Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatially variable barley root microbiomes do not imply weak host genetic control at field scale

Menghui Dong
Department of Agroecology, Aarhus University, 4200 Slagelse, Denmark https://orcid.org/0000-0002-9487-7318

Katarzyna Retzer
Department of Ecosystem Management, Climate and Biodiversity, Institute of Forest Ecology, University of Natural Resources and Life Sciences (BOKU), 1190 Vienna, Austria

Heinrich Grausgruber
Department of Agricultural Sciences, Institute of Crop Breeding and Genomics, University of Natural Resources and Life Sciences (BOKU), 3430 Tulln an der Donau, Austria https://orcid.org/0000-0003-4790-0922

Aditya Sanjay Malwe
Center for Evolutionary Hologenomics, Globe Institute, University of Copenhagen, 1353 Copenhagen K, Copenhagen, Denmark

Miguel Sanchez Garcia
Biodiversity and Crop Improvement, International Center for Agricultural Research in the Dry Areas (ICARDA), 10112 Rabat, Morocco

Andrea Visioni
Biodiversity and Crop Improvement, International Center for Agricultural Research in the Dry Areas (ICARDA), 10112 Rabat, Morocco

Karl Ludwig Körber
Department of Agroecology, Aarhus University, 4200 Slagelse, Denmark

Anna Elisabeth Backhaus
Biodiversity and Crop Improvement, International Center for Agricultural Research in the Dry Areas (ICARDA), 10112 Rabat, Morocco

Kerstin Rönsch
Eurofins Genomics Food/Environment/White Biotech Products & Services GmbH, 85560 Ebersberg, Germany

Markus Schmitt
Eurofins Genomics Food/Environment/White Biotech Products & Services GmbH, 85560 Ebersberg, Germany

Christopher James Barnes
c.barnes@sund.ku.dk

Department of Agroecology, Aarhus University, 4200 Slagelse, Denmark; Center for Evolutionary Hologenomics, Globe Institute, University of Copenhagen, 1353 Copenhagen K, Copenhagen, Denmark

Mette Vestergård

mvestergard@agro.au.dk

Department of Agroecology, Aarhus University, 4200 Slagelse, Denmark https://orcid.org/0000-0002-0054-4855

Mogens Nicolaisen

mn@agro.au.dk

Department of Agroecology, Aarhus University, 4200 Slagelse, Denmark https://orcid.org/0000-0002-0407-2488

Article

Keywords: plant microbiome, host genetics, soil spatial heterogeneity, context-dependent microbiome assembly, microbial GWAS, barley

Posted Date: October 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11145675/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

The plant microbiome has been proposed as ‘the second genome of the plant’, motivating interest in microbiome-assisted breeding. For microbiome-assisted breeding to become a reality, the plant genome must obviously have a role in regulation of the microbiome. However, agricultural fields are spatially heterogeneous, and it remains unclear how host genetics and soil heterogeneity jointly shape plant microbiomes under field conditions. Here, using 16S and ITS metabarcoding data from 2,400 root representing 386 barley lines across spring and winter field trials, we quantified host genetic effects on root-associated fungal and bacterial communities and their modulation by soil spatial heterogeneity. Barley genotype exerted a detectable and reproducible effect on overall microbial community composition and on a subset of bacterial and fungal biomarker taxa associated with plant biomass. However, these host genetic effects were strongly modulated by field-scale spatial heterogeneity. Spatial context influenced both whole-community structure and the abundance patterns of individual host-associated biomarkers. These biomarkers differed markedly in their occupancy consistency, spatial dependency, and host Single-Nucleotide Polymorphism (SNP)-by-soil spatial structure interaction patterns, supporting the view that host-associated microbiome recruitment is context dependent rather than a fixed filtering across field environments. Consistent with this, variation of SNP-associated positive biomarkers significantly correlated with barley biomass stability, but these associations were stronger at the level of total positive biomarker abundance than at the level of individuals. Together, our findings demonstrate that large-scale field experiments are essential for dissecting host-microbiome assembly patterns and reveal that host genetic control of the root microbiome in the field is detectable but strongly context dependent, with important implications for microbiome-assisted breeding.

Introduction

Host genetics is widely recognized as important determinants of microbiome assembly through resource allocation, immunity, and habitat modification[1]. Across plants and animals, hosts assemble microbiomes that differ between host species and genotypes[2,3]. However, microbiome assembly occurs within heterogeneous environments, where the effects of biotic and abiotic conditions can equal or exceed host effects[4]. How host-mediated recruitment persists under such heterogeneity remains unresolved, raising a central question in microbial ecology: to which extent can the host shape associated microbial communities despite substantial environmental heterogeneity?

This question is highly relevant in agricultural systems, where manipulating plant-associated microbiomes has emerged as a promising route to improve crop performance and resilience[5]. Beneficial microbiomes can contribute to nutrient acquisition[6], stress resilience[7], and disease suppression[8], which is reflected in the increasing interest in microbiome-assisted breeding strategies to harness the benefits of microbiomes[9]. Recent estimates suggest that approximately 90% of breeding efforts have historically overlooked environmental heterogeneity[10]. While this simplification has been sufficient for improving many host traits, it may represent a fundamental limitation for microbiome-assisted breeding, as

microbial assembly is highly responsive to local soil conditions and may challenge the stability of host-mediated recruitment.

Addressing this question requires large-scale comparisons among host genotypes under actual field conditions, yet such studies remain scarce owing to the difficulty of combining extensive plant lines containing known genotypic variation with high-throughput microbiome profiling, across replicated, spatially heterogenous plots. Existing efforts in crops, including maize[11], millet[12], and potato[13,14], have primarily focused on microbiome heritability or identification of host-associated microbial taxa, providing evidence that host genetics can influence microbiome composition. However, spatial heterogeneity is commonly treated as background variation rather than an explicit component of recruitment processes. In particular, the consequences of field-scale heterogeneity within agricultural fields, where neighbouring plants may experience distinct microbial source pools and soil conditions, remain poorly understood. Consequently, a key unresolved question is whether host control over microbiome assembly functions primarily acts as a fixed filter to soil microbes that are subject to soil heterogeneity, or as a dynamic context-dependent recruitment strategy capable of assembling alternative microbial communities across neighbouring microsites. Spatial heterogeneity may therefore be more than noise that masks host genetic effects: it may modulate these effects on microbiome recruitment, making host control itself dependent on local environmental context.

Here, we investigated how host genetics shapes root-associated microbiomes under actual agricultural field conditions, while explicitly accounting for field-scale spatial heterogeneity. Specifically, we asked to what extent host genetic effects on microbiome assembly persist across neighbouring microsites that differ in soil conditions. Using a large-scale field experiment comprising approximately 2,400 bacterial and fungal root microbiome profiles from 386 barley lines and 120 bulk soil samples collected across spring and winter field trials, we integrated spatial variation, host genetic data and microbiome profiling to quantify the relative contributions of host genetics, spatial heterogeneity and their interaction to root microbiome assembly. By combining evidence from dissimilarities between root and bulk soil communities, within-genotype beta-dispersion, and SNP-by-soil spatial structure interactions involving barley biomass and SNP-associated taxa, we further tested whether host-mediated microbiome assembly is best explained as a fixed filtering process susceptible to spatial heterogeneity, or as a context-dependent recruitment strategy through which plants assemble ecologically distinct microbial communities under different local soil conditions.

Results

Microbial community composition across panels and genotypes

To investigate how host genetics and field-scale spatial heterogeneity jointly shape root-associated microbiomes, we analysed bacterial and fungal metabarcoding data from two large barley field experiments conducted in neighbouring agricultural fields in Austria under contrasting seasonal

cultivation regimes. The Spring Panel (SpringPan) field primarily consisted of spring barley lines, and the Winter Panel (WinterPan) field primarily consisted of winter barley lines. Each field included 200 genetically distinct barley lines, with 14 shared spring lines cultivated in both experiments. The genotypes were grown in three replicated plots, and each plot was further divided into two subsamples, collected at the north and south end of the plot, respectively, resulting in six microbiome samples per genotype for downstream analysis. Root-associated microbiomes were sampled at tillering stage. Detailed field layouts, environmental characteristics and climatic conditions are provided in Supplementary Materials Fig. S1.

Principal coordinate analysis (PCoA) on microbial community composition of all samples from SpringPan and WinterPan showed two clear clusters based on panel (Fig. 1a, b), and PERMANOVA confirmed that there were significant differences between SpringPan and WinterPan for both bacterial ($R^2 = 0.147$, $p < 0.001$) and fungal ($R^2 = 0.201$, $p < 0.001$) communities. PCoA also showed that both bacterial and fungal communities had a larger separation along the PCo1 in SpringPan than in WinterPan, indicating a larger variation in microbial composition in the SpringPan across samples.

The relative abundances of the top 15 bacterial and fungal families are shown in Supplementary Materials Fig. S2 and S3. Several abundant bacterial families, including Nocardioidaceae and Sphingomonadaceae, and fungal families such as Nectriaceae and Cladosporiaceae, were consistently detected in both SpringPan and WinterPan, indicating a shared core taxonomic structure across field trials. Nevertheless, both bacterial and fungal communities also contained families that were enriched in either SpringPan or WinterPan, consistent with taxonomic turnover between the spring and winter panels.

It should be noted that SpringPan and WinterPan were sown and sampled at different time points and therefore also differ in environmental conditions in addition to the differences in genotype (winter and spring types). Therefore, it was impossible to disentangle whether observed microbial differences were driven primarily by environment or by host genetics. To further quantify the relative effects of genotype and panel (season), we used a set of 14 barley lines which were grown in both panels (‘check lines’), allowing us to estimate the effect of genotype, panel (season), and their interactions. PCoA and PERMANOVA performed on the check lines (Fig. 1c, d) revealed a significant effect of both genotype, season, and genotype × season interaction (all $p < 0.001$). The season effect, representing environmental differences such as temperature, time since sowing and small-scale soil heterogeneity, explained the largest proportion of variation in both bacterial ($R^2 = 0.175$) and fungal communities ($R^2 = 0.247$). Genotype also contributed substantially, accounting for 8.7% of bacterial and 9.2% of fungal variation. Similar to the analysis of all samples, the check lines showed larger variation in the SpringPan than in the WinterPan both in bacterial and fungal communities. Given the significant genotype × season interaction, SpringPan and WinterPan were analysed separately in all downstream analyses.

Barley genotypic effect on microbial composition

Host genotype significantly explained a substantial part of the variation in microbial community composition in both panels, but with different effect sizes (Fig. 2a). The SpringPan showed a stronger genotype effect in both bacterial and fungal communities (Bacteria-R^2 = 0.401, $p < 0.001$; Fungi-R^2 = 0.426, $p < 0.001$) than the WinterPan (Bacteria-R^2 = 0.201, $p < 0.05$; Fungi-R^2 = 0.242, $p < 0.001$).

To further validate the relationship between barley genetic variation and microbial community composition, we calculated genetic distances among barley genotypes using Prevosti's genetic distance based on the 50K Single Nucleotide Polymorphisms (SNPs) data of all barley genotypes. PCoA based on Prevosti's distance demonstrated the overall structure of barley genotypes involved in this study (Supplementary Materials Fig. S4). Consistent with the PERMANOVA results, microbial dissimilarities correlated significantly with barley genetic distance in the SpringPan for both bacteria (r = 0.0963, p = 0.0011) and fungi (r = 0.1139, p = 0.002), according to the Mantel correlation test (Supplementary Materials Fig. S5). In contrast, no significant Mantel correlation was detected in WinterPan for either bacteria (p = 0.32) or fungi (p = 0.85), despite the significant genotype effect observed in the PERMANOVA.

Spatial heterogeneity modulated genotype effect

PERMANOVA also revealed that the plot position in the field trials (row and column) significantly impacted bacterial and fungal community composition in both SpringPan and WinterPan, although with relatively small effect sizes (all $p < 0.001$, R^2 ranged from 0.004–0.011), suggesting a potential effect of small-scale spatial heterogeneity within the fields (Fig. 2a). Moreover, when the interaction between genotype and plot position (row or column) were considered, significant (all $p < 0.001$) and large effect sizes were observed for both bacterial and fungal communities in both panels. Specifically, interaction terms of Genotype×Row and Genotype×Col together explained 40.9% of the bacterial community variation in SpringPan (Genotype×Row-R^2 = 0.208, Genotype×Col-R^2 = 0.201), slightly higher than the 36.9% explained in WinterPan (Genotype×Row-R^2 = 0.178, Genotype×Col-R^2 = 0.191). For fungi, the interaction terms explained an even greater proportion of community variation. In the SpringPan, 41.8% of the fungal variation was explained by genotype–spatial interactions (Genotype×Row, R^2 = 0.207; Genotype×Column, R^2 = 0.211), compared with 40.8% in the WinterPan (Genotype×Row-R^2 = 0.201, Genotype×Col-R^2 = 0.207). As a result, 59.1–85.5% of the total variation in bacterial and fungal communities could be explained by genotype, soil spatial effects, and their interactions.

However, PERMANOVA was based on categorical genotype identities and raw spatial coordinates, which do not capture continuous host-genetic relatedness or multi-scale spatial structure. To distinguish continuous host-genetic effects from spatially structured environmental variation, we next decomposed microbiome variation into host-genetic and soil spatial structure components. Because simple plot coordinates (plot row and column position) describe only linear spatial gradients, spatial structure was represented using principal coordinates of neighbour matrices (PCNMs), which summarize multi-scale spatial patterning. Meanwhile, host genetics was represented using PCoA axes derived from SNP-based

Prevosti's genetic distances. Plot level average abundance data of microbial communities were used in the analysis, because subplot position explained very little variation in either bacterial or fungal community composition (R^2 = 0.0013–0.0071 across panels; Supplementary Materials Table S1), whereas plot-scale row and column effects were consistently larger (combined R^2 = 0.011–0.088; Fig. 2a). Moreover, several leading PCNM axes were significantly correlated with measured soil properties in the bulk soil of 60 plots across the field (see in Supplementary Materials Table S2) in each panel, including total carbon, total nitrogen, cation exchange capacity, base saturation, pH and multiple mineral elements (Supplementary Materials Fig. S6a), indicating that they captured biologically meaningful soil heterogeneity.

In both panels, spatial structure explained more microbiome variation than host genetics. In forward-selected dbRDA models, cumulative adjusted R^2 increased steeply with additional PCNM axes, but much weaker with SNP-PCoA axes, particularly in WinterPan (Supplementary Materials Fig. S6b). Variation partitioning analysis confirmed this contrast (Fig. 2b). In bacterial communities in SpringPan, spatial structure independently explained 5.5% of the variation, compared with 3.5% for host genetic structure, with only 0.6% jointly explained. In WinterPan, the independent host-genetic fraction explained 0.4%, whereas the spatial fraction explained 4.5% (Fig. 2b). Fungal communities showed the same overall pattern, with spatial fractions consistently exceeding host-genetic fractions and total explained variation generally higher than in bacteria. Together, these results indicate that spatially structured environmental heterogeneity was a stronger determinant of microbiome composition than host genetic relatedness alone.

Because the strong genotype-by-space interaction (Fig. 2a) suggested that host genetic effects were spatially contingent, we tested whether the relationship between host genetic divergence and microbiome differentiation varied across geographic distance (raw XY coordinates) or spatial context represented by PCNMs (Fig. 2c). Several PCNM axes showed significant interaction terms in both bacterial and fungal communities, indicating that host genetic effects varied across spatial contexts rather than simply declining with distance. As expected, geographic distance based on raw XY coordinates showed mostly negative interaction effects, suggesting weaker host genetic effects with increasing plot separation. By contrast, PCNM-based interactions were both positive and negative, indicating that different components of spatially structured soil heterogeneity could either weaken or strengthen host genetic effects on microbiome composition. These patterns were further visualized by stratified regressions across near, mid and far groups for XY distance and the top 10 forward-selected PCNM axes (Supplementary Materials Fig. S7).

Host selection shaped root microbiome in a context-dependent manner

To assess the soil context of host-associated root microbiome assembly, we examined the relationships between root microbial communities, soil physicochemical properties and bulk soil community

composition across 60 plots across the length and width of the fields (Fig. 3a). In SpringPan, calcium (Ca), magnesium (Mg), cation exchange capacity (CEC) and base saturation (BS%) were associated with bulk soil bacterial and/or fungal communities, whereas Ca, CEC and carbon content (C%) were associated with root-associated bacterial communities but not with fungal communities. In WinterPan, fewer associations between soil properties and microbial communities were detected in either bulk or root samples: bulk soil bacterial communities and root-associated fungal communities were significantly correlated with manganese (Mn) and nitrogen content (N%). No significant correlations were detected between bulk soil and root-associated community composition for either bacteria or fungi, indicating that root microbiome assembly was not simply a passive reflection of the surrounding soil microbial pool.

We next asked whether stronger host selection led to microbiome convergence within genotypes or instead promoted context-dependent divergence. We quantified host selection strength as the Bray-Curtis dissimilarity between each root microbiome sample and its six nearest bulk soil communities, and we quantified within-genotype divergence as distance to the genotype centroid (beta-dispersion). Supplementary Materials Fig. S8 illustrates both distance metrics using a representative example. Across bacterial communities, the relationship between host selection strength and distance to genotype centroid was weak, becoming significant only in WinterPan (Fig. 3b). By contrast, fungal communities showed consistently positive associations in both field trials, indicating that stronger host selection was associated with greater, rather than lower, within-genotype dispersion. Together, these results are consistent with host selection operating in a context-dependent manner, rather than through simple filtering of the surrounding soil community.

Barley biomass and its correlation with microbial communities

The aboveground biomass of the barley lines was recorded at the same time as the root sampling for microbial community analysis (Supplementary Materials Fig. S10). In the SpringPan, aboveground biomass ranged from 7.53 to 98.68 g per five plants, whereas in the WinterPan it ranged from 11.06 to 44.38 g. Aboveground biomass varied significantly between barley genotypes in both panels (SpringPan: ANOVA, $F = 1.748$, $p < 0.001$; WinterPan: ANOVA, $F = 1.930$, $p < 0.001$). Barley biomass also correlated significantly with both bacterial and fungal communities in SpringPan and WinterPan, supported by significant correlation with PCoA axes (all $p < 0.01$) and PERMANOVA (all $p < 0.001$) (Supplementary Materials Fig. S11).

Microbial biomarkers of biomass and their association with SNPs

We employed multiple complementary approaches to identify microbial biomarkers associated with barley biomass, ensuring the robustness of biomarker selection. Boruta, a wrapper algorithm for all relevant feature selection models, was applied to identify microbial candidates. As a result, 35 and 33

bacterial candidates and 52 and 28 fungal candidates were identified from the SpringPan and WinterPan, respectively (Fig. 4a, d). Next, an extreme gradient boosting (XGBoost) model was implemented and interpreted using SHAP values to further validate the candidate microbial biomarkers and quantify their relative contributions to barley biomass. Models built using microbial features pre-selected by Boruta showed moderate predictive performance, with test-set R^2 values of 0.35 and 0.39 for bacterial communities in the SpringPan and WinterPan, respectively, and 0.28 and 0.40 for fungal communities in the SpringPan and WinterPan. Because differences in barley biomass among genotypes are partly driven by direct host genetic effects, in addition to indirect effects mediated through the microbiome, we further applied linear mixed-effects (LME) models, incorporating barley genotype and plot position as random effects, to control for direct genetic and spatial confounding effects. This approach minimizes spurious associations and avoids falsely attributing genotype-driven biomass variation to microbial effects. Through this integrative framework, we identified 11 and 3 bacterial biomarkers and 25 and 12 fungal biomarkers with consistent significance and directionality (positive or negative effects) in SpringPan and WinterPan, respectively (Fig. 4a and d).

Finally, we performed microbiome-wide association analysis (MGWAS), using a significance threshold of $pw < 5 \times 10^{-5}$, to test whether the microbial biomarkers, previously identified as positively or negatively correlated with barley biomass, were associated with barley 50K SNP variants. The relative importance estimated by SHAP and potential contribution indicated by LME of host genetic-associated biomarkers are shown in Fig. 4b and e. A full list of results of feature selection and MGWAS can be found in Extended Data Table 1.

We detected 4 positive bacterial biomarkers in SpringPan, including ASVs identified as *Aeromicrobium, Devosia, Pseudomona*s, and *Dongia*, but no biomarker in WinterPan that had significant associations to SNP loci (Fig. 4b, c).

For fungi, 22 SNP-associated OTUs were identified in SpringPan, comprising 5 positively correlating biomarkers from *Setophoma, Monosporascus, Magnaporthiopsis, Schizothecium*, and the order Sordariales. In WinterPan, 5 fungal biomarkers were associated with SNP loci, including taxa from *Leucosporidium, Dioszegia, Alternaria*, and *Cryptococcus* (Fig. 4e, f).

Overall, a higher number of fungal taxa than bacterial taxa were identified as microbial biomarkers associated with host genetics, and more bacterial and fungal biomarkers were detected in SpringPan than in WinterPan.

Host genetic control of positive biomarkers varied across spatial contexts

We next examined whether positive host-associated microbial biomarkers showed consistent host-genetic occupancy and whether their abundances remained spatially structured across the field. We first illustrated the spatial distribution of these SNP-associated positive biomarkers in bulk soil across the 60

plots (Supplementary Materials Fig. S12). Most of these taxa were undetectable or only rarely detected in bulk soil and, when present, occurred in only a limited number of plots, indicating limited representation in the surrounding bulk soil community.

To summarize host genetic associations, we constructed a network linking barley SNP loci to positive microbial biomarkers (Fig. 5a). In SpringPan, two SNP loci emerged as major hubs. In particular, JHI-Hv50k-2016-201999 was associated with two bacterial and three fungal positive biomarkers. In contrast, several other SNP loci exhibited more restricted, kingdom-specific associations or were linked to only a limited number of biomarkers. In WinterPan, only a single SNP locus was associated with four fungal biomarkers and with no bacterial biomarkers (Fig. 5a). Notably, no biomarker-associated SNP loci were shared between SpringPan and WinterPan, indicating distinct host genetic association patterns between the two barley panels. Candidate gene annotation further showed that several focal SNPs overlapped annotated barley genes or were located near candidate genes (Extended Data Table 2).

Beyond abundance, the biomarkers associated with specific SNPs also showed substantial variation in occupancy across genotypes, defined as detection in *k* out of six replicates within a genotype (Fig. 5a). Some biomarkers were consistently detected across many genotypes, whereas others were restricted to a smaller subset despite comparable levels of genetic association significance. For example, *Pseudomonas* ASV68333 was rarely detected consistently within genotypes, whereas *Devosia* ASV137553 occurred more broadly but still showed low occupancy (≤ 3 of 6 replicates) in more than half of the genotypes in which it was detected. By contrast, fungal biomarkers generally showed higher occupancy across genotypes than bacterial biomarkers, although occupancy remained variable overall.

We further quantified the extent to which the abundances of these host genetic-associated biomarkers remained structured by soil spatial heterogeneity using a likelihood ratio test (LRT)-based model comparison framework with PCNMs. Host genetic-associated positive biomarkers remained strongly structured by soil spatial heterogeneity (Fig. 5b). Across all tested biomarkers, adding the top 10 selected PCNM axes significantly improved model fit relative to the genotype-only null model (LRT $p < 0.001$). The increase in marginal R^2 ranged from 0.072 to 0.110 for bacterial biomarkers in SpringPan, from 0.090 to 0.170 for fungal biomarkers in SpringPan, and from 0.222 to 0.236 for fungal biomarkers in WinterPan. Overall, spatial effects were stronger in fungi than in bacteria, with the strongest spatial dependency observed for fungal biomarkers in WinterPan. These effects were quantified using linear mixed-effects models (LME) in which genotype was included as a random effect and PCNM axes were used to represent soil spatial structure. Comparisons between full- and reduced-PCNM models further showed that full models generally achieved better fit according to ΔAIC, whereas top10-PCNM models were more often preferred by ΔBIC (Supplementary Materials Fig. S13), supporting the use of the reduced top10-PCNM model as a parsimonious representation of biomarker-level spatial structure.

We then tested whether host genetic effects on individual biomarkers changed across spatial contexts by modelling biomarker abundance as a function of SNP allele, top10 PCNM spatial structure, and their interaction. Several positive biomarkers showed significant SNP-by-PCNM interactions (Fig. 5c), whereas

corresponding non-significant interaction patterns are shown in Extended Data Fig. 1. These results indicate that host genetic effects on biomarker abundance were dependent on spatial context. Interestingly, the hub SNP JHI-Hv50k-2016-201999 exhibited significant interactions with multiple PCNM axes in modulating both bacterial and fungal biomarkers, implying broad context dependence of this locus. In addition, the significant interactions were distributed differently across the gradient: JHI-Hv50k-2016-201999 acted predominantly on the lower-PCNM12 side, whereas JHI-Hv50k-2016-274673 was associated with enrichment of more positive biomarkers on the higher-PCNM12 side.

We therefore asked whether such context dependence was more evident at the level of individual biomarkers or of the overall positive biomarker abundance (Fig. 5d). Relationships between host selection strength and the stability of individual biomarkers were generally weak or inconsistent, especially in fungi. By contrast, the total positive biomarker abundance showed a clearer association with host selection strength. Together, these results suggest that context dependence was more apparent at the total abundance level than for any single biomarker alone.

Discussion

Using large-scale metabarcoding data from barley roots generated under field conditions, we demonstrate that barley genotypes exert a detectable and reproducible influence on the composition of root-associated microbial communities despite the strong effect of soil heterogeneity. This finding supports the emerging view that the plant microbiome is partially determined by host genetics with feedback effects on the plant phenotype and is therefore a potential target for crop improvement[15–17]. However, our results further show that host genetic effects on microbiome assembly are strongly contingent on local environmental conditions. Rather than acting as fixed properties of plant genotypes, host effects were frequently expressed through interactions with spatially structured soil environments, indicating that host determined microbiome recruitment in the field is inherently context dependent.

Host genotype is often considered a relatively minor component compared to environmental variation in shaping plant microbiomes under field conditions. Previous studies have estimated that most root microbiome variation is environmentally determined[18], with only a relatively small fraction attributable to host genotype, and even heritable microbial taxa are often strongly associated with environmental context[19]. Our comparison between spring and winter panels are broadly consistent with this view: differences between the two panels grown in different conditions exceeded the effect of genotype alone (Fig. 1a,b).

Within individual field panels, however, host genetic effects were clearly detectable. Genotype alone explained approximately 20–43% of microbial variation (Fig. 2a), and when genotype-by-space interaction terms were included, the total explained variation exceeded 55%. Thus, host effects on microbiome assembly in the field may be substantially underestimated when considered only as a direct genotype main effect, because an important component of plant microbiome assembly is expressed through the host-by-environment term rather than genotype effects alone. Although such host-by-

environment contingencies have been more explicitly recognized in human and animal microbiome research[20–22], they have been less clearly incorporated into studies of plant microbiome assembly under field conditions. Importantly, we found that the relatively large within-genotype dispersion did not necessarily indicate weak host control, as stronger host effects in some cases coincided with similar or greater within-genotype divergence (Fig. 3b). Together with the significant genotype-by-space interactions, this suggests that an important part of host influence is expressed through the interaction with spatially structured soil heterogeneity. In this sense, the interaction effects on the microbiome may be better understood as an extended genotype effect, in which microbiome recruitment is shaped not only by genotype itself but also by how that genotype is expressed across spatially heterogeneous soil environments[23–25].

We further examined whether host-environment interactions extended to the recruitment of specific microbial taxa that may be agronomically relevant. Several host SNP-associated positive biomarkers of barley biomass were modulated by SNP-by-spatial-context interactions (Fig. 5c), indicating that the enrichment effect of individual host loci depended on local soil context. In some cases, these interaction patterns appeared potentially complementary across spatial contexts, such that context dependence was more evident at the level of individual biomarkers, whereas greater stability emerged at the level of total positive biomarker abundance (Fig. 5d). Together, these patterns suggest that host genetic effects in the field may be exerted across alternative microbial targets under different environmental conditions, rather than being expressed through uniform recruitment of the same taxa in all contexts, as both root exudation profiles and microbiome assembly are both shaped by environmental conditions [26,27].

The mechanisms underlying these host-by-space interactions remain unresolved, but at least two non-exclusive processes are likely to contribute. Soil spatial heterogeneity is expected to alter the local microbial source pool available for recruitment[28–30], while also modifying host-mediated recruitment processes, such that the same genotype may recruit different microbial assemblages under different local conditions[31–34]. Our results indicate that the first process alone is unlikely to account for the observed patterns. We observed no significant relationship between bulk and root microbiome composition (Fig. 3a), and most SNP-associated microbial biomarkers were below detection limits in bulk soil at the sequencing depth achieved in this study (Supplementary Materials Fig. S12). These observations suggest that variation in local microbial source pools alone is unlikely to explain the observed host-by-space interactions. Instead, spatially structured soil environments may influence host physiological traits, including root exudation, nutrient status, or immune-related processes[26,27], thereby altering microbiome recruitment by the same host genotype across different soil contexts.

A notable finding of our study is that host-mediated divergence from surrounding soil communities and microbiome assembly consistency can become decoupled under heterogeneous field conditions. Reduced within-host microbiome variability has often been interpreted as evidence of stronger deterministic assembly or host filtering[35–37], because strong selection is expected to homogenise microbial communities among individuals sharing the same host genotype. However, our results suggest

that these two aspects of microbiome assembly can become decoupled under heterogeneous field conditions. Genotypes exhibiting greater divergence from surrounding bulk communities also showed higher within-genotype beta-dispersion (Fig. 3b). Thus, stronger host-mediated divergence from soil communities did not necessarily result in more homogeneous microbiome assembly across space. Instead, these findings support a model of adaptive host filtering, in which host genotypes maintain selective recruitment while adjusting the taxonomic identity of recruited microbiota according to local environmental conditions (Fig. 5d). Under this framework, host plants may preserve microbiome-associated functions[38] while allowing recruitment trajectories to vary among heterogeneous environments.

Together, our findings highlight a fundamental interaction between host genetic control and environmental modulation in shaping plant-microbiome associations under field conditions. More broadly, our results indicate that environmental variation is not simply noise obscuring host genetic effects, but an integral part of how microbiome recruitment is expressed in the field. For microbiome-assisted breeding, within-genotype dispersion, sometimes interpreted as reduced microbial stability, should not necessarily be regarded as an undesirable feature. Instead, greater dispersion may in some cases reflect context-dependent recruitment of distinct microbial assemblages under different environmental conditions. This suggests that priority should not be given solely to plant loci showing strong associations with a single beneficial microbe, but also to host-microbial associations that support robust or flexible recruitment across spatially heterogeneous field environments.

Methods

Field experiment and sampling design

To evaluate host-mediated microbiome recruitment under actual field heterogeneity, we established two large barley field trials at the Raasdorf-Pysdorf site (48.236371° N, 16,587804° E) of the BOKU experimental station, Austria consisting of spring or winter barley lines and thus grown under contrasting seasonal conditions (Supplementary Materials Fig. S1a). The field trial of winter barley lines (WinterPan) was sown on October 19th, 2022. The field trial of spring barley lines (SpringPan) was sown on March 30th, 2023. Each field trial was sown with 200 different barley genotypes. SpringPan contained old and modern spring barley varieties and elite genotypes from mostly Europe and the Mediterranean Region, while WinterPan contained winter varieties from mostly Europe and West Asia. Fourteen selected varieties (spring lines) were shared across the two panels as ‘check lines’ to test the effect of different seasonal conditions. Information regarding all the barley varieties is listed in Extended Data Table 3. Each genotype was in 3 replicated plots of approximately 3 m^2. Therefore, each field trial had in total 600 plots, with a layout of 30 rows and 20 columns. For root sampling for microbiome analysis, each plot was subsequently split into 2 subplots (northern and southern end) to assess within plot variation generating a total of 6 replicates per variety. The two field trials (SpringPan and WinterPan) were next to

each other but experienced different patterns of rainfall, soil and canopy temperature, and solar radiation due to the different time period of cultivation (Supplementary Materials Fig. S1c).

Barley root and bulk soil sampling

Root sampling for microbiome analyses was conducted when barley plants in the individual plots reached BBCH 27, corresponding to late tillering and prior to the onset of stem elongation. From each subplot, 5 roots were collected randomly by digging up roots of individual plants from 0–10 cm depth with a hand shovel. The roots were shaken to remove loosely adhering soil and pooled together as one sample and kept cold during transport to storage at -20°C until microbiome analysis. In total, 2,400 root samples (200 lines × 6 replicates × 2 field trials) comprising roots from a total of 12,000 barley plants were collected. To quantify shoot biomass, aboveground tissues of the same five barley plants used for root microbiome analysis were harvested at the time of sampling. The plant material was dried at 40°C and 15% relative humidity for 5 days, before measuring shoot biomass using an analytical balance and expressed as dry weight (g) per sample. Besides root samples, 60 bulk soil samples were collected in the unplanted areas next to specific plots (see layout in Supplementary Materials Table S2) from each field trial to determine the soil spatial heterogeneity in terms of soil properties and local microbial source pool.

Sample preparation

The collected barley roots with adhering soils were shipped to Aarhus University (Flakkebjerg, Denmark) at -20°C. Upon arrival, all samples were immediately stored at -20 °C until further processing. Subsequently, samples were freeze-dried (ScanVac CoolSafe freeze dryer, Leogane JSCo., Allerød, Denmark) for 72 hours and sorted manually to remove non-root materials, such as leaf debris, seeds, and stems. The roots with adhering soil were then placed in sterile 90 mL plastic containers and ground with ten sterile steel balls (ø 5mm) at 1,500 strokes per min for 3 minutes using a GenoGrinder (SPEX SamplePrep, USA). The resulting powder was stored at room temperature until DNA extraction.

Soil chemical analyses

Soil chemical properties were determined for bulk soil collected from the selected 60 plots in each field. Soil samples were sieved to 2 mm prior to analysis. For total carbon (C) and nitrogen (N), a 0.5 g homogenized subsample was dried at 105°C for 24 h, and total C and N contents (%) were determined from a 250 mg subsample using a TruSpec CHN analyser (LECO). Soil pH was measured in either water or $CaCl_2$ suspension, as indicated for each variable. Exchangeable cations were determined according to EN ISO 11260 using an unbuffered 0.1 M $BaCl_2$ extraction. Briefly, 2.5 g of air-dried fine soil was extracted with 50 mL of 0.1 M $BaCl_2$, allowed to stand overnight, shaken for 2 h on a rotary shaker, and then filtered. The filtrates were analysed by ICP-OES for Al, Fe, Ca, Na, K, Mg and Mn. For the unbuffered

$BaCl_2$ extracts, pH was additionally recorded. Effective cation exchange capacity (CEC) was calculated as the sum of exchangeable cations, and base saturation (BS) was calculated as the percentage of CEC occupied by base cations (Ca^{2+}, Mg^{2+}, K^+ and Na^+).

DNA extraction

Total DNA of barley root and bulk soil samples was extracted using the LGC sbeadex™ mini plant kit (LGC Genomics, Teddington, UK) in combination with KingFisher™ Flex (KingFisher, London, UK). Briefly, 100 mg of ground material was weighed into a 1.5 mL sterile Eppendorf tube and extracted into a 55 µL elution buffer solution following the manufacturer's protocol. All samples were extracted in a random order to control batch effects. The extracted DNA samples were stored at -20 ℃ until use.

Metabarcode sequencing

To determine the bacterial and fungal community composition, Illumina MiSeq sequencing was performed on both root and bulk soil samples. The libraries of bacterial and fungal DNA were generated by a two-step dual-indexing PCR strategy, as previously described[39]. For bacteria, the V3/V4 region of 16S ribosomal RNA (rRNA) was amplified using primers 341F (S-DBact-0341-b-S-17: 5'-CCTACGGGNGGCWGCAG-3') and 785R (S-D-Bact-0785-a-A-21: 5'-GACTACHVGGGTATCTAATCC-3') [40]. For fungi, the internal transcribed spacer 2 (ITS2) region of the fungal rRNA gene was amplified using primers fITS7 (5'-GTGARTCATCGAATCTTTG-3') and ITS4 (5'-TCCTCCGCTTATTGATATGC-3') [41]. In the first (profiling) PCR, tagged locus-specific primers were used to amplify the target regions. Reactions were performed using MyTaq HS DNA Polymerase (Bioline, London, UK), the supplied reaction buffer, dNTPs, tagged primers, and template DNA in a total reaction volume of 5 µL. The PCR program was as follows: 3 min at 95°C followed by 27 cycles of 95°C for 30 s, 55°C/57°C (bacteria/fungi) for 30 s, and 72°C for 30 s, and final elongation at 72°C for 5 min. Following the first PCR, amplicons were purified and normalized. Sample-specific unique dual index (UDI) sequences were added to generate sequencing-ready libraries compatible with the Illumina MiSeq plat-form, enabling multiplex sequencing and subsequent sample demultiplexing. Indexed libraries were pooled in equimolar amounts, purified, and sequenced on an Illumina MiSeq instrument (Eurofins Genomics, Ebersberg, Germany) using MiSeq Reagent Kit v3 chemistry with paired-end 2 × 300 bp sequencing.

Bioinformatics analysis

Amplicon sequencing data were processed using Quantitative Insights into Microbial Ecology 2 (QIIME 2, version 2025.4)[42]. For bacteria, amplicon sequence variants (ASVs) were determined by DADA2[43] from raw paired-end reads using the "qiime dada2 denoised-paired" function. Forward and reverse reads were trimmed by 17 and 21 bp at the 5' end, respectively, to remove primer sequences, and truncated at 300

bp (forward) and 200 bp (reverse) to remove low-quality bases. Taxonomic assignment of bacterial ASVs was performed using a naïve Bayes classifier implemented in the QIIME 2 by the “feature-classifier” function against the Silva Ref NR99 [release 138.1] database [44]. Non-bacterial reads, including chloroplast and mitochondrial reads, were removed prior to downstream analyses. Samples containing fewer than 3,000 reads and ASVs detected in fewer than five samples were excluded from subsequent analyses.

For fungi, ASVs were initially generated using the same DADA2 workflow as described for bacteria, except that forward and reverse reads were trimmed at 19 and 20 bp and truncated at 300 and 220 bp. Because the fungal ITS regions often exhibit substantial intra-genomic and intra-specific sequence variation, ASV-level resolution may artificially split biologically similar taxa and inflate estimates of diversity. To obtain a more conservative representation of fungal taxa for ecological analyses, ITS2 regions were extracted from DADA2-derived sequences using ITSx [45], and sequences shorter than 90 bp were discarded. The remaining sequences were subsequently clustered into operational taxonomic units (OTUs) at 98.5% sequence similarity using VSEARCH. This widely used threshold[46] was adopted as a pragmatic approximation of species-level Species Hypotheses (SHs) implemented in the UNITE database[47] and to reduce artificial inflation of fungal diversity resulting from within-species sequence variation. The resulting OTUs were subsequently taxonomically annotated using the IdTaxa function from the DECIPHER package in R against the UNITE February 2025 training set[48]. Non-fungal sequences were removed prior to downstream analyses. Samples containing fewer than 4,000 reads and OTUs detected in fewer than five samples were excluded from subsequent analyses.

Microbiome community composition analyses

For community-level analyses, bacterial ASV and fungal OTU count tables were converted to relative abundances and Hellinger-transformed, and Bray-Curtis dissimilarities were calculated for downstream multivariate analyses. Community composition was tested against genotype identity and spatial variables using permutational multivariate analysis of variance (PERMANOVA). In the field panels, models included genotype identity, row and column position, subplot side, and genotype-by-space interaction terms.

Host genetic structure and distance-based analyses

For spring barley, previously generated genotype data barley from the Barley Infinium iSelect 50K single nucleotide polymorphism (SNP) array (Illumina, San Diego, CA, USA) were obtained as described by Bouhlal et al.[49]. Lines lacking genotype data were genotyped separately using the same SNP array, with leaf samples collected at the ICARDA glasshouse and submitted to LGC Genomics (Hoddesdon, UK) for genotyping. Winter barley genotype data were generated separately within the present study using the same SNP array. To provide functional context for focal loci, candidate gene annotation was conducted

using the barley Morex V3 reference genome annotation (RefSeq assembly GCF_904849725.1). For each focal SNP, overlapping genes, the nearest gene, and genes located within plus/minus 50 kb, 100 kb, and 250 kb were identified based on physical position, and available functional descriptions were retrieved from the corresponding reference gene models where possible.

Single Host genetic structure was quantified from genome-wide SNP data using Prevosti genetic distances. Principal coordinate analysis (PCoA) was performed on the resulting distance matrices to summarize host genetic variation. Shared check lines were excluded from trial-specific analyses. To test whether the relationship between host genetic divergence and microbiome differentiation changed across space, pairwise microbial Bray-Curtis dissimilarities were regressed against pairwise host genetic distances, and comparisons were further stratified into three spatial classes defined by the lower, intermediate and upper tertiles of either pairwise Euclidean distances calculated from field coordinates or scores of significant PCNM axes.

Spatial structure and variation partitioning

To characterize spatial structure beyond simple row and column effects, principal coordinates of neighbour matrices (PCNMs) were calculated from plot coordinates[50]. For subplot-level analyses, row coordinates were adjusted according to subplot side before Euclidean distance calculation and PCNM derivation. Associations between PCNM axes and measured soil properties were assessed using Spearman correlations. To compare the relative contributions of host genetic structure and spatial heterogeneity to microbiome variation, distance-based redundancy analysis (dbRDA; capscale[51], Bray-Curtis distance) with forward selection (ordiR2step, 999 permutations)[52] was used to identify significant PCNM axes and significant host-genetic PCoA axes for each panel and organism. Variation partitioning (varpart)[53] was then used to estimate the independent and shared fractions of community variation explained by host genetic and spatial components.

Spatial-context dependence of host genetic effects

To test whether host genetic effects on microbiome differentiation changed across spatial contexts, pairwise microbial community dissimilarity was modelled as a function of host genetic distance, spatial context, and their interaction. Spatial context was represented either by raw plot-coordinate (XY) distance or by each of the top 10 forward-selected PCNM axes for the corresponding panel and organism. Host genetic distance, microbial community dissimilarity, and spatial variables were standardized as z-scores, and linear models were fitted with a host genetic distance × spatial context interaction term. The resulting interaction coefficients were used to assess whether host genetic effects became stronger or weaker across more dissimilar spatial contexts. For visualization, pairwise comparisons were additionally grouped into near, mid, and far spatial-context classes based on tertiles of XY distance or PCNM values.

Host selection strength and within-genotype community dispersion

To test whether stronger host selection was associated with microbiome convergence or divergence within genotypes, we quantified host selection strength at the sample level as the Bray-Curtis dissimilarity between each root microbiome sample and its six nearest bulk soil samples. This metric was used to represent the extent to which the root microbiome diverged from the local surrounding soil community. Within-genotype community dispersion was quantified as the distance of each sample to its genotype centroid (beta-dispersion[54]) based on Bray-Curtis dissimilarities among root microbiome samples. Relationships between host selection strength and distance to genotype centroid were then tested using linear mixed-effects models with genotype included as a random intercept. For visualization, both variables were standardized as z-scores, and sensitivity analyses were additionally performed after removing the upper and lower 1% of values as outliers identified by Mahalanobis distance within each community-by-trial subset (chi-square cutoff, upper 1%).

Biomarker identification

Microbiome biomarkers associated with barley biomass were identified using a two-step feature-selection framework. First, taxa with relative abundance > 0.05% in at least 10% of samples were retained and screened using Boruta[55]. Confirmed features were then used to train XGBoost[56] regression models predicting above-ground biomass, with the optimal number of boosting rounds selected by five-fold cross-validation with early stopping. SHAP[57] values were used to quantify feature importance and effect direction.

Candidate biomarkers were further evaluated using linear mixed-effects models (lme4[58]) based on rclr-transformed abundances. For each candidate taxon, above-ground biomass was modelled as a function of biomarker abundance, with genotype, row and column included as random intercepts. Taxa were retained as host-associated biomarkers when the direction of the mixed-model slope was concordant with the SHAP-derived effect direction.

Microbiome GWAS

For microbiome genome-wide association analyses (MGWAS), rclr-transformed biomarker abundances were averaged across replicate samples within genotype and used as genotype-level phenotypes. SNP genotypes were converted from HapMap to PLINK format and filtered for genotype missingness < 20% and minor allele frequency > 1%. A kinship matrix was estimated from genome-wide SNPs, and the first 10 genotype principal components were included as covariates to account for population structure. GWAS was then performed separately for each biomarker using GEMMA linear mixed models[12,59].

Spatial dependence of positive SNP-associated biomarkers

To test whether positive host-associated biomarkers retained spatial structure after accounting for host genotype identity, we fitted linear mixed-effects models to rclr-transformed biomarker abundances with genotype included as a random intercept. Spatial effects were represented by forward-selected PCNM axes from the corresponding organism-panel combination. For each biomarker, we compared a null model including only genotype identity with spatial models including either all selected PCNMs or the top 10 selected PCNMs. Spatial support was assessed using likelihood-ratio tests together with differences in marginal R^2, AIC and BIC.

SNP-by-PCNM interaction analyses

To test context-dependent host recruitment at the single-biomarker level, rclr-transformed abundances of positive SNP-associated biomarkers were modelled as a function of associated SNPs and spatial context represented by PCNM axes. For each biomarker, up to two significant associated SNPs and the top 10 significant PCNM axes from the corresponding organism-panel combination were included as main effects, together with all SNP-by-PCNM interaction terms. Reduced models including only SNP and PCNM main effects were compared with full models additionally including SNP-by-PCNM interaction terms using ANOVA, and interaction strength was summarized as the change in adjusted R^2. P values of interactions were adjusted for multiple testing using the Benjamini-Hochberg procedure.

To visualize context-dependent recruitment patterns, we plotted fitted allele-specific regressions of biomarker abundance against individual PCNM axes for biomarkers showing SNP-by-PCNM interactions. For each panel, the x-axis represents the focal PCNM axis, the y-axis represents rclr-transformed biomarker abundance, and separately fitted lines correspond to alternative allelic states at the associated SNP. These plots were used to illustrate how allele-specific biomarker responses diverged across the same spatial gradient.

To assess whether host selection was associated with stability of microbiome recruitment, we further quantified the abundance of positive host-associated biomarkers at both the pooled and individual-biomarker levels. For each genotype, mean abundance, standard deviation, and coefficient of variation were calculated across replicate samples and related to host selection strength using correlation and regression-based summaries.

Declarations

Funding

This article is published as part of the BarleyMicroBreed project, that has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101060057. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Acknowledgements

We gratefully acknowledge Louise Reinbach Rasmussen, Stina Sjöberg, Tamás Plaszkó, and Somayeh Gholizadeh for their substantial assistance with sample processing and DNA extraction.

Conflict of Interest

The authors declare no competing interests.

Author Contributions

M.N., M.V., and C.J.B. conceived the study. M.D. designed and performed the analyses, interpreted the results, and wrote the original draft of the manuscript. A.S.M., C.J.B., and K.L.K. contributed to data analysis and interpretation. Katarzyna Retzer and H.G. designed the field experiment, collected the field samples, measured plant biomass, and performed soil chemical analyses. M.S.G., A.V., and A.E.B. provided a subset of the barley lines and performed SNP-chip genotyping. Kerstin Rönsch and M.S. conducted the microbiome sequencing. M.N., M.V., and C.J.B. supervised the study. All authors reviewed and revised the manuscript and approved the final version.

Data Availability

Raw amplicon sequencing reads have been deposited in the NCBI Sequence Read Archive under BioProject accession PRJNA1499680. The records are currently under embargo and will be made publicly available upon publication. The HapMap-formatted SNP genotype data, barley above-ground biomass data and sample metadata underlying the analyses have been deposited in Zenodo under DOI [10.5281/zenodo.22143284]. These files are also currently under embargo and will be made publicly available upon publication. Confidential access details for both data deposits have been provided separately to the editors and reviewers.

Code Availability

Custom scripts used for the analyses are available from the corresponding authors upon reasonable request.

References

1. Zhang J, Liu W, Bu J, Lin Y, Bai Y (2023) Host genetics regulate the plant microbiome. Curr Opin Microbiol 72:102268
2. Week B, Russell SL, Schulenburg H, Bohannan BJM, Bruijning M (2025) Applying evolutionary theory to understand host–microbiome evolution. Nat Ecol Evol 9:1769–1780
3. Ferretti P, Johnson K, Priya S, Blekhman R (2026) Genomics of host–microbiome interactions in humans. Nat Rev Genet 27:62–80
4. Couch CE, Epps CW (2022) Host, microbiome, and complex space: Applying Population and Landscape Genetic Approaches to Gut Microbiome Research in Wild Populations. J Hered 113:221–234
5. Compant S et al (2025) Harnessing the plant microbiome for sustainable crop production. Nat Rev Microbiol 23:9–23
6. Singh BK, Hu HW, Macdonald CA, Xiong C (2025) Microbiome-facilitated plant nutrient acquisition. Cell Host Microbe 33:869–881
7. Hong Y, Zhou Q, Hao Y, Huang AC (2022) Crafting the plant root metabolome for improved microbe-assisted stress resilience. New Phytol 234:1945–1950
8. Liu S et al (2026) Host-mediated rhizosphere microbiome transfer suppresses *Fusarium oxysporum* in banana. New Phytol 250:2462–2476
9. Shi S, Bastías DA, Wang H, Faville M, O'Callaghan (2026) M. A roadmap for plant–microbiome breeding to enhance plant stress tolerance. Trends Microbiol 34:614–626
10. Raza S et al (2025) Future crop breeding needs to consider future soils. Nat Plants 11:939–941
11. He X et al (2024) Heritable microbiome variation is correlated with source environment in locally adapted maize varieties. Nat Plants 10:598–617
12. Wang Y et al (2022) GWAS, MWAS and mGWAS provide insights into precision agriculture based on genotype-dependent microbial effects in foxtail millet. Nat Commun 13:5913
13. Song Y et al (2025) Seed tuber microbiome can predict growth potential of potato varieties. Nat Microbiol 10:28–40
14. Zhao T et al (2025) Unveiling potato cultivars with microbiome interactive traits for sustainable agricultural production. Plant Cell Environ 1–15
15. Compant S et al (2024) Harnessing the plant microbiome for sustainable crop production. Nat Rev Microbiol 23:9–23
16. Kumar SM et al (2026) Microbiome-assisted plant breeding: integrating host-microbiome interactions into crop improvement. Arch Microbiol 208:242
17. Araujo ASF, Costa RM, Lopes AC (2026) A. Plant breeding trials should include the belowground microbiome. npj Sci Plants 2:15
18. Xun W et al (2024) Dissection of rhizosphere microbiome and exploiting strategies for sustainable agriculture. New Phytol 242:2401–2410

19. He X et al (2024) Heritable microbiome variation is correlated with source environment in locally adapted maize varieties. Nat Plants 10:598–617
20. Manasa S, Panigrahy L, Panigrahi M, K. C. S., Rout GR (2021) Overview of cold stress regulation in plants. Bot Rev 88:359–387
21. Wang MH, Achkar JP (2015) Gene-environment interactions in inflammatory bowel disease pathogenesis. Curr Opin Gastroenterol 31:277–282
22. Gubert C et al (2022) Gene-environment-gut interactions in Huntington's disease mice are associated with environmental modulation of the gut microbiome. iScience 25:103687
23. Cadwell K, Loke P (2025) Gene–environment interactions shape the host–microbial interface in inflammatory bowel disease. Nat Immunol 26:1023–1035
24. Veach AM et al (2019) Rhizosphere microbiomes diverge among *Populus trichocarpa* plant-host genotypes and chemotypes, but it depends on soil origin. *Microbiome 2019 7:1* 7, 76
25. Turner TR, James EK, Poole PS (2013) The plant microbiome. Genome Biol 14:209
26. Jing J, Cong WF, Bezemer TM (2022) Legacies at work: plant–soil–microbiome interactions underpinning agricultural sustainability. Trends Plant Sci 27:781–792
27. Hassani MA, Durán P, Hacquard S (2018) Microbial interactions within the plant holobiont. Microbiome 6:58
28. Ulbrich TC, Rivas-Ubach A, Tiemann LK, Friesen ML, Evans SE (2022) Plant root exudates and rhizosphere bacterial communities shift with neighbor context. Soil Biol Biochem 172:108753
29. Bulgarelli D et al (2012) Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. Nature 488:91–95
30. Lundberg DS et al (2012) Defining the core Arabidopsis thaliana root microbiome. Nature 488:86–90
31. Ling N, Wang T, Kuzyakov Y (2022) Rhizosphere bacteriome structure and functions. Nat Commun 13:836
32. Bouwmeester H, Dong L, Wippel K, Hofland T, Smilde A (2025) The chemical interaction between plants and the rhizosphere microbiome. Trends Plant Sci 30:1002–1019
33. Dilla-Ermita CJ, Lewis RW, Sullivan TS, Hulbert SH (2021) Wheat genotype-specific recruitment of rhizosphere bacterial microbiota under controlled environments. Front Plant Sci 12:718264
34. Spooren J, van Bentum S, Thomashow LS, Pieterse CMJ, Weller DM, Berendsen RL (2024) Plant-Driven Assembly of Disease-Suppressive Soil Microbiomes. Annu Rev Phytopathol 62:1–30
35. Wagner MR, Lundberg DS, del Rio TG, Tringe SG, Dangl JL, Mitchell-Olds T (2016) Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. Nat Commun 7:12151
36. Arnault G, Mony C, Vandenkoornhuyse P (2023) Plant microbiota dysbiosis and the Anna Karenina Principle. Trends Plant Sci 28:18–30
37. Li T, Gao J (2023) Attribution of dispersal limitation can better explain the assembly patterns of plant microbiota. Front Plant Sci 14:1168760

38. Hu J, Vandenkoornhuyse P, Khalfallah F, Causse-Védrines R, Mony C (2023) Ecological corridors homogenize plant root endospheric mycobiota. New Phytol 237:1347–1362
39. Taketani RG et al (2026) Host plant selects bacterial rhizosphere microbiome function whereas community structure is determined by soil legacy. ISME Commun 6:83
40. Kudjordjie EN, Sapkota R, Steffensen SK, Fomsgaard IS, Nicolaisen M (2019) Maize synthesized benzoxazinoids affect the host associated microbiome. Microbiome 7:1–17
41. Klindworth A et al (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41:e1
42. Ihrmark K et al (2012) New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. FEMS Microbiol Ecol 82:666–677
43. Siddique AB, Unterseher M (2016) A cost-effective and efficient strategy for Illumina sequencing of fungal communities: A case study of beech endophytes identified elevation as main explanatory factor for diversity and community composition. Fungal Ecol 20:175–185
44. Wu L et al (2015) Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. BMC Microbiol 15:3
45. Bolyen E et al (2019) Author Correction: Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2 (Nature Biotechnology, 37, 8, (852–857). 10.1038/s41587-019-0209-9). *Nat. Biotechnol.* 37, 1091 (2019)
46. Callahan BJ et al (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583
47. Quast C et al (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596
48. Bengtsson-Palme J et al (2013) Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol Evol 4:914–919
49. Větrovský T et al (2020) GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. Sci Data 7:228
50. Kõljalg U et al (2013) Towards a unified paradigm for sequence-based identification of fungi. Mol Ecol 22:5271–5277
51. Abarenkov K et al (2024) The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered. Nucleic Acids Res 52:D791–D797
52. Bouhlal O et al (2022) CGIAR Barley Breeding Toolbox: A diversity panel to facilitate breeding and genomic research in the developing world. Front Plant Sci 13:1034322
53. Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol Modell 153:51–68

54. Anderson MJ, Willis TJ (2003) Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. Ecology 84:511–525
55. Blanchet FG, Legendre P, Borcard D (2008) Forward selection of explanatory variables. Ecology 89:2623–2632
56. Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. Ecology 73:1045–1055
57. Anderson MJ (2006) Distance-Based Tests for Homogeneity of multivariate dispersions. Biometrics 62:245–253
58. Kursa MB, Rudnicki WR (2010) Feature selection with the boruta package. J Stat Softw 36:1–13
59. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 785–794ACM
60. Lundberg SM, Lee S-I (2017) A Unified Approach to Interpreting Model Predictions. Adv Neural Inf Process Syst 30
61. Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear lixed-effects models Using lme4. J Stat Softw 67:1–48
62. Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nat Genet 44:821–824

Figures

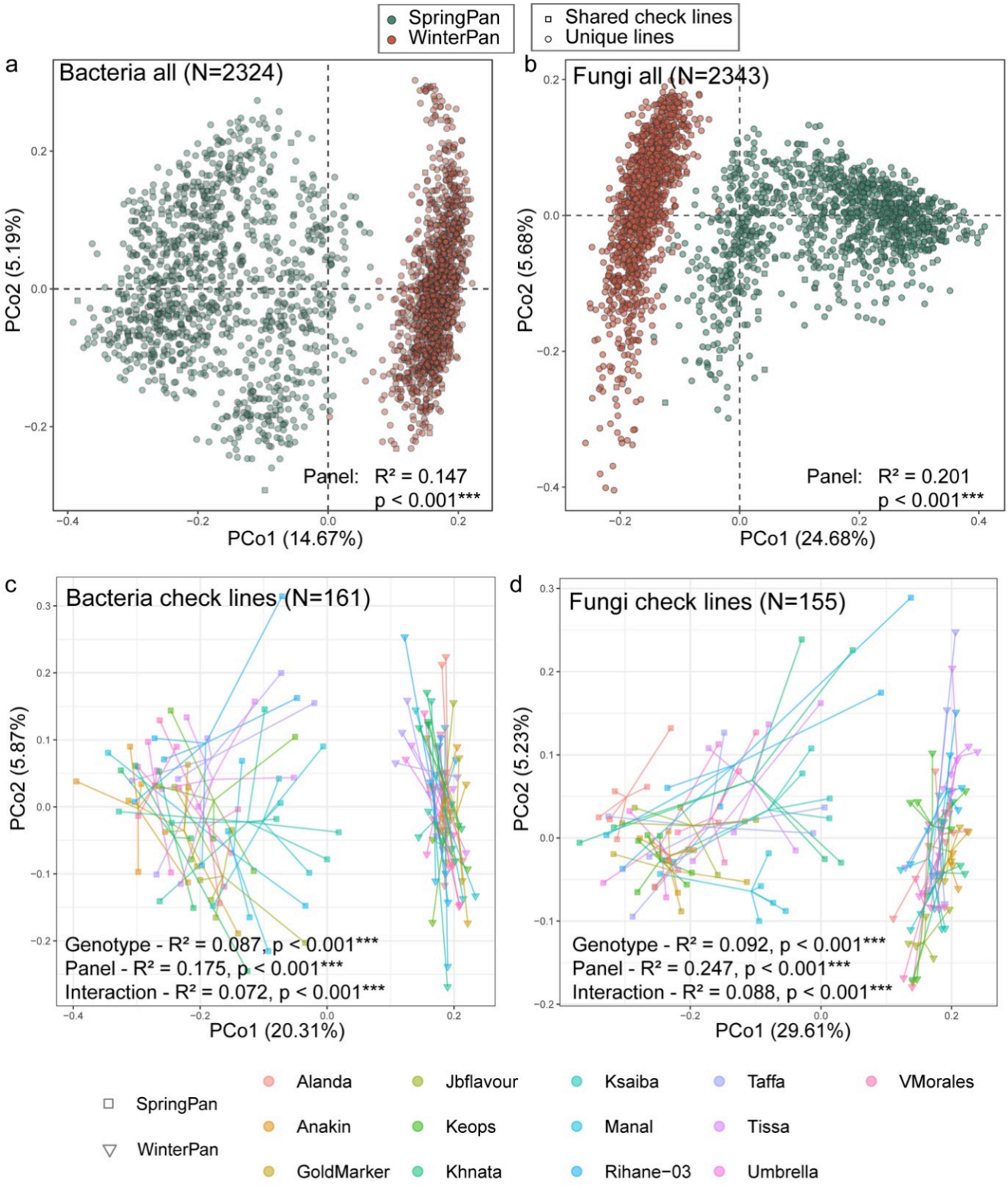


Figure 1

Principal coordinate analysis (PCoA) of microbial community composition of spring and winter panels in all varieties (a, b) and check lines (c, d). Check lines are the selected spring barley lines that were cultivated in both SpringPan and WinterPan. Statistical information is based on PERMANOVA analyses based on Bray Curtis dissimilarities of Hellinger transformed bacterial and fungal relative abundances.

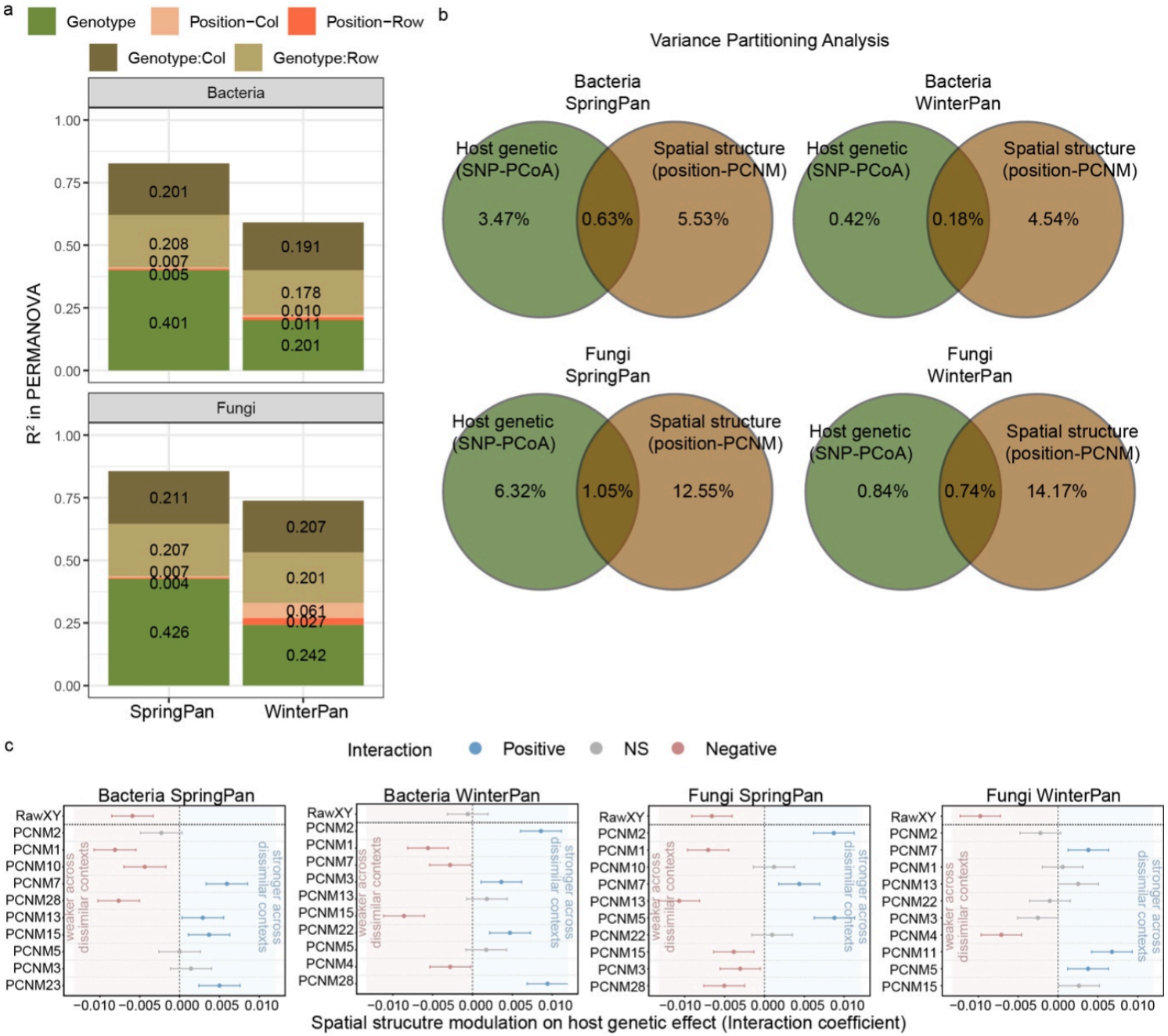


Figure 2

a) PERMANOVA of genotype effect, spatial effect, and their interactions on bacterial and fungal communities in SpringPan and WinterPan. PERMANOVA analyses were based on Bray Curtis dissimilarities of Hellinger transformed bacterial and fungal relative abundances. The spatial effect was evaluated as the effect of plot position in the column (Position-Col) and row (Position-Row) matrix in the field trials. Check lines were excluded to avoid bias when comparing within-panel community variation.
b) Variation partitioning analysis of host genetic and soil spatial effects on microbial community composition. Host genetic structure was represented by PCoA axes derived from Prevosti's genetic dissimilarities calculated from SNP data, and soil spatial structure was represented by PCNM axes. For each component, only axes retained as significant in forward-selected db-RDA models were included in the variation partitioning analysis.

(c) Spatial-context dependence of host genetic effects on microbiome composition. To aid interpretation, spatial context can be viewed as differences in raw plot-coordinate (XY) distance or along each of the top 10 forward-selected PCNM axes, which were further visualized by grouping pairwise comparisons into spatial-context classes in Supplementary Materials Fig. S7. The interaction coefficients shown here were estimated from linear models testing whether the relationship between host genetic distance and microbial community dissimilarity changed across these spatial contexts. Positive coefficients indicate that host genetic effects became stronger across more dissimilar spatial contexts, whereas negative coefficients indicate that host genetic effects became weaker across more dissimilar spatial contexts. Points show interaction estimates and horizontal bars show 95% confidence intervals.

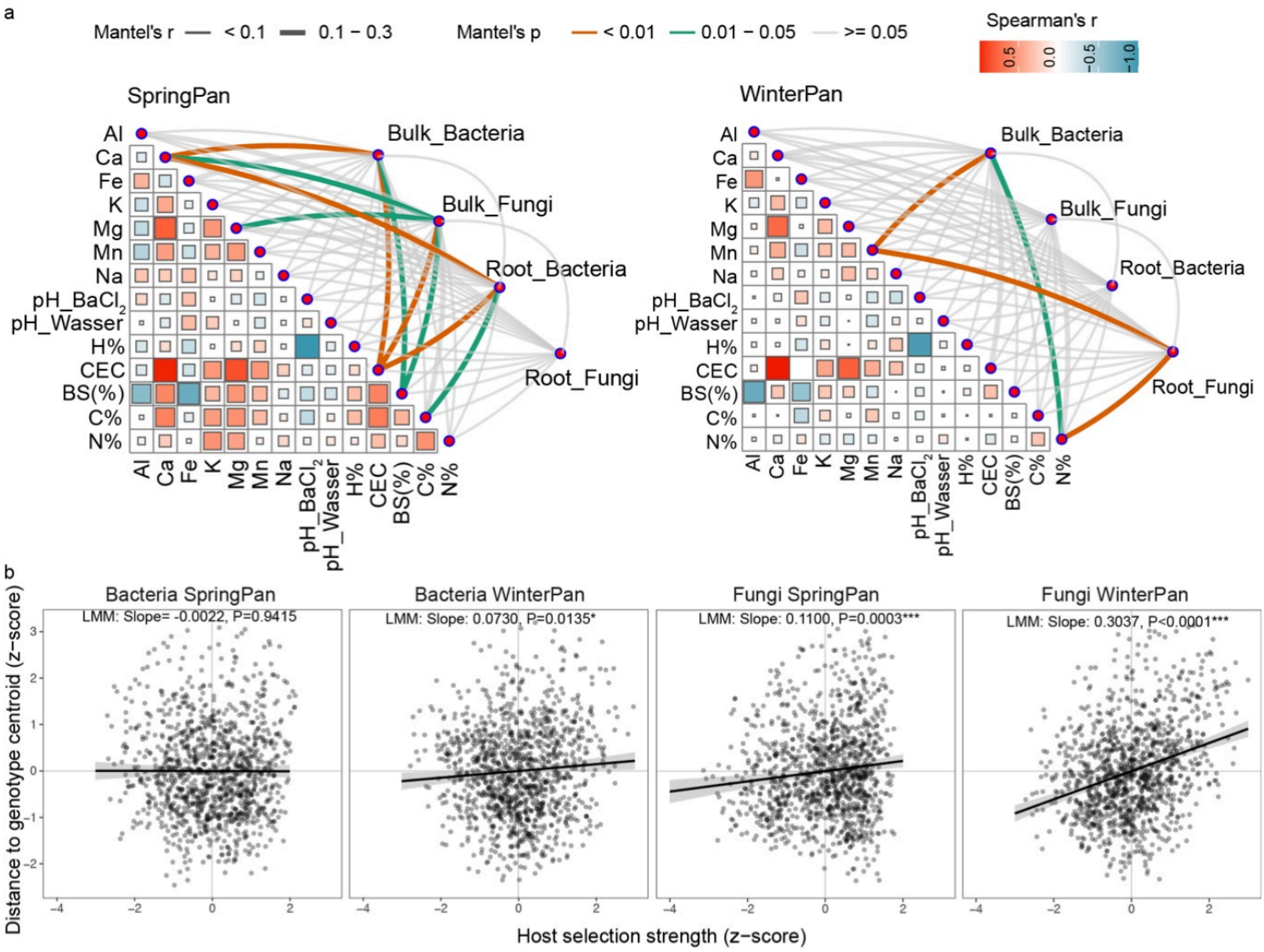


Figure 3

a) Mantel correlations between soil properties and bacterial and fungal community composition in bulk and root-associated soils. Analyses were conducted on 60 plots across the field with matched soil property and bulk soil community measurements.

b) Linear mixed-effects regression of distance to the genotype centroid against host selection strength. Shown after removal of flagged outliers. Host selection strength was defined as the Bray-Curtis distance between each root sample and its six nearest bulk soil samples, reflecting divergence of the root microbiome from the local soil community. Distance to the genotype centroid was estimated by beta-dispersion based on Bray-Curtis dissimilarities and used as a measure of within-genotype variation. Both variables are shown as z-scores. Model: Distance to genotype centroid ~ host selection strength + (1 | Genotype). Corresponding sensitivity analyses using the full dataset are provided in Supplementary Materials Fig.9

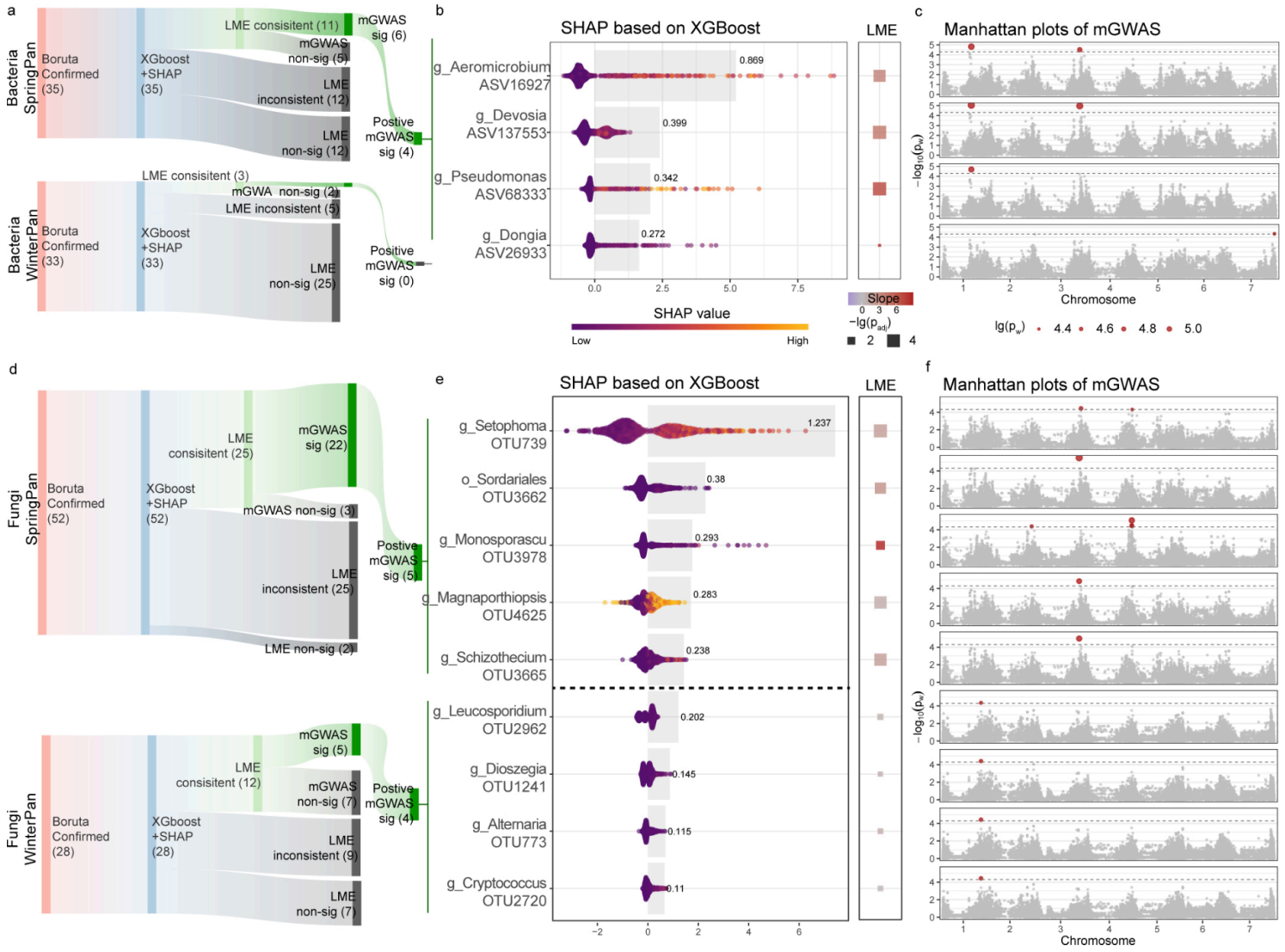


Figure 4

a, d) Workflow for identifying bacterial and fungal biomarker taxa. We first applied Boruta feature selection to identify taxa significantly associated with barley biomass. The contribution of these taxa to barley biomass was then quantified using SHapley Additive exPlanations (SHAP) values derived from an XGBoost model, followed by linear mixed effect modelling with genotype included as a random effect. Taxa showing consistent directions of effect (positive or negative) across SHAP and linear mixed-effects analyses were retained for MGWAS to test associations with barley SNP loci. Finally, only taxa that

positively contributed to barley biomass and were significantly associated with one or more barley SNP loci were retained as positive host-associated biomarkers. Numbers indicate the number of taxa remaining after each filtering step.

b, e) Model-based estimation of correlations between bacterial and fungal taxa, and barley biomass. XGBoost models were trained using a randomly selected 80% subset of the data. SHAP values were subsequently used to evaluate the relative importance of predictors and their positive or negative contributions to biomass prediction. Model performance was assessed using the remaining 20% of the data. Only positive biomarkers with consistent significance and directionality in SHAP and LME and which were significantly associated with one or more SNP loci are shown. The taxa were ranked according to SHAP value that shows the relative importance of the taxa to biomass. For LME, the size of the square indicates the significance while the color indicates the slope (effect size) of the regressions.
c, f) Microbiome-wide association study (MGWAS) of bacterial and fungal biomarkers. Centered log-ratio (CLR) transformed abundances of microbial biomarkers were used as phenotypes in the MGWAS. SNP loci with $pw < 5 \times 10^{-5}$ were considered significant and are highlighted in red. The size of the dots indicates the significance of the association between bacterial taxa and barley SNP.

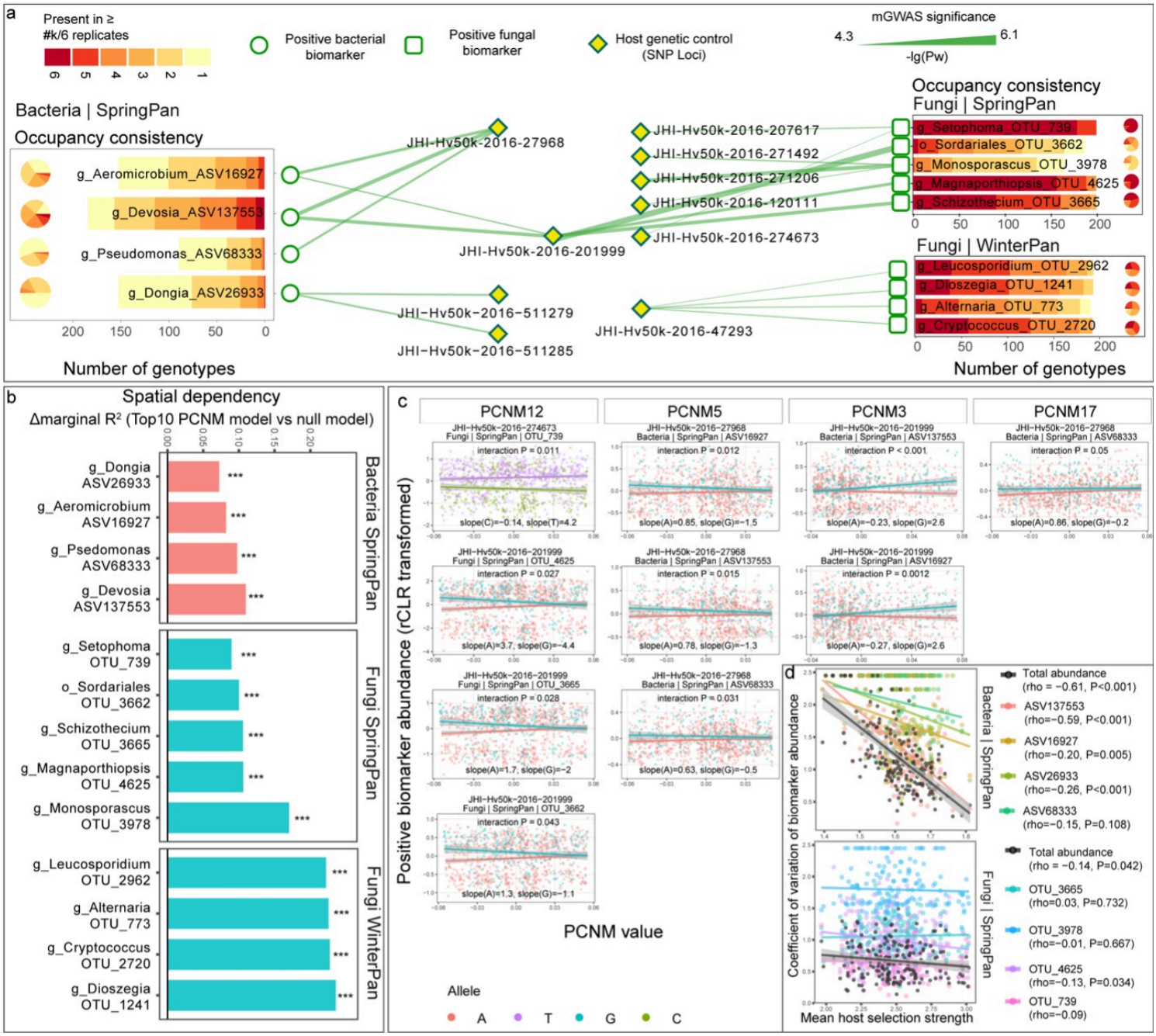


Figure 5

a) Network showing MGWAS associations between SNP loci and positive microbial biomarkers together with biomarker occupancy across host genotypes. Edges represent significant associations identified by MGWAS. Edge widths were correlated to association strength (-log10(p)). The bar plots summarize biomarker occupancy across genotypes as the number of genotypes in which a biomarker was detected in *k* out of six replicates, and the pie charts show the relative proportion of biomarkers at different occupancy levels. To minimize bias from sequencing depth, occupancy was calculated from the rarefied dataset.

b) Soil spatial effect on host-genetic associated biomarkers. Soil spatial effects were quantified for each microbial biomarker using a likelihood-ratio-test (LRT) based model comparison. A linear mixed-effects model including the top ten significant PCNMs as fixed effects and genotype as a random effect was compared with a genotype-only null model. LRT statistics were used to assess the significance and magnitude of soil spatial effects on individual biomarkers.

c) Allele-specific biomarker responses across spatial structure represented by PCNM axes. Representative biomarkers with significant SNP-by-PCNM interaction terms. For each biomarker, rclr-transformed abundance was modelled as a function of the associated SNP allele, the focal PCNM axis, and their interaction. Points represent individual samples and fitted lines show allele-specific regression slopes, illustrating how host genetic effects on biomarker abundance vary across spatial gradients.

d) Relationship between host selection strength and biomarker stability. Host selection strength was quantified as the Bray-Curtis dissimilarity between each root sample and its six nearest bulk soil samples. Biomarker stability was quantified as the coefficient of variation (CV) of biomarker abundance across replicate samples within genotype. Black lines indicate the association between host selection strength and the total abundance of positive biomarkers, whereas coloured lines indicate the corresponding relationships for individual biomarkers. Correlation was estimated by Spearman's correlation test.

Supplementary Files

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

- ExtendedDataTable1BiomarkermGWASsummary.xlsx
- ExtendedDataTable2Figure5A10SNPannotations.xlsx
- ExtendedDataTable3listofgenotypes.xlsx
- SupplemantaryMaterials.docx
- ExtendedDataFig1.pdf