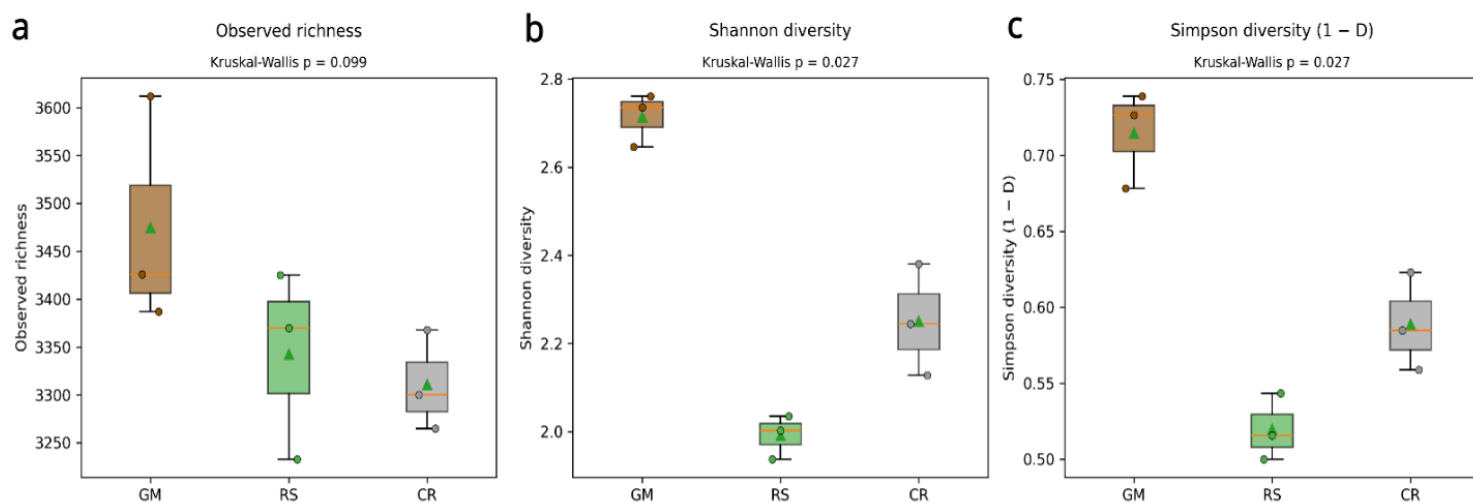


Figure 5

Genus-level alpha diversity across goat manure (GM), maize rhizosphere soil (RS) and control soil (CR). Boxplots show (a) observed richness, (b) Shannon diversity and (c) Simpson diversity ($1 - D$).

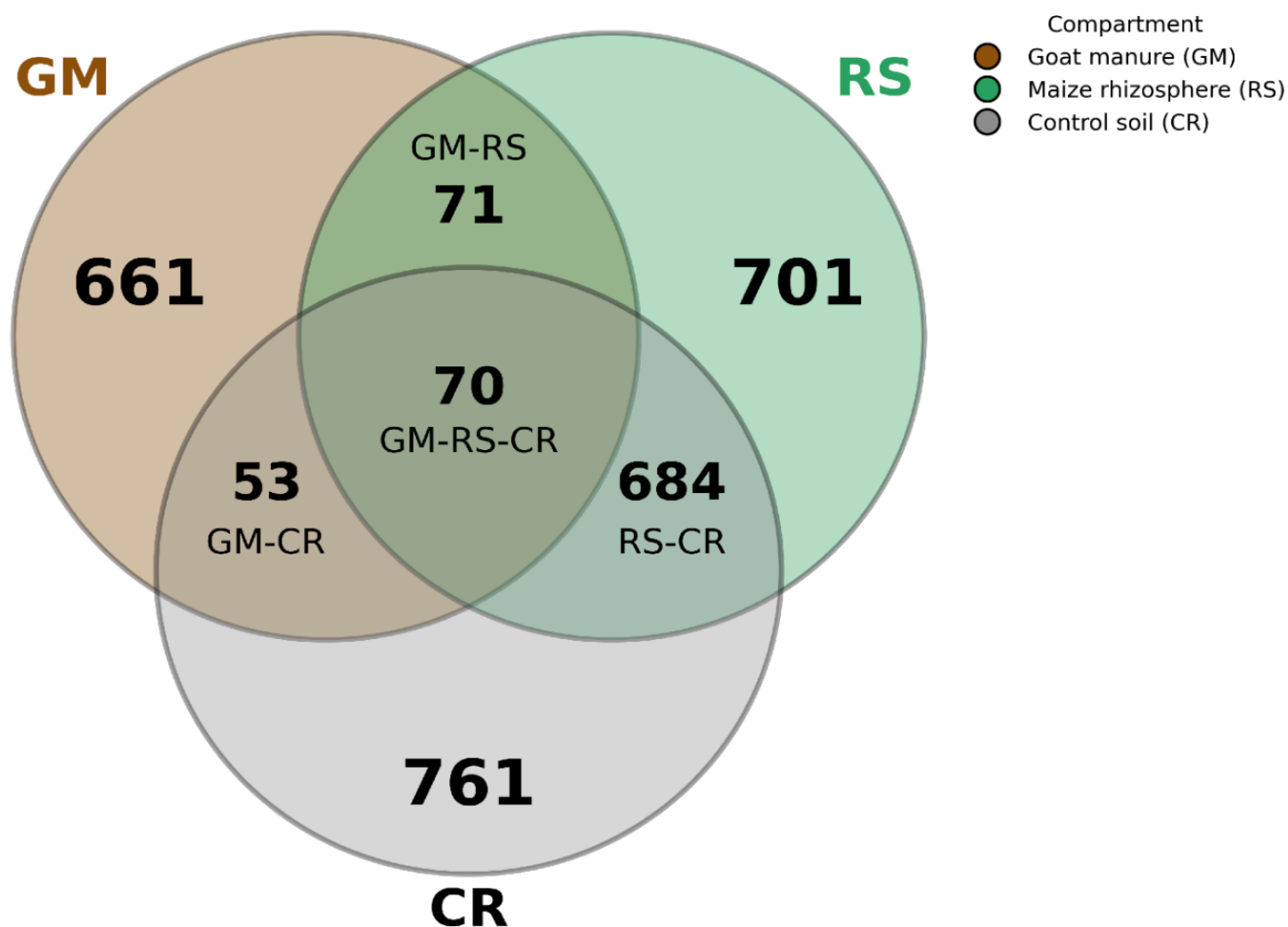


Figure 6

Shared prevalent genera across goat manure (GM), maize rhizosphere soil (RS) and control soil (CR). Features were considered prevalent when detected at relative abundance $\geq 1 \times 10^{-5}$ in all three samples within a compartment. Values are mutually exclusive Venn regions; the central value represents genera shared across all compartments.

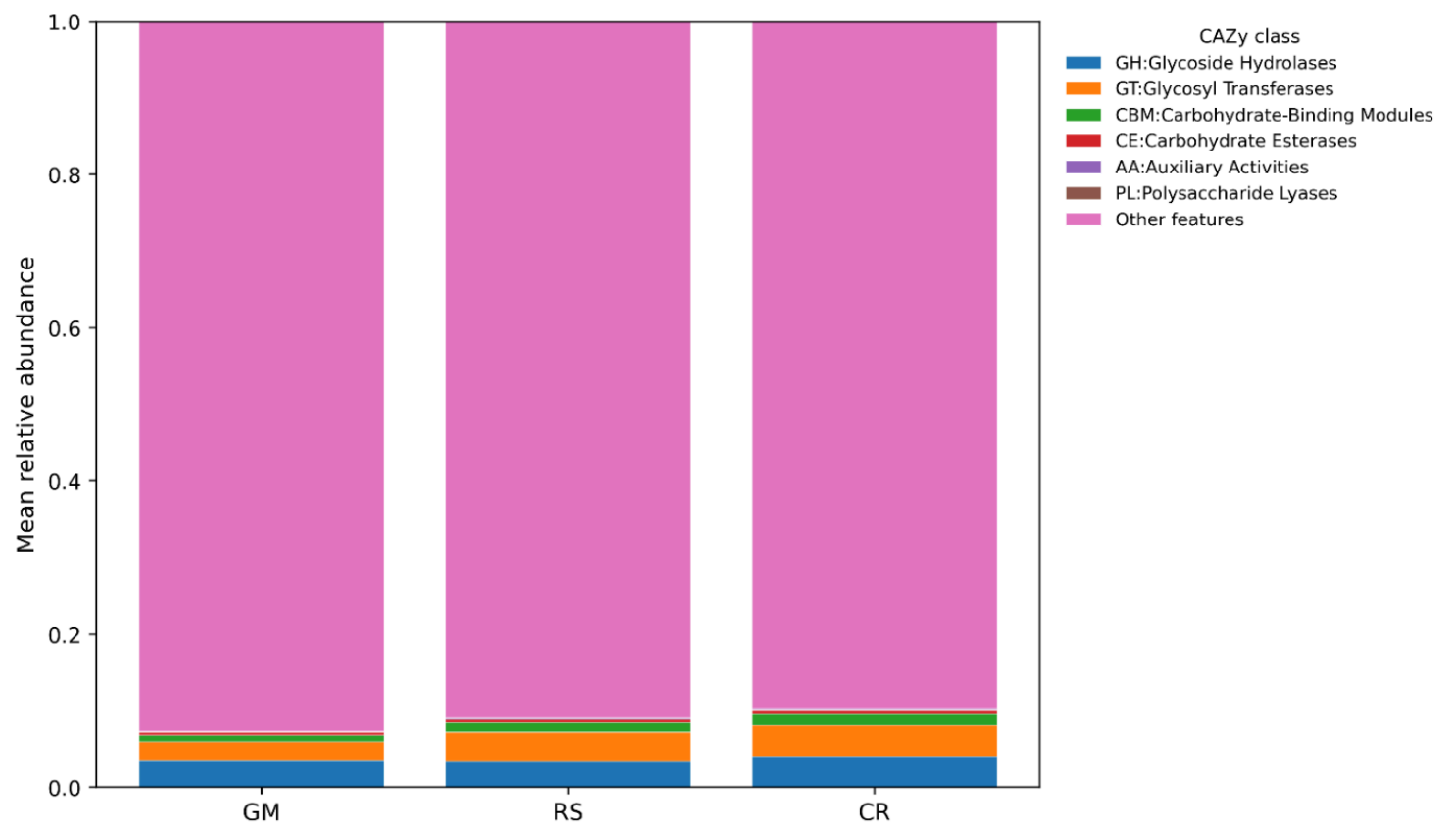


Figure 7

Mean CAZy level 1 composition in goat manure (GM), maize rhizosphere soil (RS) and control soil (CR). Bars show the six resolved CAZy classes and the residual unresolved fraction.

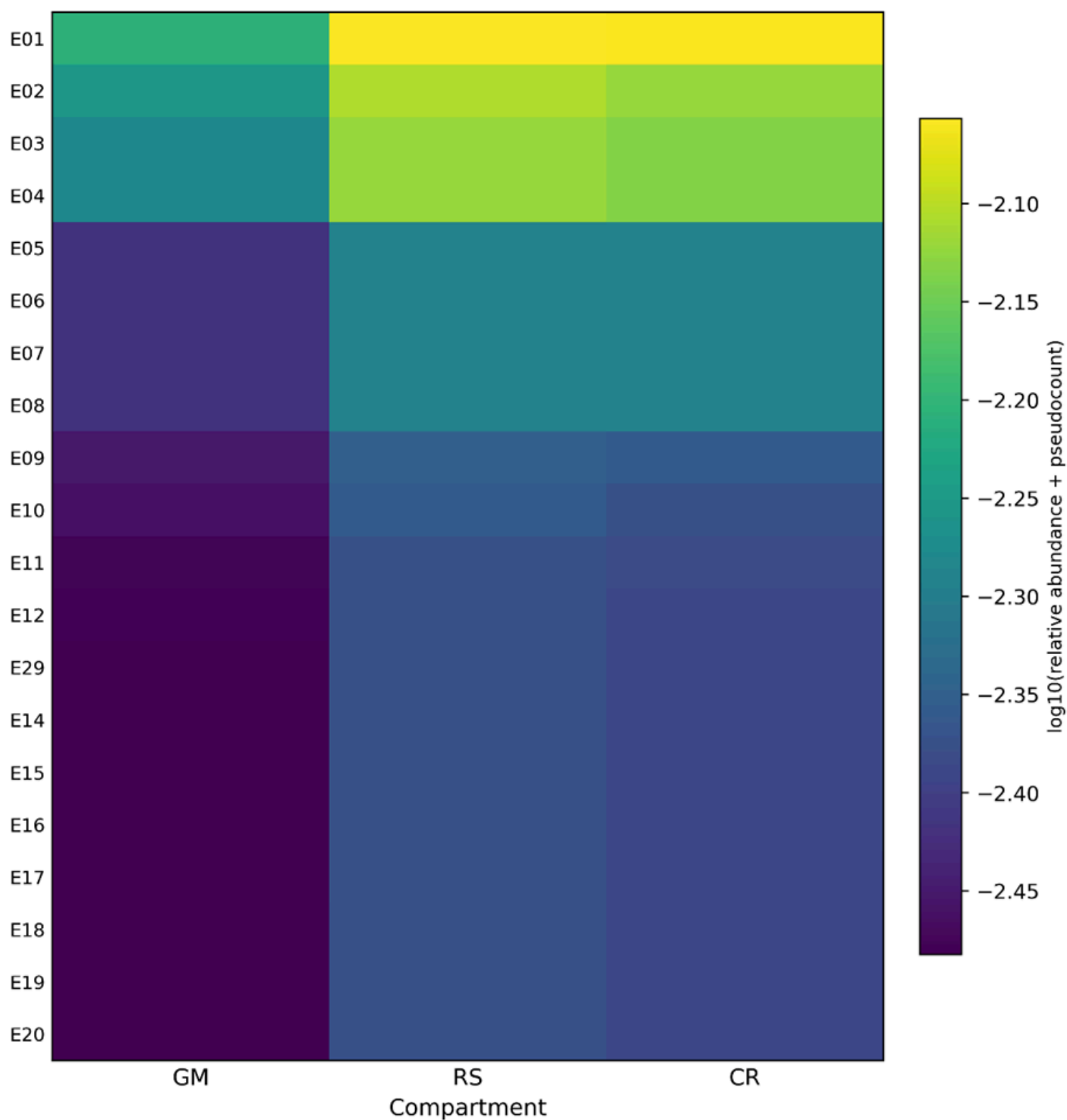


Figure 8

Heatmap of the 20 most abundant resolved EC-level CAZy features across goat manure (GM), maize rhizosphere soil (RS) and control soil (CR). Values represent log₁₀-transformed mean relative abundance after addition of a pseudocount. Feature codes are define in Supplementary Table S5.

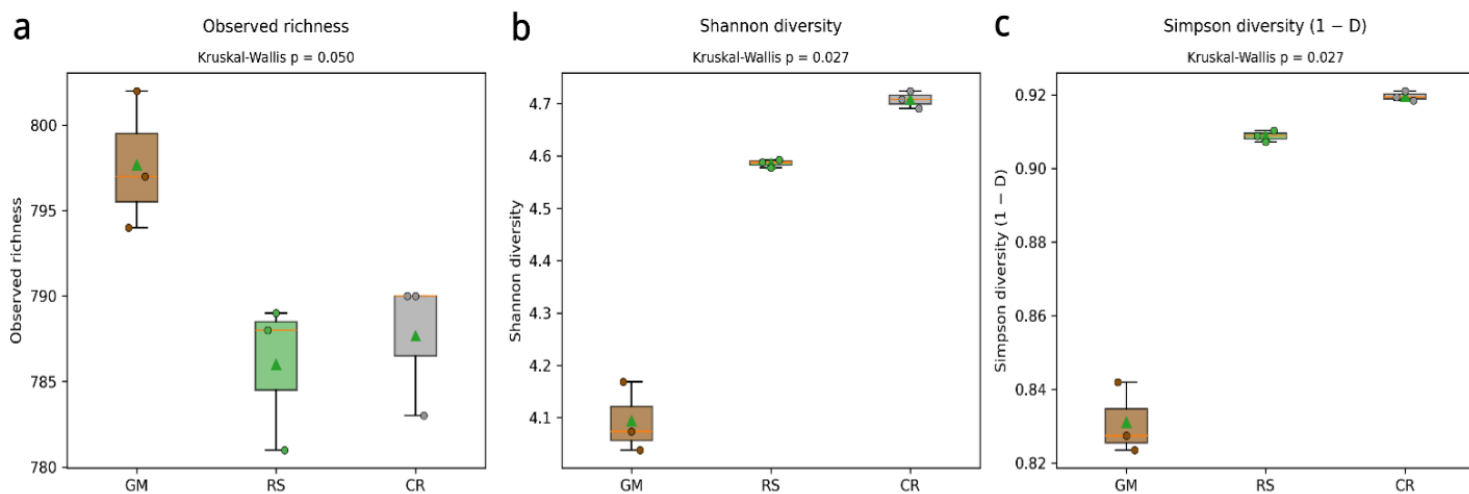


Figure 9

Alpha diversity of EC-resolved CAZy profiles across goat manure (GM), maize rhizosphere soil (RS), and control soil (CR). Boxplots show (a) observed richness, (b) Shannon diversity, and (c) Simpson diversity ($1 - D$). Points represent composite samples, and Kruskal-Wallis p-values are shown above each panel.

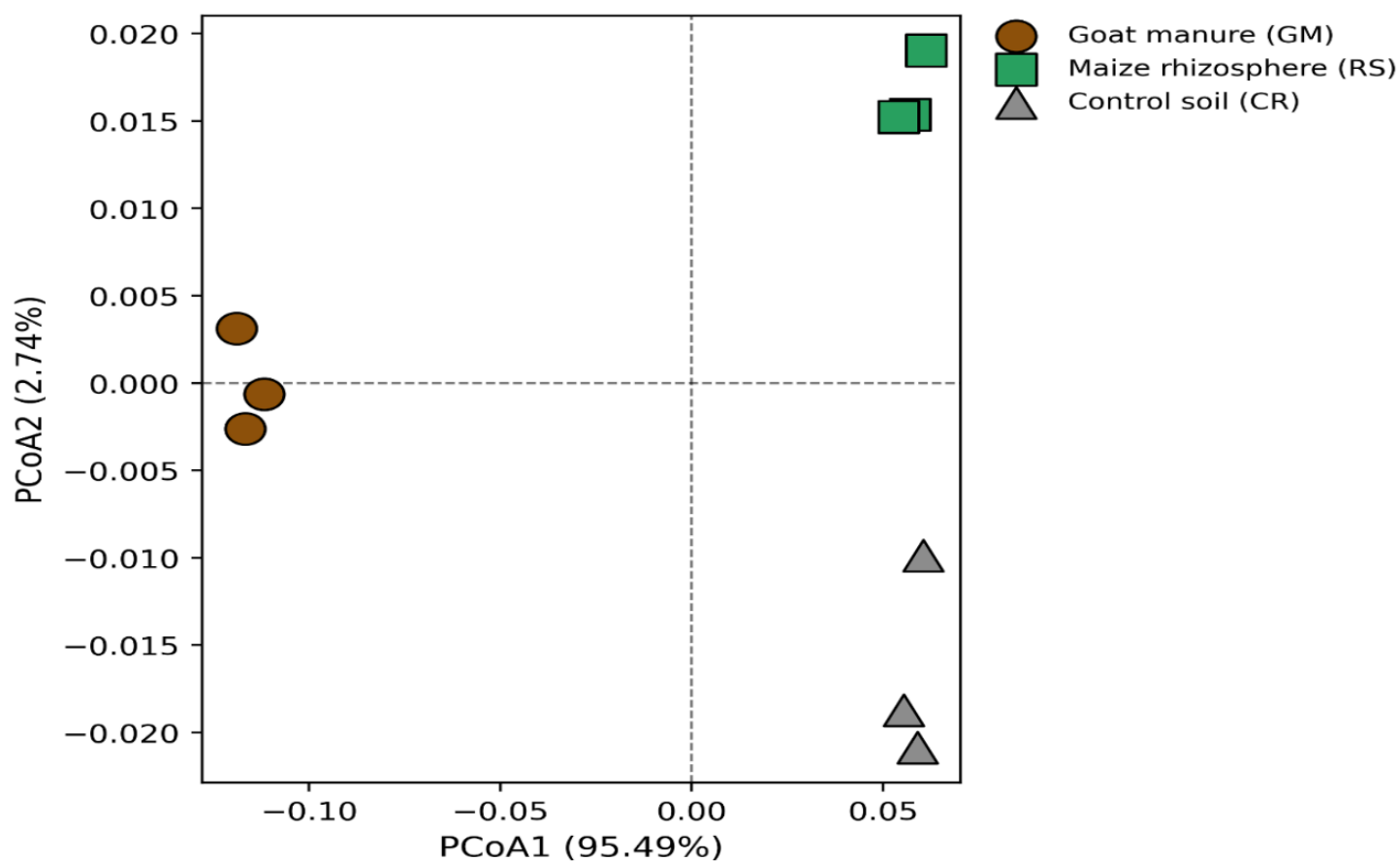


Figure 10

Bray-Curtis principal coordinate analysis of EC-resolved CAZy composition in goat manure (GM), maize rhizosphere soil (RS), and control soil (CR). Each point represents one composite sample; axis labels show the percentage of variation explained.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ThebeetalScientificReportsSupplementaryMaterial.docx](#)