IDENTIFYING PLANT GENES SHAPING MICROBIOTA COMPOSITION IN THE BARLEY RHIZOSPHERE

Carmen Escudero-Martinez[1,11], Max Coulter[1,2,11], Rodrigo Alegria Terrazas[1,3], Alexandre Foito[4], Rumana Kapadia[1], Laura Pietrangelo[1,5], Mauro Maver[1,6,7], Rajiv Sharma[8], Alessio Aprile[1,9], Jenny Morris[4], Pete Hedley[4], Andreas Maurer[10], Klaus Pillen[10], Gino Naclerio[5], Tanja Mimmo[6,7], James Abbott[2], Robbie Waugh[1,4], Geoffrey Barton[2*] and Davide Bulgarelli[1*]

[1]University of Dundee, Plant Sciences, School of Life Sciences, Dundee, UK; [2]University of Dundee, Computational Biology, School of Life Sciences, Dundee, UK; [3]Mohammed VI Polytechnic University, Agrobiosciences Program, Plant & Soil Microbiome Subprogram, Bengurir, Morocco; [4]The James Hutton Institute, Invergowrie, UK; [5]Department of Biosciences and Territory, University of Molise, Campobasso, Italy; [6]Faculty of Science and Technology, Free University of Bozen-Bolzano, Bolzano, Italy; [7]Competence Centre for Plant Health, Free University of Bozen-Bolzano, Bolzano, Italy; [8]Scotland's Rural College, Edinburgh, UK; [9]Department of biological and environmental sciences and technologies, University of Salento, Lecce, Italy; [10]Institute of Agricultural and Food Science and Plant Breeding, Martin Luther University, Halle-Wittenberg, Germany; [11]These authors contributed equally: Carmen Escudero-Martinez and Max Coulter.

*Correspondence: g.j.barton@dundee.ac.uk or d.bulgarelli@dundee.ac.uk

Abstract

A prerequisite to exploiting soil microbes for sustainable crop production is the identification of the plant genes shaping microbiota composition in the rhizosphere, the interface between roots and soil. Here we used metagenomics information as an external quantitative phenotype to map the host genetic determinants of the rhizosphere microbiota in wild and domesticated genotypes of barley, the fourth most cultivated cereal globally. We identified a small number of loci with a major effect on the composition of rhizosphere communities. One of those, designated the *QRMC-3HS* locus, emerged as a major determinant of microbiota composition. We then subjected soil-grown sibling lines harbouring contrasting alleles at *QRMC-3HS* and hosting contrasting microbiotas to comparative root RNA-seq profiling. This allowed us to identify three primary candidate genes, including a Nucleotide-Binding-Leucine-Rich-Repeat (*NLR*) gene in a region of structural variation of the barley genome. Our results provide novel insights into the footprint of crop improvement on the plant's capacity of shaping rhizosphere microbes.

Introduction

Plants thrive in association with diverse microbial communities, collectively referred to as the plant microbiota. This is capable of impacting the growth, development and health of their hosts[1–4]. The rhizosphere, the interface between the roots and soil[5], is a key microhabitat for the plant microbiota. For instance, similar to probiotics of the microbiota populating the digestive tract of vertebrates[6], microbes inhabiting the rhizosphere can promote plant growth by facilitating mineral nutrient uptake and pathogen protection[1,7–10].

These interactions do not represent stochastic events but are controlled, at least in part, by the plant genome[11,12]. Resolving the host genetic control of the microbes thriving at the root-soil interface therefore represents one of the pre-requisites for the rational manipulation of the plant microbiota for agriculture[13]. This is particularly relevant for the microbiota associated with crop wild relatives, which, having evolved under marginal soil conditions, may represent an untapped resource for low-input agriculture[14,15]. However, despite a footprint of domestication and crop selection having been identified in the taxonomic composition of the rhizosphere microbiota in multiple plant species[16–23], host genes underpinning this diversification remain poorly understood.

Barley is the fourth-most cultivated cereal globally[24] and an attractive experimental model to study plant-microbe interactions in the light of domestication and crop selection. We previously demonstrated that wild genotypes and ‘elite’ cultivated varieties host contrasting rhizosphere microbiotas[25,26]. Here, capitalising on an experimental population between barley genotypes at opposing ends of the domestication framework[27] and utilising state-of-the-art genomic[28] and transcriptomic[29] resources we gained novel insights into plant-microbiota interactions. We addressed three interconnected questions. Can we genetically map host determinants of the barley rhizosphere microbiota composition? If so, what kind of genes are putatively implicated in this process? And, finally, can we identify signatures of selection on these genes?

Results

The composition of the bacterial microbiota displays quantitative variation in the barley rhizosphere that appears to be controlled by a limited number of loci

We grew 52 genotypes of the progeny of a segregating population between the elite cultivar 'Barke' and a wild ancestor accession called HID-144 (see Methods), hereafter designated 'elite' and 'wild' respectively, in a soil previously used for investigating the barley rhizosphere microbiota [26,30,31] under controlled environmental conditions (see Methods). At early stem elongation (Supplementary Fig. 1), plants were removed from their substrates, and the rhizosphere fractions alongside bulk soil controls were subjected to an amplicon sequencing survey of the 16S rRNA gene.

Having defined a threshold for PCR reproducibility of individual amplicon sequencing variants (ASVs) (Supplementary Fig. 2), representing the terminal taxonomic nodes in our sequencing profiling, we inspected the impact of the parental lines on the composition of the bacterial microbiota. This allowed us to identify 36 ASVs discriminating between the parental lines (Wald test, individual *P*-value <0.05, FDR corrected, Fig. 1 a). Of these taxa, 27 ASVs, representing 5.39% of the reads, were enriched in and discriminated between the wild genotype and the elite variety (Wald test, individual *P*-value <0.05, FDR corrected, Fig. 1 a). Conversely, 9 ASVs representing 2.74% of the reads were enriched in and discriminated between the elite variety and the wild genotype (Wald test, individual *P*-value <0.05, FDR corrected, Fig. 1 a). This differential microbial enrichment between the parental lines was associated with a taxonomic diversification at phylum level: while the wild-enriched profiles encompassed several ASVs classified as Bacteroidota (33.9 % of the enriched reads), Firmicutes (32.8 %), Proteobacteria (27.9 %) Acidobacteria (3.7 %) and Myxococcota (1.3 %); the elite parent enriched predominantly for members of the phylum Actinobacteria (55.8 %), followed by members of Firmicutes (17.8 %), Proteobacteria (15.3 %), Bacteroidota (10.8 %) and Acidobacteria (0.3 %) (Supplementary Fig. 3). Next, we extended our survey to the entire segregating population. We made two observations. First, and consistent with previous investigations in the same reference soil[26,30,31], the rhizosphere communities were significantly different from unplanted soil controls as illustrated by sample separation along the x-axis of the Canonical analysis of Principal coordinates (CAP) (Adonis

test between bulk soil and rhizosphere samples, *P*-value <0.001, 5,000 permutations, Fig. 1 b). Despite identifying a significant impact of genotype on the composition of the rhizosphere samples, we failed to partition this variation into discrete classes (Adonis test, R^2 genotype among rhizosphere samples = 0.04, *P*-value 0.0006, 5,000 permutations, Fig. 1 b). Thus, while samples corresponding to the elite genotype clearly segregated from the wild genotype along the y-axis accounting for the second source of variation in a constrained ordination, individual segregants were distributed between the parental lines. This distribution mirrored the increased proportion of elite genome expected in the original back-crossed BC_1S_3 population, with the majority of microbiota profiles of segregating individuals spatially closer to the elite genotype (Fig. 1 b). These observations suggest that microbiota variation in the barley rhizosphere can be used as a trait in quantitative genetic studies.

We therefore used taxa that were differentially recruited between the parental lines and their abundances in the segregating population as “quantitative phenotypes” to search for significant associations with genetic markers located throughout the barley genome. To ascertain the phylogenetic congruence of the observed microbial trait we repeated this analysis at different taxonomic levels with sequencing reads agglomerated at genus and family level, respectively. For several bacteria we had previously characterised as being differentially abundant between the parental lines, we identified significant associations with individual homozygous or heterozygous alleles at multiple loci across the barley genome. These associations are supported either by marker regression or by a minimum LOD score of 3.43 at ASV, 3.56 at genus and 3.65 at family levels based on a LOD genome-wide significance threshold (false positive alpha level = 0.2; 1,000 permutations) (Fig. 2, Supplementary Fig. 4, Supplementary Table 1, Supplementary Database 1 and 2). However, one locus that mapped between 38.7 and 40.6 centimorgan (cM) on chromosome 3H was associated with the recruitment of phylogenetically unrelated bacteria at multiple taxonomic levels. The locus was identified as *QRMC.BaH144-3HS* where *QRMC* stands for QTL – Rhizosphere Microbiota Composition, *BaH144* corresponds to the cross Barke x HID144 and *3HS* the short arm of chromosome 3H, hereafter referred to as *QRMC-3HS*. All the bacteria recruited at *QRMC-3HS* were significantly enriched in the wild parent. We observed up to four unrelated ASVs representing

5.68% of the enriched bacterial reads linked to the *QRMC-3HS* locus classified as *Variovorax sp*., *Holophaga sp*., *Sorangium sp*. and *Tahibacter sp*. When taxa were agglomerated at the genus level, the number of significant associations increased to five, including the genus Rhodanobacter. The same analysis computed at family level revealed a congruent phylogenetic pattern associated with this locus represented by Comamonadaceae (the family of the genus Variovorax), Holophagaceae (Holophaga), Polyangiaceae (Sorangium) and Rhodanobacteraceae (Rhodanobacter). (Supplementary Table 1 and 2). *QRMC-3HS* was the only QTL recurrently found at different taxonomic levels presenting associations with up to five taxa across analyses with a LOD threshold established with more stringent criteria (false positive alpha level = 0.05; 1,000 permutations) (Supplementary Table 1 b). These results indicate that *QRMC-3HS* represents a major plant genetic determinant of microbiota recruitment in the barley rhizosphere.

Wild introgressions at *QRMC-3HS* are associated with compositional changes in rhizosphere bacterial microbiota

To validate the results of the mapping exercise, we tested whether barley lines harbouring contrasting alleles, *i.e.*, either 'elite' or 'wild', at *QRMC-3HS* would be associated with distinct microbial phenotypes. We generated two sibling lines designated 124_17, harbouring the elite alleles at *QRMC-3HS*, and 124_52, harbouring wild alleles at *QRMC-3HS* (Supplementary Fig. 1) by selfing the progeny of line HEB_15_124 which we identified as being heterozygous at the locus of interest (see Methods). Besides the genetic differences at *QRMC-3HS*, the derived lines share 95.5 % and 93.3 % of their genomes, respectively, with the elite parent based on molecular marker profiling using the barley 50k iSelect SNP Array[32] (Supplementary Database 3). They also represent *bona fide* progenies of the population investigated in this study (Supplementary Table 3).

We grew these sibling lines, along with the elite genotype and bulk soil controls, in the same experimental set-up described for the mapping experiment. We quantified 16S rRNA gene copy numbers for these rhizosphere and bulk soil samples as a first step towards a comparative microbiota profiling of the new material. This quantification showed no statistical differences of 16S RNA gene copy number among the tested genotypes (Kruskal-Wallis test and Dunn post-hoc test, *P*-value =

0.05, FDR corrected, Supplementary Fig. 5 a). We next inspected three ecological indices of alpha diversity, *i.e.*, within sample microbial diversity, namely observed ASVs, Chao1 and Shannon indices, proxies for microbiota richness and evenness. This analysis did not reveal significant differences between the communities inhabiting the rhizosphere of the sibling lines and those of Barke (ANOVA followed by Tukey post-hoc test, *P*-value > 0.05, FDR corrected, Supplementary Fig. 6). Conversely, we did identify a significant host-genotype component when we inspected beta-diversity, which is the between sample microbial diversity, a proxy for microbiota composition (Adonis test, R^2 genotype = 0.1188, *P*-value = 0.005, 5,000 permutations). This was manifested by the separation of the communities associated with the tested genotypes, in particular those of line 124_52 from those of the parental line Barke along the axis accounting for the major variation in a CAP (Fig. 3 a). We were, however, unable to determine which individual ASVs were responsible for the observed differentiation at the threshold imposed in the mapping experiment (Wald test, individual *P*-value > 0.05, FDR corrected).

These results nevertheless indicate that genetic variation at *QRMC-3HS* is associated with a significant shift in community composition in the rhizosphere. This trait is not driven by 16S rRNA gene copy number variation nor by differences in community richness and evenness. However, a wild Introgression at *QRMC-3HS* is not sufficient to trigger differential enrichments of individual bacteria.

Genetic variation at *QRMC-3HS* does not perturb the composition of the barley fungal microbiota

To investigate whether *QRMC-3HS* could shape the fungal communities inhabiting the rhizosphere, we carried out a similar sequencing survey using the rRNA ITS region. In common with the observed results for the bacterial counterpart, the evaluation of ITS copy number variation did not reveal significant differences between the sibling lines and the parental line Barke (Kruskal-Wallis test and Dunn post-hoc test, individual *P*-value > 0.05, FDR corrected, Supplementary Fig. 5 b).

Next, we generated an amplicon sequencing library using primers targeting the rRNA ITS region and identified a total of 216 fungal ASVs after applying filtering criteria (see Methods). When we implemented a beta-diversity analysis of the ITS library, we failed to identify a significant effect of the host genotype on these communities

(Adonis, *P*-value = 0.9628, 5,000 permutations). This was further manifested by the lack of spatial separation among microbiota samples of different genotypes in a CAP (Fig. 3 b) Likewise, no differentially recruited ASVs were identified in pair-wise comparisons using DESeq2 (Wald test, individual *P*-values > 0.05, FDR corrected).

These observations suggest that the selection pressure exerted by *QRMC-3HS* on the barley microbiota is predominantly confined to its bacterial members.

QRMC-3HS does not impact other root and yield traits.

To gain mechanistic insights into the plant traits associated with microbiota diversification, we examined root macro architecture, as morphological differences in barley roots can alter microbial composition in the rhizosphere[18,30]. When we measured nine different root morphology parameters of plants grown in the same soil used for microbiota characterisation, no significant differences were found between the sibling lines and the elite genotype (Kruskal–Wallis and post-hoc Dunn's test, *P*-values > 0.05, Supplementary Table 4).

Next, we grew the sibling lines and the elite genotype in sterile sand and determined the elemental composition of carbon and nitrogen in their exudates, as both of these elements represent another possible driver of microbial recruitment in the rhizosphere[33–35]. We selected two different timepoints: 2- and 3-weeks post-transplantation, to study the patterns of exudation. The former timepoint is critical for the establishment of the bacterial community in cereals[36], while the latter corresponds to the onset of stem elongation when the rhizosphere is harvested for microbial profiling[26,30,31]. As plants were supplemented with a nutrient solution (Methods), we used wash-through from unplanted pots as controls. The carbon content was significantly higher in the planted samples compared with the unplanted one regardless of the timepoint. For instance, unplanted samples wash-through contained just 1.5-3.6 % in carbon content weight (w/w), while that of plant exudates was 23-31 % in weight, although no significant differences among genotypes were identified (Kruskal–Wallis and post-hoc Dunn's test, *P*-value ≤0.05) (Supplementary Fig. 7 a). No significant effect was identified among time-points, regardless of the type of specimen investigated, *i.e.*, unplanted wash-through or planted exudates (Kruskal–Wallis and post-hoc Dunn's test, *P*-value >0.05) (Supplementary Fig. 7 a). Likewise, nitrogen content at the earliest timepoint (2 weeks) ranged from 2.8 to 6.0

% (w/w) and was not statistically different between soil wash-through or planted samples (ANOVA, *P*-value >0.05) (Supplementary Fig. 7 b). We could, however, differentiate the soil wash-through at the later timepoint (3 weeks), with higher nitrogen content of 10-18 % (w/w) in the soil wash-through, compared to the exudates of planted samples ranging from 4-6 % (w/w) compatible with the plant's uptake of this mineral from the nutrient solution. Within these latter specimens, no significant differences among the tested genotypes were identified (Kruskal–Wallis and post-hoc Dunn's test, *P*-value >0.05) (Supplementary Fig. 7 b).

We next explored the primary metabolism of the sibling lines and the elite genotype exudates at the onset of stem elongation stage (3 weeks) using gas chromatography–mass spectrometry (GC/MS). The metabolites recovered belong to categories such as amino acids, organic acids, carbohydrates and other polar compounds (Supplementary Fig. 8, Supplementary Database 4). Amongst carbohydrates, fructose and glucose represented the largest fraction of the exudates (Supplementary Fig.8 a). We found the majority of the compounds were classified as amino acids, with L-valine, L-leucine, isoleucine, L-glutamic and L-aspartic acid the more abundant (Supplementary Fig. 8 b). The main organic acid retrieved was succinic acid (Supplementary Fig. 8 c), while gamma-aminobutyric acid (GABA) was the most abundant in the 'other polar compounds' category. All of these compounds were present in comparable relative amounts regardless of the genotype, and the genotype effect on the metabolic composition of the exudates was consequently not significant (Adonis, *P*-value ≥0.05, 5,000 permutations).

To investigate any potential effect of *QRMC-3HS* on yield, we grew the sibling lines along with the elite cultivar Barke under the same conditions as the microbiota profiling to measure the thousand-grain weight (TGW) and main tiller grain weight in four independent experiments (Supplementary Fig. 9). Despite a batch effect identified for one of the replicated experiments, we observed a congruent trend where the elite material had higher yield than the sibling lines, regardless of their allelic composition at QRMC-3HS (ANOVA post-hoc Tukey, *P*-value ≤0.05) (Supplementary Fig. 9 a, b). We interpret this as indicating that *QRMC-3HS* alone may not be linked to the yield traits measured. Genes responsible for the seed dispersal attribute of wild barley spikes, designated brittle rachis[37] (HvBtr1–HvBtr2 locus at 40,451,507 – 40,710,518 bp on chromosome 3H) map physically near to

QRMC-3HS (33,181,340 – 36,970,860 bp on cultivar Barke) and may be implicated in the reduced yield observed in the sibling lines. We therefore used molecular markers (see Methods) to ascertain haplotype composition for these two genes. This revealed that both sibling lines and the wild parental line carry the wild allele as they do not show mutations in *Btr1* or *Btr2* (Supplementary Fig. 10 a, b). Conversely, our elite parent Barke and four other elite lines used controls, displayed a mutation in either of these genes, consistent with previous observations (Supplementary Fig. 10 a).

The sibling lines at *QRMC-3HS* and the cultivar Barke display distinct root transcriptional profiles

To further dissect the genetic mechanisms behind the differences in microbiota recruitment observed between the sibling lines 124_17 and 124_52, we conducted a comparative RNA-seq experiment using root tissue from the sibling lines and Barke. A total of 15 RNA-seq libraries were sequenced, with five biological replicates for each of the three genotypes (Supplementary Table 5). Three comparisons were made: 124_52 vs Barke, 124_17 vs Barke and 124_52 vs 124_17 using the BaRTv2[29] Barke transcriptome as a reference.

Consistent with the high genotypic similarity between the sibling lines and Barke, only 84 BaRTv2 genes were found to be differentially expressed (DE) in the 124_52 – Barke comparison, whilst 37 DE genes were identified in the 124_52 - 124_17 comparison. Interestingly, no DE genes were identified in the 124_17- Barke comparison, and all but three of the DE genes identified in the 124_52 - 124_17 comparison were also found in the 124_52 - Barke comparison (EdgeR, Individual *P*-value < 0.01, FDR corrected, Fig. 4). These results agree with the expectation that lines with the Barke *QRMC-3HS* allele (*i.e.*, the 124_17 sibling line and Barke) have similar transcriptional profiles, with changes in transcription possibly reflecting changes in microbiota compliment (see Fig. 3)

A contrasting microbial phenotype was observed between the sibling lines 124_52 and 124_17, despite their similarity at the genetic level (Fig. 3, Supplementary Database 1). Therefore, we decided to focus on the 34 DE genes shared between the 124_52 - 124_17 and 124_52 – Barke comparisons (Fig. 4), to identify gene products potentially shaping the bacterial microbiota. The full list of 34 DE genes is

found in Supplementary Database 3. Of the 34 DE genes in the 124_52 - 124_17 comparison, only two map within *QRMC-3HS*. The first of these genes is of unknown function, while the second is annotated as a nuclear binding leucine-rich-repeat like (*NLR*).

To understand how underlying genetic differences between 124_52 and 124_17 related to gene expression changes, the allelic composition of these two lines were compared at chromosome 3H (Fig. 5 b), and on the other 6 barley chromosomes (Supplementary Fig 11-16). We mined for regions of contrasting allelic composition in each of the seven chromosomes, and once identified, these were related back to the expression changes of genes expressed in the dataset (Fig. 5, Supplementary Fig.11 - 16). We found that the majority (31 of the 34 DE genes) were found on chromosome 3H, and that 28 of these were present at regions of the chromosome with contrasting alleles between 124_52 and 124_17 (Fig. 5). The three DE genes identified on other chromosomes were all found in regions where alleles between 124_52 and 124_17 do not differ. These results suggest that DE between 124_52 and 124_17 is predominantly due to cis-regulation or non-orthologous gene variation (presence/ absence variation), and that the number of trans-regulated genes is relatively small. A prediction of this observation is that differences in rhizosphere microbial phenotype between the two lines are not likely due to a large-scale reprogramming of the transcriptome. This is also reflected in the lack of differentially expressed transcription factors in the 34 124_52 - 124_17 DE genes (Supplementary Database 3).

Identification and prioritisation of *QRMC-3HS* candidate genes

A total of 59 genes were identified in the BaRTv2 gene/transcript[29] annotation within the bounds of the *QRMC-3HS* locus (identified as 33,181,340 – 36,970,860bp on chromosome 3H in cultivar Barke, Supplementary Database 4). Of these, 25 were found to be expressed in the RNA-seq dataset and were prioritised, as they are likely to be expressed in root tissue (Supplementary Database 4). As previously described, two out of these 25 genes were found to be DE in the 124_52 - 124_17 comparison subset of 34 genes (Supplementary Database 3).

To further prioritise candidate genes, variant calling was carried out to identify likely high impact and non-synonymous variants between lines 124_52 and 124_17. The variants were annotated in relation to the BaRTv2.18 annotation using SnpEff[38]. A detailed annotation of SNPs identified in each gene is shown in Supplementary Database 5. A total of 545 variants were identified across the 59 BaRTv2 genes annotated within *QRMC-3HS*. However, many of these variants were found in genes not expressed in our RNA-seq data, or were annotated as having low impact, meaning they are either synonymous changes or located in non-coding (5'/3' UTR) regions of genes and were therefore not considered as priority candidates. Two genes carried high impact variants. BaRT2v18chr3HG123110, of unknown function, has a frameshift variant and a missing stop codon. BaRT2v18chr3HG123140, annotated as a putative Xyloglucan endotransglucosylase/hydrolase enzyme (XTH) carries a frameshift variant close to the 5' end of the coding sequence (CDS) (Supplementary Fig. 17). Fourteen other genes had moderate impact variants, all of which have missense (non-synonymous) SNPs (Supplementary Database 5).

In summary, three genes in the *QRMC-3HS* locus were found to either be differentially expressed between two pair-wise comparisons, *i.e.*, 124_52 vs 124_17, and 124_52 vs. Barke, or have high impact variants and are therefore considered as primary candidates for shaping the barley rhizosphere microbiota (Supplementary Database 5).

Structural variation at *QRMC-3HS* in the barley pan-genome

The recent publication of the barley pan-genome[28] enabled us to investigate potential structural variants (SVs) at *QRMC-3HS*. These may affect gene presence or expression, and therefore may impact on candidate gene prioritisation. The genome sequence for Barke is represented in the pan-genome, although our wild parent is not. We initially compared the sequence across the *QRMC-3HS* locus in cv. Barke to the corresponding sequence in cv. Morex (Fig. 6 a). The alignment showed conserved synteny across the *QRMC-3HS* locus except close to the distal end, where a region of dissimilarity of approximately 480kb (Barke 3H:36,582,968 – 37,063,927bp) was identified (Fig. 6 a). To explore whether this was unique to the Barke - Morex comparison, we compared Barke to 14 other lines in the pan-genome (Supplementary Fig. 18). Comparisons of Barke with Golden Promise (Fig. 6 b),

Hockett and HOR13942 (Supplementary Fig. 18) showed continuous synteny across *QRMC-3HS*, whilst the other 12 comparisons, including that with the only wild genotype in the pan-genome (B1K-04-12) showed a break in synteny similar to that observed in Morex (Supplementary Fig. 18).

The putative *NLR* gene, BaRT2v18chr3HG123500, which was DE between 124_17 and 124_52 (Fig. 5, Supplementary Database 3), has a physical position on chromosome 3H at 36,880,423 – 36,890,887 bp, within this region of dissimilarity (Supplementary Databases 3, 4, Fig. 6 a). According to the pan-genome annotation, an ortholog of this gene is not present in Morex, RGT Planet or B1K-04-12. A closer look at the CPM values of the candidate *NLR* revealed that this gene is expressed at very low levels in 124_52 (Fig. 7 a).

To determine whether this low expression is due to the absence of the gene, we designed PCR primers specific to a region of the predicted gene BaRT2v18chr3HG123500 (Fig. 7 b). We further predicted, based on sequence comparisons (Fig. 6, Supplementary Fig. 18), that the gene would be absent from Morex and RGT Planet, but present in genotypes carrying an elite *QRMC-3HS* locus (Barke, Hockett, Golden Promise and 124_17) and so these were included as positive and negative controls, in addition to the wild parental line HID144. PCR results showed that an amplicon derived from the putative *NLR* gene was not detectable in RGT Planet and Morex, as anticipated from sequence comparisons, while it was present in Barke, Hockett and Golden Promise (Fig. 7 c). The amplicon was found to be present in both 124_52 and 124_17 as well as HID144, albeit with a different size product in HID144 and 124_52. This suggests that the difference in *NLR* gene expression between 124_52 and 124_17 may not be due to presence/absence but other polymorphisms in its genomic sequence. Regardless, the pan-genome comparisons identify the region at the distal end of the *QRMC-3HS* locus around the putative NLR as a region of sequence divergence.

Discussion

In the present study, we combined microbiota abundance and quantitative genetics to identify regions of the barley genome responsible for rhizosphere microbiota recruitment. Our results demonstrate that the taxonomic composition of the

rhizosphere microbiota can be treated as a quantitative trait whose genetic basis appear to display a footprint of human selection in the barley genome.

Our genetic mapping experiment demonstrated that the heritable component of the barley microbiota in the rhizosphere is controlled by a relatively low number of loci. This is congruent with observations of the bacterial communities inhabiting the phyllosphere of the model plant *A. thaliana*[39], the staple crop maize[40] and the cereal sorghum[41]. One of the loci identified in our study, designated *QRMC-3HS*, displays an association with several phylogenetically unrelated bacteria, with the notable exception of members of Actinobacteria. This mirrors the differential enrichment observed for members of this phylum between elite and wild barley material when grown in the same soil[26], and other modern/ancestral pairs[18]. A prediction from this observation is that a subset of soil bacteria may have evolved the capacity to engage with barley via a *QRMC-3HS*-mediated mechanism. This is in agreement with comparative bacterial genomics data which indicates that taxonomically diverse bacteria can share the same adaptive mechanisms to the plant environment[42,43]. An alternative, and not mutually exclusive scenario, is that *QRMC-3HS* mediates the recruitment of the so called 'hub' of the microbiota, *i.e.*, individual microorganisms capable of regulating the proliferation of other members of the community, as observed in Arabidopsis[44] and maize[45]. Mining metagenome-assemblies that are significantly associated to plant loci (Oyserman *et al.*, accompanying manuscript) as well as tapping into recently developed synthetic communities of the barley microbiota[46] will enable these scenarios to be investigated experimentally.

The development of powerful genetic and genomic resources allowed us to characterise *QRMC-3HS* at an unprecedented level. We made three important observations. First, the sibling lines harbouring contrasting homozygous alleles at *QRMC-3HS* host contrasting bacterial microbiotas. Despite these lines not triggering the significant enrichment of individual taxa observed in their parental lines, our approach indicates that host genetic composition is sufficient to predict an impact on overall rhizosphere community structure. Besides validating our genetic mapping, this observation is aligned to recent observations of sorghum genotypes[41]. Second, the same lines allowed us to determine that allelic variation at *QRMC-3HS* does not perturb the composition of fungal members of the community. Although this feature is distinct to observations in a genome-wide investigation of root-associated

communities of Arabidopsis[47], our observation is consistent with recently reported results for the rhizosphere microbiota of maize, where individual host genes shaped the bacterial but not the fungal microbiota[10]. Third, once we characterised these lines for additional traits that could be intuitively considered to be implicated in shaping the microbiota in the barley rhizosphere, such as root system architecture[30] and rhizodeposition of primary metabolites[48], we failed to identify significant associations between these traits and allelic composition at *QRMC-3HS*. While differences in the genetic background of the tested plants prevent us from drawing firm conclusions when considering these studies, our observations suggest that *QRMC-3HS* may code for a novel component of the host genetic control of barley microbiota recruitment.

We therefore employed a root RNA-seq experiment to gain mechanistic insights into the regulation of the microbiota mediated by *QRMC-3HS*. One of the candidate genes found to be significantly up-regulated in plants harbouring the elite allele at locus *QRMC-3HS* putatively encodes an NLR protein[49]. This class of protein represents one of the two main groups of immune receptors capable of selectively recognising and terminating microbial proliferation via effector recognition[50]. The gene we identified encodes a predicted protein consisting of a Rx-type coiled-coil-nucleotide-binding site-leucine-rich repeat (CC-NLR) domains, containing a putative integrated domain encompassing ankyrin repeats anchoring an NPR1-like (NONEXPRESSOR OF PATHOGENESIS-RELATED GENES 1) domain. This type of integrated *NLR* gene has recently been identified in the wheat genome[51,52]. The integrated domains may work as decoys, mimicking an effector target and enabling microbial recognition[53,54]. The *NPR1* gene is a key transcriptional regulator for plant defence responses related to the hormone salicylic acid (SA)[55]. Besides its canonical role in pathogen protection[56,57], it is important to note that *npr1* mutants, impaired in SA perception, fail to recruit a root microbiota comparable with their cognate wild-type plants[58]. A so-called resistance gene analogue sharing structural features with *bona fide NLR* genes has been identified among the candidate genes underpinning the establishment of the bacterial microbiota in the sorghum rhizosphere[41], further suggesting the possible significance of these genes for bacterial recruitment in the rhizosphere. A distinctive feature of the *NLR* gene identified in our study is that it lies in a region of structural variation of the barley genome: for instance, the cultivar

Morex, often used as a reference for microbiota investigations[25,26,31] lacks a copy of this gene. The use of the single barley reference genome available prior to 2020[59,60], would have prevented us from identifying this priority candidate gene.

Two other genes within the *QRMC-3HS* were considered among our primary candidates. The first is a gene that is differentially regulated between sibling lines harbouring contrasting microbiotas. As it encodes an unknown protein, we cannot hypothesis its mechanistic contribution to plant-microbe interactions in the rhizosphere. The second is a xyloglucan endotransglucosylase/hydrolase enzyme (XTH), characterised by a frame shift variant close to the 5' end of the CDS in line 124_52. XTH enzymes are widely conserved across plant lineages where they are responsible for cleavage and/or re-arrangement of xyloglucans[61,62], the most abundant hemicellulosic polysaccharides in primary cell walls[63]. In Arabidopsis, cell wall features are a recruitment cue for nearly half of the endogenous root microbiota[64] and cell wall modifications underpin some of the gene ontology categories identified in genome-wide association mapping experiments conducted with this plant[39]. A recent investigation conducted using so called 'split-root' approach demonstrated that genes encoding XTH were down-regulated in roots exposed to a high-density microbiome (*i.e.*, akin to natural soil conditions)[65]. Despite not identifying a significant phenotype of the macroscale root system architecture of our sibling lines, whereby XTH may play a primary role, this gene may still contribute to microbiota recruitment via modification of cell wall polysaccharides, a critical checkpoint in molecular plant-microbe interactions[66].

Taken together our data indicate that loci controlling microbiota recruitment in the barley rhizosphere were subjected to human selection during the development of current barley varieties. As microbiota profiling did not feature in these selection programmes, it is legitimate to hypothesize that genes shaping rhizosphere microbial communities are linked with other agronomic traits. The observation that *QRMC-3HS* is adjacent to a major QTL for yield-related traits previously identified on chromosome 3H using the same genetic material (*QRMC-3HS*; 38.7-40.6 cM; yield QTL, 40.7-43.9 cM)[67–69] may support this scenario.. Selection for yield traits may have inadvertently introduced a gene impacting the microbiota. An alternative, and not mutually exclusive, scenario is that disease resistance may have been the trait under agronomic selection. This would be in line with a recent investigation which

demonstrated that bacteria isolated from the barley rhizosphere mediated the establishment of both pathogenic and mutualistic fungi in roots[71]. In this scenario, selection at *QRMC-3HS* may contribute to the fine-tuning of these multitrophic interactions. Recent innovations in barley genetics[72–75] will facilitate the development of refined genetic material required to probe these scenarios experimentally.

In conclusion, by characterising an experimental population between wild and elite barley genotypes for rhizosphere microbiota composition, we have identified a putative major plant genetic determinant of the barley microbiota on chromosome 3H. Within the associated interval we have discovered three priority candidate genes, coding for an uncharacterised protein, an NLR and an XTH enzyme, respectively. These are putatively required for microbiota establishment in wild and cultivated barley genotypes. These latter two proteins have previously been implicated as putative regulators of the microbiota in other plant species. A striking observation derived from our investigation is that one of these candidate genes, the *NLR*, exists in a highly dynamic region of the barley genome, suggesting that selection for agronomic traits may have led to a divergent microbiota in elite cultivars. We therefore advocate the use of dedicated plant genetic resources to resolve plant-microbiota interactions at the gene level and accelerate their applications for sustainable crop production.

Figure Legends

Fig. 1: Barley microbiota composition displays a quantitative variation in a segregating population between wild and elite parental lines. **a**) Ternary plot depicting microbiota composition in the elite and wild genotypes as well as bulk soil samples. Each dot illustrates an individual ASV; the size of the dots is proportional to ASV's abundance while their position reflects the microhabitat where bacteria were predominantly identified. Individual dots are color-coded according to their significant enrichment in the rhizosphere of either parental line (Wald Test, Individual *P*-values <0.05, FDR corrected). **b**) Canonical Analysis of Principal Coordinates computed on Bray-Curtis dissimilarity matrix. Individual dots in the plot denote individual biological replicates whose colour depict sample type in the bottom part of the figure. The number in the plot depicts the proportion of variance (R^2) explained by the factor 'Sample' within the rhizosphere microhabitat, *i.e.*, Elite, Wild or Segregant. The

asterisk associated to the R^2 value denote its significance, *P*-value 'Sample' = 0.0006; Adonis test 5,000 permutations.

Fig. 2: Genetic map of the barley determinants of the rhizosphere bacterial microbiota. Circos plot depicting **a**) the seven barley chromosomes **b**) grey connector lines link the physical position of SNPs with the genetic position in cM as indicated in the outer part of the ring; numbers in black within the individual chromosome define genetic positions (cM) significantly associated (using the function *scanone* implementing interval mapping with a single-QTL model, expectation-maximization algorithm, LOD genome-wide significance threshold 20 % adjusted per taxa, 1,000 permutations) to the differential enrichment of individual **c**) ASVs, **d**) genus or **e**) family, respectively. Different shapes depict taxonomic assignment at phylum level. Shapes are colour-coded according to the microbiota of the parental line where individual taxa were identified.

Fig. 3: Wild alleles at locus *QRMC-3HS* are associated with a shift in the composition of the bacterial, but not fungal, microbiota. Canonical Analysis of Principal Coordinates computed on Bray-Curtis dissimilarity matrix of **a**) bacterial or **b**) fungal ASVs' abundances. Sample type is depicted in the bottom part of the figure. The number in the plots show the proportion of variance (R^2) explained by the factors 'Batch' and 'Genotype', respectively. Asterisks associated to the R^2 value denote its significance. **a**) *P*-value 'Batch' = 0.2775; *P*-value 'Genotype' = 0.0050; Adonis test 5,000 permutations. **b**) *P*-value 'Batch' = 0.0268; *P*-value 'Genotype' = 0.9628; Adonis test 5,000 permutations.

Fig. 4: The sibling lines harbouring contrasting alleles at locus *QRMC-3HS* and the cultivar Barke display distinct root transcriptional profiles. Venn diagram showing the number of differentially expressed genes among pairs of comparisons between the sibling lines 124_52, 124_17 and their elite parent Barke (EdgeR pair-wise comparison, individual *P*-values <0.01, FDR corrected).

Fig. 5: Differentially expressed genes mapping at locus *QRMC-3HS*. a) Dots depict individual genes and their expression pattern in the pair-wise comparison 124_17 vs. 124_52 ($\log_2$ Fold-Change), colour-coded according to their significance as illustrated at the bottom of the figure (EdgeR, individual *P*-values <0.01, FDR corrected). **b**) Projection of the individual genes on the structures of chromosome 3H

for the lines 124_17 and 124_52, respectively, color-coded according to allelic composition as indicated in the key at the bottom of the figure. The physical location of locus *QRMC-3HS* is highlighted in pale pink.

Fig. 6: Locus *QRMC-3HS* defines an area of structural variation in the barley genome. Alignment visualisation of the sequence at and surrounding the *QRMC-3HS* locus comparing **a**) the cultivars Barke and Morex and **b**) Barke to cultivar Golden Promise. The *QRMC-3HS* locus is shown in white, while purple dots represent sequencing matches longer than 1,000 bp and > 95% identity. The gap in the diagonal in **a**) denotes a disruption of synteny between the two genotypes. Numbers on the axis denote the physical interval, in bp, analysed in the given genomes.

Fig. 7: The *NLR* gene associated with genotype-dependent transcriptional and genomic variations. a) Boxplot showing the root RNA-seq *NLR* expression across the elite and the sibling lines in normalized counts per million. **b**) Schematic representation of the *NLR* gene transcripts inferred from the RNA-seq data depicting predicted protein domains from InterProScan.. Arrowhead depicts predicted amplicon site of the PCR marker **c**) PCR amplicons covering the intron between the LRR and the ankyrin domains in the indicated genotypes-Negative control (NC).

Methods

Plants

Barley plants from family 15 of the nested-association mapping population (NAM) HEB-25[27] were used in this investigation. We developed the sibling lines 124_52 and 124_17, with contrasting haplotypes wild and elite at the *QRMC-3HS* locus, by selfing the line HEB_15_124 which was heterozygous at the locus of interest. All lines used were genotyped using a combination of KASP markers and Infinium iSelect 9K and 50K SNPs arrays platforms. Barley plants used in this study along with the genetic information are summarized in Supplementary Database 1 and 3.

Plants growth conditions and rhizosphere fractionation

Barley seeds were surface sterilized and pre-germinated as previously reported[26]. Germinated seeds with comparable rootlets were sown in individual 400 mL pots filled with a sieved (15 mm) reference agricultural soil previously used for barley-microbiota investigation and designated Quarryfield[26,30,31]. The number of replicates varies according to the experiment: the mapping experiment n=4 for the parental lines and n=2 for each of the segregating lines, whereas in the sibling lines we used n=10. Unplanted soil pots were included in each experiment and designated bulk soil controls. Plants were grown until stem elongation (~5 weeks, BBCH15 cereal growth stage) under controlled environmental conditions as described in[26,30,31]. At this developmental stage, plants were uprooted from the soil, stems were detached from the uppermost 6 cm of the root system which, upon removal of large soil aggregates, was subjected to a combination of washing and vortexing to dislodge rhizosphere fractions as previously described[26,30,31].

Assessment of root and yield traits

Agronomic traits related to yield, brittle rachis and root architecture were assessed for the sibling lines. Main tiller seeds grown in Quarryfield soil were used to measure yield with a MARVIN seed analyser (Supplementary Fig. 9). Brittle rachis in *Btr1* and *Btr2* gene mutations were assessed using KASP markers[76]. Root architecture factors were studied for plants at early stem elongation for consistency with the microbial rhizosphere experiments (Supplementary Fig.1). Roots were thoroughly washed and kept in phosphate buffered saline solution (PBS) until processing. Root systems were scanned and analysed using WinRHIZO software (Regent Instruments Inc.). Shoot and root dry weights were determined by drying the samples in the oven at 70 °C for 48 h. Specific root length (cm/g) and specific root area (cm^2/g) were calculated according to [18]. Normality was assessed by Shapiro-Wilk test. Significance was tested with a Kruskal-Wallis with post-hoc Dunn test (Bonferroni corrected) or an ANOVA test followed by a Tukey post hoc test according to normality.

Barley root exudates metabolic profiling

Barley plants were germinated as described above. After germination, 3 barley plantlets were sown in a 400 mL plastic pot filled with approx. 300 g of pre-sterilized silver sand and organised in the glasshouse in randomized blocks design (n=15 per

genotype/block). Barley nutrient solution 100% was applied to each pot (50 mL at 48 h intervals) and in the last week a 25 % nitrogen barley nutrient solution[77] was applied. After 2 and 3 weeks and following the randomized block design, barley roots were carefully taken out of the pots, and the sand around the roots washed off with water. Using a plastic jar (100mL vol), root exudates were collected using 3 plants from the same genotype per jar. The washed plant roots were submerged in 50 mL sterilized distilled water and were left to exude for 6 to 7 h. Unplanted controls were generated by washing trough the unplanted sand with sterilized distilled water collecting 50 mL, which were processed identically to the exudates. The root exudates (100 mL) and unplanted controls were collected in clean plastic jars and filtered (cellulose Whatman No. 42) and 100 µl of 2 mg/mL erythritol solution (extraction standard for GC/MS) was added to each jar. The exudate solution was frozen at -80 °C and subsequently concentrated to powder form by freeze-drying for 4 days. The experiment was harvested on four consecutive days approximately between 11 AM and 6 PM. Freeze-dried exudates (5 mg) were pooled per genotype and block (n=4) and analysed by an Elemental analyser for total carbon and nitrogen quantification using the Dumas method[78], while 20 mg of the same samples were used to perform a semiquantitative GC/MS analysis as previously described[79].

Metabolite profiles were acquired using a GC–MS (DSQII Thermo-Finnigan, UK) system carried on a DB5-MS column (15 m x 0.25 mm x 0.25 µm; J&W, Folsom, CA, USA). Data were acquired using the XCALIBUR (Thermo Scientific, Waltham, MA, USA) software package V. 2.0.7. Semiquantitative data was acquired by integrating selected ion chromatographic peaks.

Bacterial and fungal DNA quantification

Bacterial and fungal DNA fractions were quantified in the rhizosphere of the sibling lines by quantitative real-time polymerase chain reaction. Rhizosphere DNA samples were diluted to 10 ng/µL and serial dilutions applied. A final concentration of 0.1 ng/µL was employed for both the Femto Fungal DNA Quantification Kit and Femto Bacterial DNA Quantification Kit (Zymo Research) according to the manufacturer protocol. The sibling lines DNA samples were randomized in the 96 well plates, using a minimum of 10 biological replicates per treatment. Quantification was performed in a StepOne thermocycler (Applied Biosystems by Life Technology).

Preparation of 16 rRNA gene and / ITS amplicon pools and Illumina 16S rRNA and ITS gene amplicon sequencing

16S rRNA and ITS data from the barley rhizospheres were determined using previously described protocols[26,30,31]. The 515F (GTGCCAGCMGCCGCGGTAA)-806R (GGACTACHVGGGTWTCTAAT) primer pair[80] were used for amplifying 16S rRNA sequences, while the PCR primers ITS1f (CTTGGTCATTTAGAGGAAGTAA)-ITS2 (GCTGCGTTCTTCATCGATGC) were used for the ITS library.[81,82].

Paired-end Illumina sequencing was performed using the Illumina MiSeq system (2x 150 bp reads) as indicated in [80]. Library pool quality was assessed using a Bioanalyzer (High Sensitivity DNA Chip; Agilent Technologies) and quantified using a Qubit (Thermo Fisher) and qPCR (Kapa Biosystems, Wilmington, USA). Amplicon libraries were spiked with 15% of a 4 pM phiX control solution. The resulting high-quality libraries were run at 10 pM final concentration.

Amplicon sequencing reads processing

Quality assessment and DADA2 version 1.10[83] and R 3.5.1[84] was used to generate the ASVs following the basic methodology outlined in the 'DADA2 Pipeline Tutorial (1.10)' and it is explained in detail in[31]. Subsequently, sequences classified as 'Chloroplast' or 'Mitochondria' from the host plant were pruned in silico. Additional pruning was carried out, removing ASVs matching a list of potential contaminants of the lab[85]. Next, we merged the library used for genetic mapping with the library of the sibling lines to perform simultaneous processing of both libraries creating a single new Phyloseq object. Further filtering criterion was applied, low count ASVs were pruned from the merged libraries (at least 20 reads in 2% of the samples), retaining 93% of the initial reads (Supplementary Fig. 1). This dataset was rarefied at equal sequencing depth across samples (10500 reads) and agglomerated at genus and family taxonomic levels. Finally, the resulting Phyloseq object was subsetted by the corresponding library for downstream analyses.

The mapping 16S rRNA gene amplicon library merged with the sibling lines library allowed us to identify 2,189 individual ASVs from a total of 8,219,883 sequencing reads after filtering. While the sibling lines ITS rRNA amplicon sequencing library was generated identifying 216 individual ASVs from a total of 4,641,285 sequencing reads after filtering.

Calculation of alpha-, beta-diversity indices and differential abundance between rhizospheres

Alpha-diversity richness was estimated as described in[31]. Beta-diversity analysis was carried out by calculating the dissimilarities among microbial communities using the rarefied data with the Bray-Curtis index as described in[31]. For ITS, the Bray-Curtis dissimilarity matrix was square root transformed to allow visualization since all the samples appeared agglomerated in the PCoA visualization. DESeq2 was used to perform microbial differential abundance analysis to identify genera differentially enriched between pair-wise comparisons by Wald test (False Discovery Rate, FDR < 0.05)[86].

QTL mapping of bacterial microbiota phenotype

Following the analysis of microbial differential abundance (DESeq2) between wild and elite parental lines, ASVs enriched in the wild or elite parents were recapitulated in the segregant population for further mapping. Association between microbial abundances and loci of the barley genome was conducted using the package R/qtl[87] and the function *scanone*, with the expectation-maximization (EM) algorithm implementing interval mapping considering only a single-QTL model. The LOD genome-wide significance threshold was set at 20 % adjusted per taxa using 1,000 permutations. The loci, shown in Fig. 2, were selected based on marker regression analysis or their LOD scores genome adjusted per taxa (Supplementary Table 1) (functions *makeqtl*, *fitqtl* and *plotPXG*). The delimitation of the different loci was performed by applying the Bayes credible interval method with confidence intervals at 95 % (function *bayestint*[88]) (Supplementary Table 2). The percentage of explained variance (R^2) was calculated per individual phenotype (taxa), at the flanking maker of the interval upper part, corresponding to each of the individual taxa mapping at this position, which is summarized in Supplementary Table 2.

Transcriptomic analyses (RNA-seq) of the sibling lines

Roots from cv. Barke and the sibling lines were processed as described above. Briefly, the different genotypes (n=10) were grown in a random matrix in pots filled with Quarryfield soil and maintained in the glasshouse for 5 weeks in a randomised arrangement. The uppermost 6 cm of the root system were processed as described in ‘rhizosphere fractionation’. For harvesting root samples, following vortexing the

root system to remove the soil/rhizosphere fractions in PBS, roots were collected with sterile forceps, excess PBS gently removed using a clean paper towel and immediately flash-frozen in liquid nitrogen until processed. All the root systems were collected in three consecutive days between 10 AM and 4 PM, reflecting the time necessary to process root samples.

RNA was extracted from individual root systems with the Macherey-Nagel™ NucleoSpin™ RNA, Mini Kit (Thermo Fisher, USA) following the manufacturer's protocol, including the Plant RNA Isolation Aid Invitrogen (Thermo Fisher, USA) for the sample lysis step using 90 µL of Plant RNA Isolation Aid, mixed with 870 µL of RA1 buffer and 40 µL of dithiothreitol (DTT). RNA quality was assessed using an Agilent 2100 Bioanalyzer or TapeStation (Agilent, USA). Samples for sequencing were selected based on microbiota profiles and quality of the RNA sample.

Approximately, 2 µg of total RNA per sample (n=18) was submitted to Genewiz for Illumina sequencing. Total RNA (300 ng) was further purified using the NEBNext mRNA Magnetic Isolation Module (NEB). Library preparation was carried out using the NEBNext® Ultra™ II Directional RNA Library Prep with Sample Purification Beads and indexed with NEBNext Multiplex Oligos for Illumina (96 Unique Dual Index Primer Pairs set 1). Next-Generation sequencing was carried out on an Illumina NovaSeq 6000 using an SP, 300 cycles, flow cell with 2 x 150 bp paired-end reads. The library was stranded with a sequencing depth of 40M reads per sample.

Differential expression analysis of RNA-seq

Downstream data pre-processing and analysis for both transcript quantification and variant calling was caried out using snakemake[89].

Raw reads were trimmed using trim galore (https://github.com/FelixKrueger/TrimGalore) version 0.6.6 with parameters "-q 20 –Illumina –paired".

Transcript quantification was carried out using Salmon[90] version 1.4.0 using parameters "-l A --seqBias –posBias –validateMappings" with BaRTv2.18[29] as the reference transcriptome. Analysis of RNA-seq quantifications was carried out using a custom modified version of the 3D RNAseq pipeline[91]. The tximport R package

version 1.10.0 was used to import transcript TPM values and generate gene TPM values[92] . Low expressed transcripts and genes were filtered based on analysing the data mean-variance trend. The expected decreasing trend between data mean and variance was observed when transcripts which had < 3 of the 15 samples with counts per million reads (CPM) of 2 removed, which provided an optimal filter of low expression. A gene was counted as expressed and included in the downstream differential expression (DE) analysis if any of its transcripts met the above criteria. The TMM method was used to normalise the gene and transcript read counts to -CPM[93]. The R package umap (https://cran.r-project.org/web/packages/umap/vignettes/umap.html) implementing the umap algorithm[94] was used to visualise the expression data. It was found that sample date influenced gene expression and so this was incorporated into the EdgeR linear model as a block effect.

DE analysis was carried out using the R package EdgeR[95] version 3.32.0. The EdgeR generalised linear model quasi-likelihood (glmQL) method was used, with genotype and date of sampling used as terms in the model (~0 + genotype + sampling.date). Contrast groups were set to 124_17 – Barke, 124_52 – Barke and 124_52 – 124_17. *P*-values were corrected using the Benjamini – Hochberg method to correct false discovery rate[96]. Genes were considered to be DE if they had an adjusted *P*-value < 0.01 and a Log2FC >=1 or <=-1.

GO enrichment analysis was carried out using GO terms from the BaRTv2 transcriptome annotation[29]. A fisher's exact test was used to test for significant enrichment. The Benjamini – Hochberg method was used to correct the false discovery rate[96].

Variant calling

For variant calling the trimmed Illumina reads were combined according to genotype (Barke, 124_17 or 124_52) using the Linux cat command (forward and reverse reads in separate files). Mapping was carried against the barley cv. Barke genome[28] with STAR version 2.7.9[97]. To aid with read mapping, a BaRTv2.18 gtf file was used with the “genomeGenerate” mode. After an initial round of mapping was carried out,

splice junctions from each of the genotypes were collated and filtered using a custom script, removing splice junctions with non-canonical dinucleotide sequences, those with a read depth < 4 and a max overhang < 10 bp. The filtered splice junction set was used as input for a second round of mapping.

Mapped read files (.bam files) were pre-processed prior to variant calling using Opossum[98] with settings "SoftClipsExist True". Variant calling was carried out using Platypus[99] with the barley Barke genome as reference, and with settings "--filterDuplicates 0 --minMapQual 0 --minFlank 0 --maxReadLength 500 --minGoodQualBases 10 --minBaseQual 20". VCF files were merged using bcftools merge and filtered to remove variants outside the *QRMC-3HS* locus (Barke chr3H: 33,181,340 – 36,970,860). The resulting *QRMC-3HS* locus VCF was filtered to only keep variants above the quality threshold, and where genotypes Barke and 124_17 were called as reference allele and 124_52 was called as the alternative allele. InterProScan version 5.48-83.0 (version 83.0 data) was used to predict functional domains of predicted proteins from transcripts.

MuMMER alignment

In order to ascertain the physical position of the *QRMC-3HS* locus in each genome, the sequences of two markers flanking the *QRMC-3HS* locus (SCRI_RS_141171 and SCRI_RS_154747) were aligned to reference genomes from the pan-genome using BLAST[28,100] with default parameters. The best alignment for each flanking marker was selected as the physical position in each case. For the purposes of visualisation, these numbers were rounded to the nearest Mb. The sequence +/- 2Mb on either side of the flanking markers was extracted from each genome using a custom python script. The program nucmer from the MuMMER suite[101] was used to align the *QRMC-3HS* locus sequence from each of the pan-genome lines against the *QRMC-3HS* locus sequence of Barke with settings "dnadiff". The resulting delta file was filtered using delta-filter with the settings "-l 95 -l 1000 -g", resulting in all alignments with less than 95% identity and lengths of less than 1,000 bp being removed. The program mummerplot was then used to create figures.

Data Availability

The sequences reported in this manuscript have been deposited in the European Nucleotide Archive, https://www.ebi.ac.uk/ (accession number PRJEB48857).

Code Availability

The codes to reproduce the figures and statistical analyses reported in this manuscript are available at GitHub https://github.com/BulgarelliD-Lab/Microbiota_mapping.

References

1. Berendsen, R. L. *et al.* Disease-induced assemblage of a plant-beneficial bacterial consortium. *ISME J. 2018 126* **12**, 1496–1507 (2018).
2. Hacquard, S., Spaepen, S., Garrido-Oter, R. & Schulze-Lefert, P. Interplay Between Innate Immunity and the Plant Microbiota. *Annu. Rev. Phytopathol.* **55**, 565–589 (2017).
3. Lu, T. *et al.* Rhizosphere microorganisms can influence the timing of plant flowering. *Microbiome* **6**, (2018).
4. Verbon, E. H. & Liberman, L. M. Beneficial Microbes Affect Endogenous Mechanisms Controlling Root Development. *Trends Plant Sci.* **21**, 218–229 (2016).
5. York, L. M., Carminati, A., Mooney, S. J., Ritz, K. & Bennett, M. J. The holistic rhizosphere: integrating zones, processes, and semantics in the soil influenced by roots. *J. Exp. Bot.* **67**, 3629–3643 (2016).
6. Kamada, N., Chen, G. Y., Inohara, N. & Núñez, G. Control of Pathogens and Pathobionts by the Gut Microbiota. *Nat. Immunol.* **14**, 685 (2013).
7. Lugtenberg, B. & Kamilova, F. Plant-Growth-Promoting Rhizobacteria. *http://dx.doi.org/10.1146/annurev.micro.62.081307.162918* **63**, 541–556 (2009).
8. Alegria Terrazas, R. *et al.* Plant-Microbiota Interactions as a Driver of the Mineral Turnover in the Rhizosphere. *Adv. Appl. Microbiol.* **95**, 1–67 (2016).
9. Stringlis, I. A. *et al.* MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E5213–E5222 (2018).
10. Yu, P. *et al.* Plant flavones enrich rhizosphere Oxalobacteraceae to improve

maize performance under nitrogen deprivation. *Nat. plants* **7**, 481–499 (2021).

11. Bulgarelli, D. *et al.* Structure and Functions of the Bacterial Microbiota of Plants. (2013). doi:10.1146/annurev-arplant-050312-120106
12. Hacquard, S. *et al.* Microbiota and host nutrition across plant and animal kingdoms. *Cell Host Microbe* **17**, 603–616 (2015).
13. Schlaeppi, K. & Bulgarelli, D. The Plant Microbiome at Work. *http://dx.doi.org/10.1094/MPMI-10-14-0334-FI* **28**, 212–217 (2015).
14. Escudero-Martinez, C. & Bulgarelli, D. Tracing the evolutionary routes of plant-microbiota interactions. *Curr. Opin. Microbiol.* **49**, 34–40 (2019).
15. Cordovez, V., Dini-Andreote, F., Carrión, V. J. & Raaijmakers, J. M. Ecology and Evolution of Plant Microbiomes. *https://doi.org/10.1146/annurev-micro-090817-062524* **73**, 69–88 (2019).
16. Edwards, J. *et al.* Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E911–E920 (2015).
17. Leff, J. W., Lynch, R. C., Kane, N. C. & Fierer, N. Plant domestication and the assembly of bacterial and fungal communities associated with strains of the common sunflower, Helianthus annuus. *New Phytol.* **214**, 412–423 (2017).
18. Pérez-Jaramillo, J. E. *et al.* Linking rhizosphere microbiome composition of wild and domesticated Phaseolus vulgaris to genotypic and root phenotypic traits. *ISME J. 2017 1110* **11**, 2244–2257 (2017).
19. Pérez-Jaramillo, J. E. *et al.* Deciphering rhizosphere microbiome assembly of wild and modern common bean (Phaseolus vulgaris) in native and agricultural soils from Colombia. *Microbiome* **7**, 1–16 (2019).
20. Spor, A. *et al.* Domestication-driven changes in plant traits associated with changes in the assembly of the rhizosphere microbiota in tetraploid wheat. *Sci. Reports 2020 101* **10**, 1–12 (2020).
21. Hassani, M. A., Özkurt, E., Franzenburg, S. & Stukenbrock, E. H. Ecological assembly processes of the bacterial and fungal microbiota of wild and domesticated wheat species. *Phytobiomes J.* **4**, 217–224 (2020).
22. Tkacz, A. *et al.* Agricultural Selection of Wheat Has Been Shaped by Plant-Microbe Interactions. *Front. Microbiol.* **11**, 132 (2020).
23. Wipf, H. M. L. & Coleman-Derr, D. Evaluating domestication and ploidy effects on the assembly of the wheat bacterial microbiome. *PLoS One* **16**, e0248030 (2021).

24. Newton, A. C. *et al.* Crops that feed the world 4. Barley: a resilient crop? Strengths and weaknesses in the context of food security. *Food Secur. 2011 32* **3**, 141–178 (2011).
25. Davide Bulgarelli, A. *et al.* Structure and Function of the Bacterial Root Microbiota in Wild and Domesticated Barley Resource Structure and Function of the Bacterial Root Microbiota in Wild and Domesticated Barley. *Cell Host Microbe* **17**, 392–403 (2015).
26. Alegria Terrazas, R. *et al.* A footprint of plant eco-geographic adaptation on the composition of the barley rhizosphere bacterial microbiota. *Sci. Reports 2020 101* **10**, 1–13 (2020).
27. Maurer, A. *et al.* Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genomics 2015 161* **16**, 1–12 (2015).
28. Jayakodi, M. *et al.* The barley pan-genome reveals the hidden legacy of mutation breeding. *Nat. 2020 5887837* **588**, 284–289 (2020).
29. Coulter, M. *et al.* BaRTv2: A highly resolved barley reference transcriptome for accurate transcript-specific RNA-seq quantification. *bioRxiv* 2021.09.10.459729 (2021). doi:10.1101/2021.09.10.459729
30. Robertson-Albertyn, S. *et al.* Root hair mutations displace the barley rhizosphere microbiota. *Front. Plant Sci.* **8**, 1094 (2017).
31. Maver, M. *et al.* Applications of the indole-alkaloid gramine modulate the assembly of individual members of the barley rhizosphere microbiota 2 3. doi:10.1101/2020.12.07.414870
32. Bayer, M. M. *et al.* Development and evaluation of a barley 50k iSelect SNP array. *Front. Plant Sci.* **8**, 1792 (2017).
33. Bais, H. P., Weir, T. L., Perry, L. G., Gilroy, S. & Vivanco, J. M. THE ROLE OF ROOT EXUDATES IN RHIZOSPHERE INTERACTIONS WITH PLANTS AND OTHER ORGANISMS. *http://dx.doi.org/10.1146/annurev.arplant.57.032905.105159* **57**, 233–266 (2006).
34. Vives-Peris, V., de Ollas, C., Gómez-Cadenas, A. & Pérez-Clemente, R. M. Root exudates: from plant to rhizosphere and beyond. *Plant Cell Reports 2019 391* **39**, 3–17 (2019).
35. Fagorzi, C. *et al.* Nonadditive Transcriptomic Signatures of Genotype-by-

Genotype Interactions during the Initiation of Plant-Rhizobium Symbiosis. *mSystems* **6**, (2021).

36. Li, X., Rui, J., Mao, Y., Yannarell, A. & Mackie, R. Dynamics of the bacterial community structure in the rhizosphere of a maize cultivar. *Soil Biol. Biochem.* **68**, 392–401 (2014).
37. Takahashi R, H. J. Linkage study of two complementary genes for brittle rachis in barley. *Ber Ohara Inst Landwirtsch Biol Okayama Univ* **12**–**99**, 105 (1964).
38. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin).* **6**, 80–92 (2012).
39. Horton, M. W. *et al.* Genome-wide association study of Arabidopsis thaliana leaf microbial community. *Nat. Commun. 2014 51* **5**, 1–7 (2014).
40. Wallace, J. G., Kremling, K. A., Kovar, L. L. & Buckler, E. S. Quantitative genetics of the maize leaf microbiome. *Phytobiomes J.* **2**, 208–224 (2018).
41. Deng, S. *et al.* Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *ISME J. 2021 1511* **15**, 3181–3194 (2021).
42. Bai, Y. *et al.* Functional overlap of the Arabidopsis leaf and root microbiota. *Nat. 2015 5287582* **528**, 364–369 (2015).
43. Levy, A., Conway, J. M., Dangl, J. L. & Woyke, T. Elucidating Bacterial Gene Functions in the Plant Microbiome. *Cell Host Microbe* **24**, 475–485 (2018).
44. Agler, M. T. *et al.* Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. *PLOS Biol.* **14**, e1002352 (2016).
45. Niu, B., Paulson, J. N., Zheng, X. & Kolter, R. Simplified and representative bacterial community of maize roots. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E2450–E2459 (2017).
46. Robertson-Albertyn, S. *et al.* A genome-annotated bacterial collection of the barley rhizosphere microbiota. *bioRxiv* 2021.03.10.434690 (2021). doi:10.1101/2021.03.10.434690
47. Bergelson, J., Mittelstrass, J. & Horton, M. W. Characterizing both bacteria and fungi improves understanding of the Arabidopsis root microbiome. *Sci. Rep.* **9**, (2019).
48. Mwafulirwa, L. *et al.* Identification of barley genetic regions influencing plant-microbe interactions and carbon cycling in soil. (123AD). doi:10.1007/s11104-

021-05113-6

49. van Wersch, S., Tian, L., Hoy, R. & Li, X. Plant NLRs: The Whistleblowers of Plant Immunity. *Plant Commun.* **1**, 100016 (2020).

50. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nat. 2006 4447117* **444**, 323–329 (2006).

51. Bailey, P. C. *et al.* Dominant integration locus drives continuous diversification of plant immune receptors with exogenous domain fusions. *Genome Biol.* **19**, 1–18 (2018).

52. Wang, H., Zou, S., Li, Y., Lin, F. & Tang, D. An ankyrin-repeat and WRKY-domain-containing immune receptor confers stripe rust resistance in wheat. *Nat. Commun. 2020 111* **11**, 1–11 (2020).

53. Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T. & Dodds, P. N. A novel conserved mechanism for plant NLR protein pairs: The 'integrated decoy' hypothesis. *Front. Plant Sci.* **5**, 25 (2014).

54. Wu, C. H., Krasileva, K. V., Banfield, M. J., Terauchi, R. & Kamoun, S. The"sensor domains"of plant NLR proteins: More than decoys? *Front. Plant Sci.* **6**, 134 (2015).

55. Wu, Y. *et al.* The Arabidopsis NPR1 protein is a receptor for the plant defense hormone salicylic acid. *Cell Rep.* **1**, 639–647 (2012).

56. Chen, H. *et al.* A Bacterial Type III Effector Targets the Master Regulator of Salicylic Acid Signaling, NPR1, to Subvert Plant Immunity. *Cell Host Microbe* **22**, 777-788.e7 (2017).

57. Zavaliev, R., Mohan, R., Chen, T. & Dong, X. Formation of NPR1 Condensates Promotes Cell Survival during the Plant Immune Response. *Cell* **182**, 1093-1108.e18 (2020).

58. Lebeis, S. L. *et al.* PLANT MICROBIOME. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349**, 860–864 (2015).

59. Mayer, K. F. X. *et al.* A physical, genetic and functional sequence assembly of the barley genome. *Nat. 2012 4917426* **491**, 711–716 (2012).

60. Mascher, M. *et al.* A chromosome conformation capture ordered sequence of the barley genome. *Nat. 2017 5447651* **544**, 427–433 (2017).

61. Eklöf, J. M. & Brumer, H. The XTH Gene Family: An Update on Enzyme Structure, Function, and Phylogeny in Xyloglucan Remodeling. *Plant Physiol.*

153, 456–466 (2010).

62. Fu, M.-M. *et al.* Genome-Wide Identification, Characterization and Expression Analysis of Xyloglucan Endotransglucosylase/Hydrolase Genes Family in Barley (Hordeum vulgare). *Mol. 2019, Vol. 24, Page 1935* **24**, 1935 (2019).
63. Ezquer, I., Salameh, I., Colombo, L. & Kalaitzis, P. Plant Cell Walls Tackling Climate Change: Biotechnological Strategies to Improve Crop Adaptations and Photosynthesis in Response to Global Warming. *Plants 2020, Vol. 9, Page 212* **9**, 212 (2020).
64. Bulgarelli, D. *et al.* Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* **488**, 91–95 (2012).
65. Korenblum, E. *et al.* Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 3874–3883 (2020).
66. Vorwerk, S., Somerville, S. & Somerville, C. The role of plant cell wall polysaccharide composition in disease resistance. *Trends Plant Sci.* **9**, 203–209 (2004).
67. Sharma, R. *et al.* Genome-wide association of yield traits in a nested association mapping population of barley reveals new gene diversity for future breeding. *J. Exp. Bot.* **69**, 3811–3822 (2018).
68. Nice, L. M. *et al.* Mapping Agronomic Traits in a Wild Barley Advanced Backcross–Nested Association Mapping Population. *Crop Sci.* **57**, 1199–1210 (2017).
69. Xu, X. *et al.* Genome-Wide Association Analysis of Grain Yield-Associated Traits in a Pan-European Barley Cultivar Collection. *Plant Genome* **11**, 170073 (2018).
70. Koppolu, R. *et al.* Six-rowed spike4 (Vrs4) controls spikelet determinacy and row-type in barley. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13198–13203 (2013).
71. Mahdi, L. K. *et al.* The fungal root endophyte Serendipita vermifera displays inter-kingdom synergistic beneficial effects with the microbiota in Arabidopsis thaliana and barley. *ISME J. 2021* 1–14 (2021). doi:10.1038/s41396-021-01138-y
72. Watson, A. *et al.* Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat. plants* **4**, 23–29 (2018).
73. Garcia-Gimenez, G. *et al.* Targeted mutation of barley (1,3;1,4)-β-glucan

synthases reveals complex relationships between the storage and cell wall polysaccharide content. *Plant J.* **104**, 1009–1022 (2020).

74. Riaz, A. *et al.* Advances in Genomics-Based Breeding of Barley: Molecular Tools and Genomic Databases. *Agron. 2021, Vol. 11, Page 894* **11**, 894 (2021).

75. Becker, M. & Hensel, G. Genome editing of barley. *Genome Ed. Precis. Crop Breed.* 325–340 (2021). doi:10.1201/9781003048237-14

76. Pourkheirandish, M. *et al.* Evolution of the Grain Dispersal System in Barley. *undefined* **162**, 527–539 (2015).

77. Terrazas, R. A. *et al.* Nitrogen availability modulates the host control of the barley rhizosphere microbiota. *bioRxiv* 605204 (2020). doi:10.1101/605204

78. Standard operating procedure for soil total nitrogen - Dumas dry combustion method. Available at: https://www.fao.org/publications/card/en/c/CB3646EN/. (Accessed: 13th December 2021)

79. Foito, A., Byrne, S. L., Shepherd, T., Stewart, D. & Barth, S. Transcriptional and metabolic profiles of Lolium perenne L. genotypes in response to a PEG-induced water stress. doi:10.1111/j.1467-7652.2009.00437.x

80. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J. 2012 68* **6**, 1621–1624 (2012).

81. White, T.J., Bruns, T.D., Lee, S.B. and Taylor, J.W. (1990) Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. In Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J., Eds., PCR Protocols A Guide to Methods and Applications, Academic Press, New York, 315-322. - References - Scientific Research Publishing. Available at: https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=1321167. (Accessed: 19th November 2021)

82. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nat. 2017 5517681* **551**, 457–463 (2017).

83. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).

84. R Core Team (2018) R A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. - References - Scientific Research Publishing. Available at:

https://www.scirp.org/(S(lz5mqp453edsnp55rrgjct55))/reference/ReferencesPapers.aspx?ReferenceID=2342186. (Accessed: 19th November 2021)

85. Pietrangelo, L., Bucci, A., Maiuro, L., Bulgarelli, D. & Naclerio, G. Unraveling the composition of the root-associated bacterial microbiota of Phragmites australis and Typha latifolia. *Front. Microbiol.* **9**, 1650 (2018).

86. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).

87. Broman, K. W., Wu, H., Saunak Sen, ´ & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinforma. Appl. NOTE* **19**, 889–890 (2003).

88. Sen, Ś. & Churchill, G. A. A statistical framework for quantitative trait mapping. *Genetics* **159**, 371 (2001).

89. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).

90. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods 2017 144* **14**, 417–419 (2017).

91. Guo, W. *et al.* 3D RNA-seq: a powerful and flexible tool for rapid and accurate differential expression and alternative splicing analysis of RNA-seq data for biologists. *RNA Biol.* **18**, 1574–1587 (2021).

92. Soneson, C., Matthes, K. L., Nowicka, M., Law, C. W. & Robinson, M. D. Isoform prefiltering improves performance of count-based methods for analysis of differential transcript usage. *Genome Biol.* **17**, 1–15 (2016).

93. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 1–13 (2010).

94. Mcinnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection Software • Review • Repository • Archive. (2018). doi:10.21105/joss.00861

95. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).

96. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *https://doi.org/10.1214/aos/1013699998* **29**, 1165–1188 (2001).

97. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
98. Oikkonen, L. & Lise, S. Making the most of RNA-seq: Pre-processing sequencing data with Opossum for reliable SNP variant detection. *Wellcome Open Res.* **2**, (2017).
99. Platypus: A Haplotype-Based Variant Caller For Next Generation Sequence Data — Wellcome Centre for Human Genetics. Available at: https://www.well.ox.ac.uk/research/research-groups/lunter-group/lunter-group/platypus-a-haplotype-based-variant-caller-for-next-generation-sequence-data. (Accessed: 19th November 2021)
100. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
101. Marçais, G. *et al.* MUMmer4: A fast and versatile genome alignment system. *PLOS Comput. Biol.* **14**, e1005944 (2018).

Acknowledgements

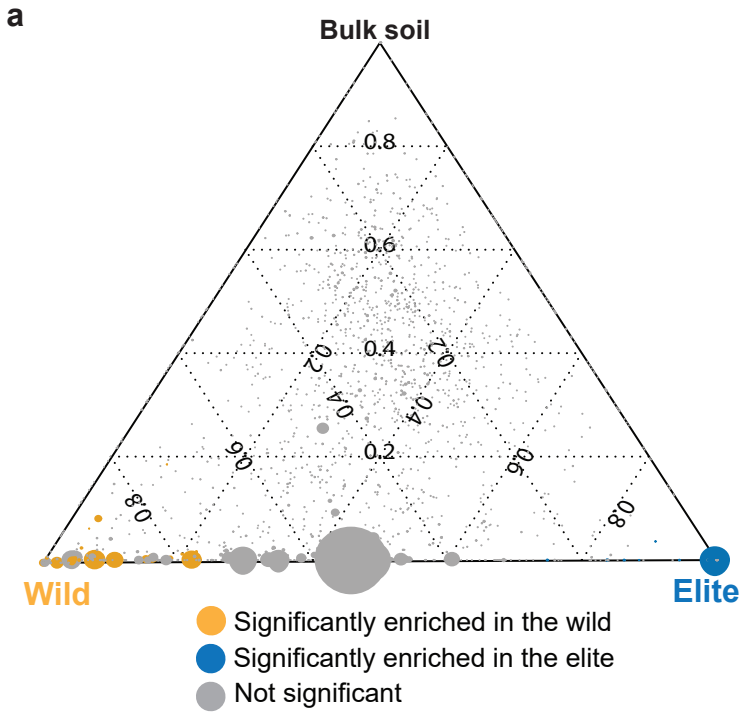
We thank Malcolm Macaulay and Jim Wilde (The James Hutton Institute, Invergowrie, UK) for the technical assistance during the execution of the experiments. We also thank Nils Stein and Mary Ziems (IPK, Gatersleben, Germany) who kindly provided us with seeds of the barley pangenome collection. This work was supported by a Royal Society of Edinburgh Personal Research Fellowship to DB and an UKRI grant (BB/S002871/1) to DB, GB and RW.

Author Contributions

C.E.M., R.A.T., R.W. and D.B designed the experiments. C.E.M., M.C., R.S., J.A., G.B. and D.B. conceived the data analyses. C.E.M., R.A.T., R.K., L.P., M.M, A. A., performed the experiments. C.E.M., M.C., A.F., R. K., A. A., R.S. and J.A. analysed the data. J.M. and P.H. generated the amplicon sequencing libraries. A.M and K. P. provided access to the HEB-25 seed material and analysed their genomic information. C.E.M, M.C and D.B wrote the first version of the manuscript, all authors

collegially commented on and contributed to the submitted version. C.E.M and M.C contributed equally to this work.

Fig.1: Microbiota composition displays a quantitative variation in a barley segregant population between wild and elite parental lines



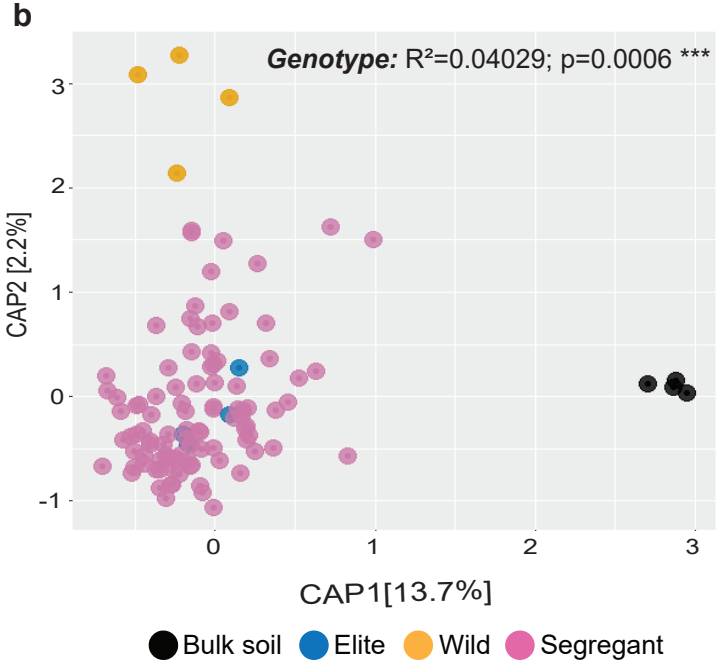
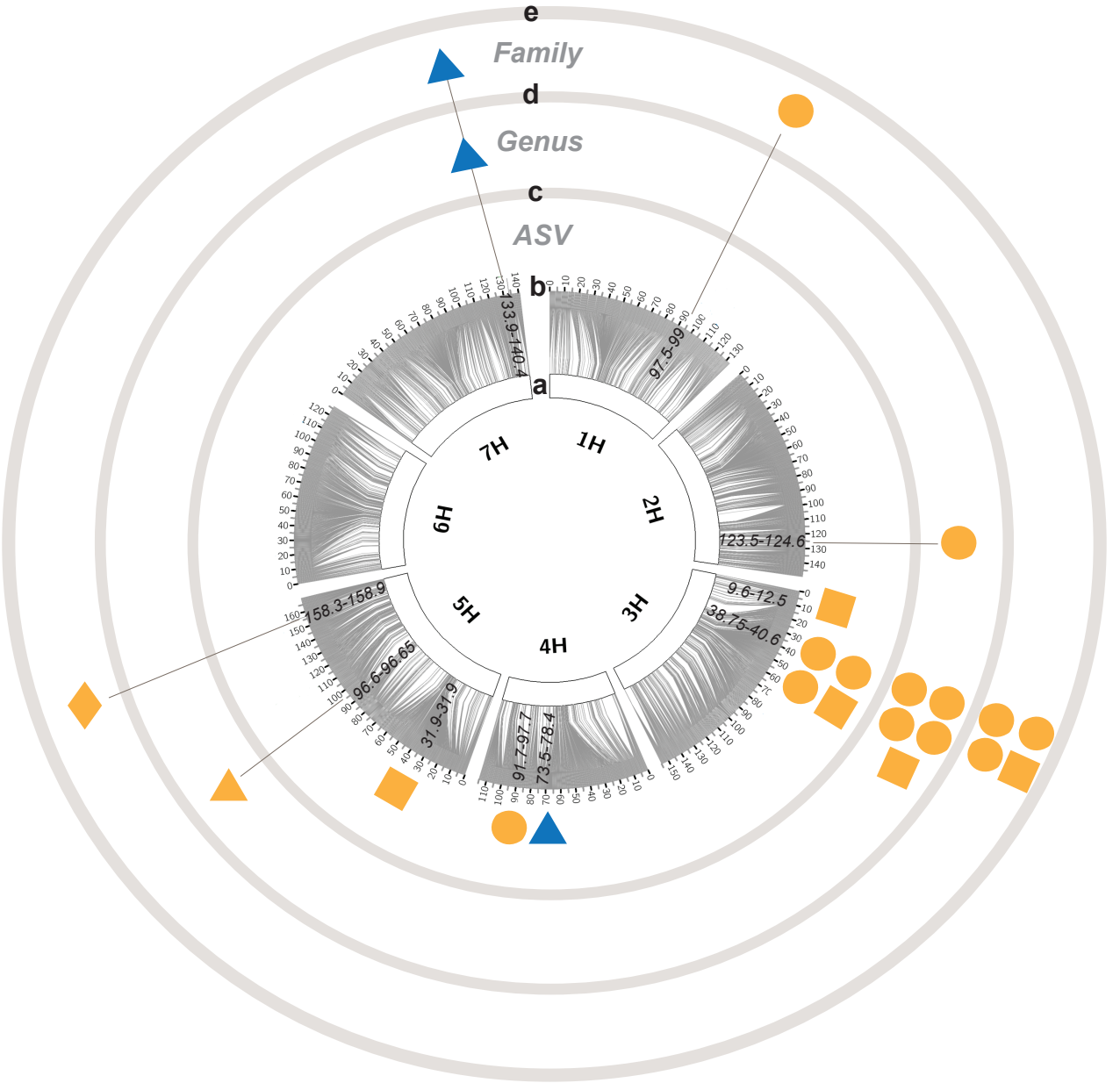


Fig.2: Genetic map of the barley determinants of the rhizosphere bacterial microbiota



Enriched in wild Enriched in elite

Taxonomy

Proteobacteria Actinobacteria Acidobacteria Bacteroidota

Fig. 3 : Wild alleles at locus *QRMC-3HS* are associated to a shift in the composition of the bacterial, but not fungal, microbiota

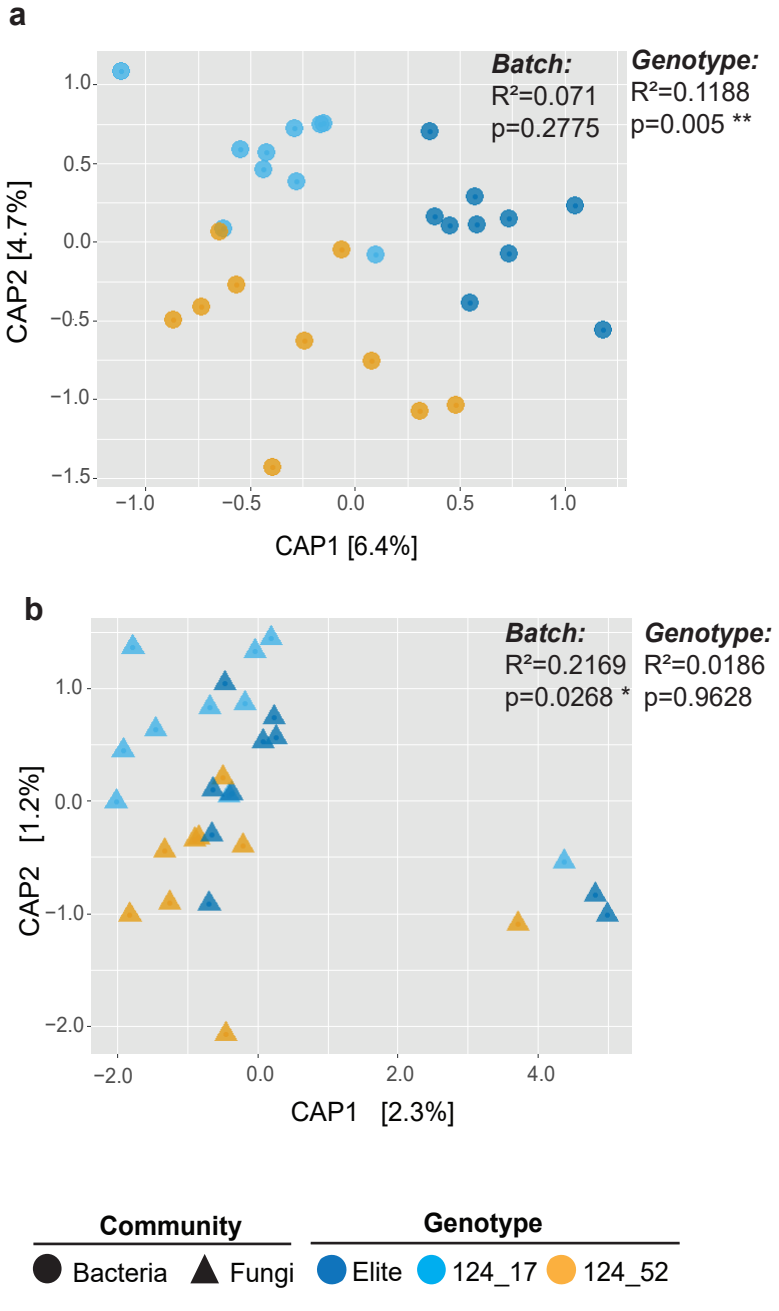


Fig. 4: Sibling lines harboring contrasting alleles at locus *QRMC-3HS* and the cultivar Barke display distinct root transcriptional profiles

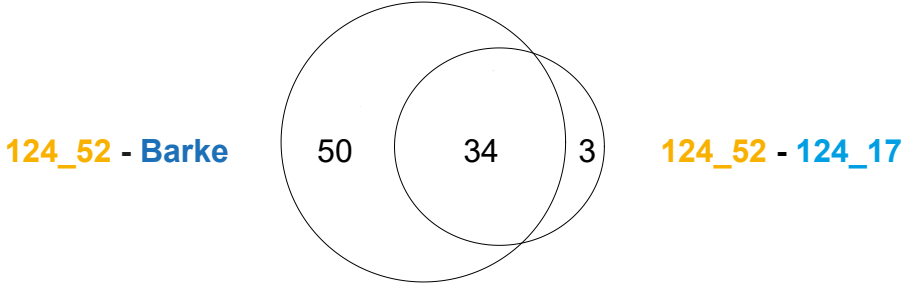


Fig. 5: Differentially expressed genes mapping at locus *QRMC-3HS*

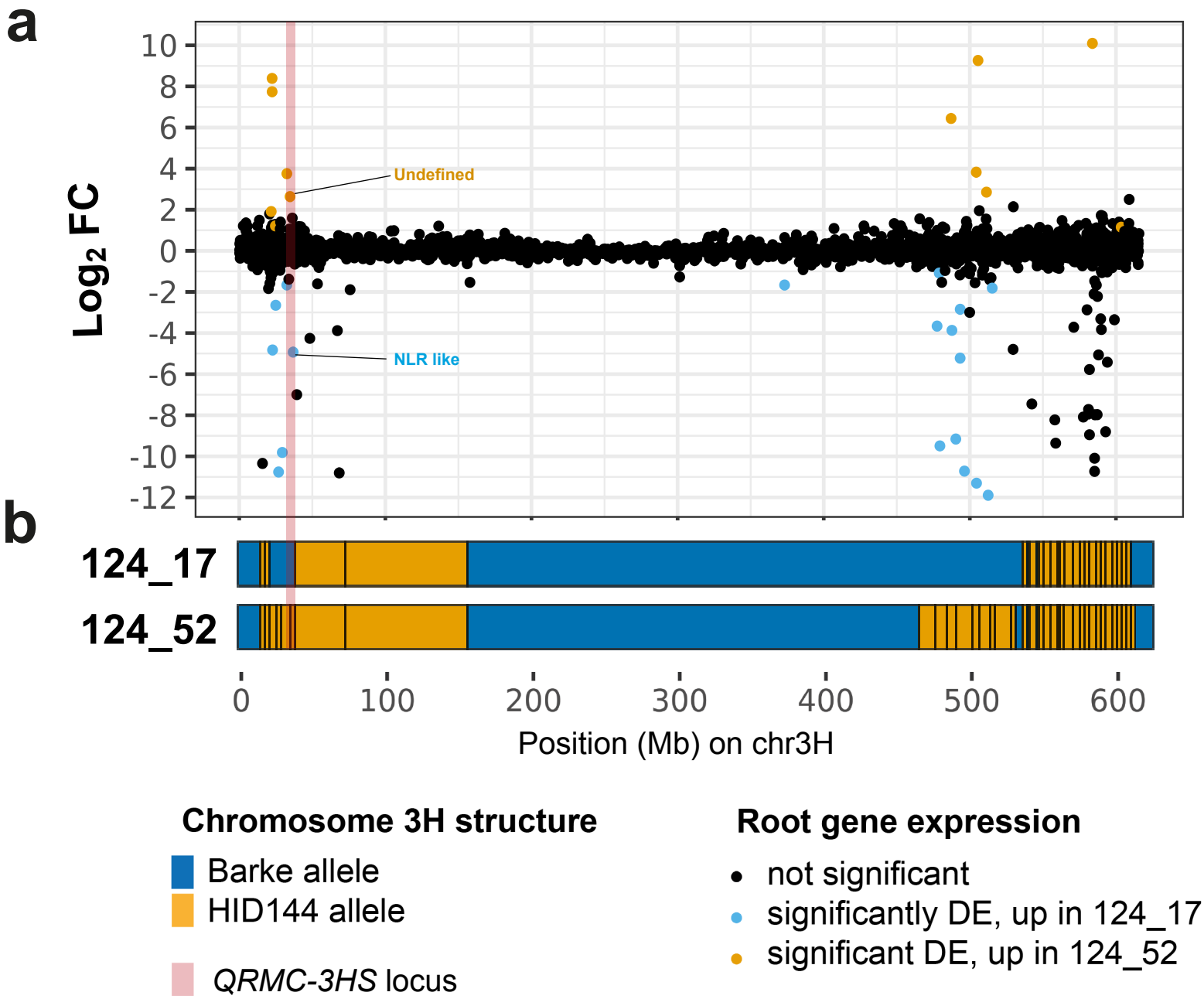


Fig. 6 : Locus *QRMC-3HS* defines an area of structural variation in the barley genome

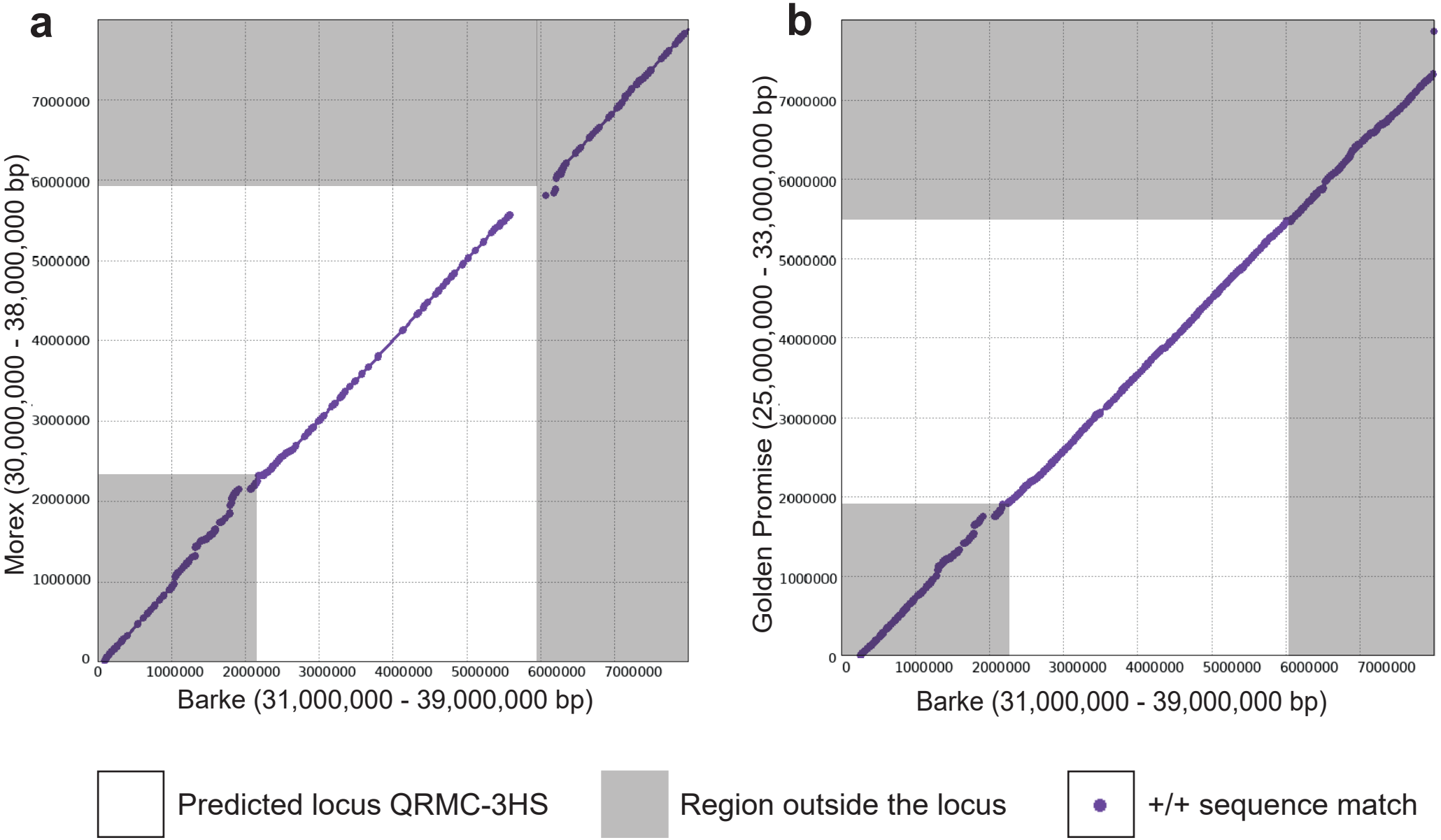
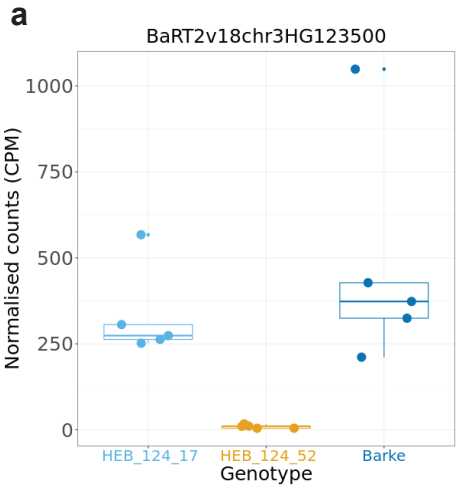
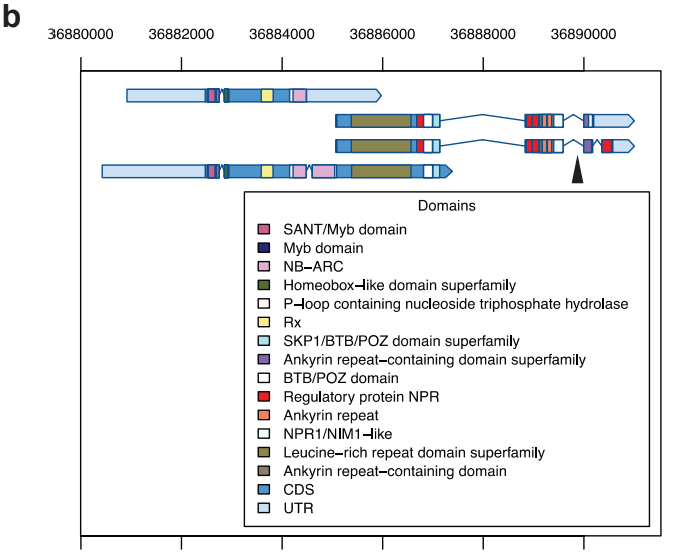
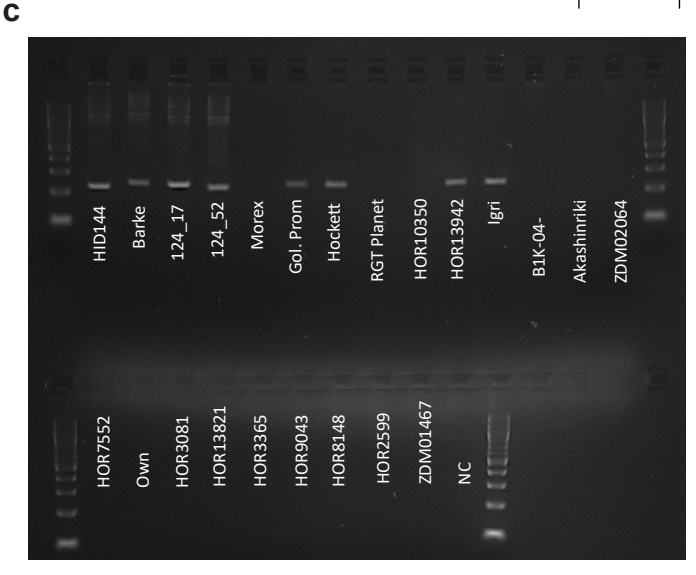



Fig. 7: The *NLR* gene is associated with genotype-dependent transcriptional and genomic variations