

RESEARCH

Open Access



# Host genotype, soil composition, and geo-climatic factors shape the fonio seed microbiome

Naheed Tabassum<sup>1,2</sup>, Hanin Ibrahim Ahmed<sup>1,2</sup>, Sabiha Parween<sup>1,2</sup>, Arsheed H. Sheikh<sup>1,2</sup>, Maged M. Saad<sup>1,2\*</sup>, Simon G. Krattinger<sup>1,2\*</sup> and Heribert Hirt<sup>1,2\*</sup>

## Abstract

**Background** Fonio (*Digitaria exilis*), an orphan millet crop, is the oldest indigenous crop in West Africa. Although the yield is low due to pre-domestication characteristics, the quick maturation time, drought tolerance, and the ability to thrive on poor soils make fonio a climate-smart crop. Being holobionts, plants evolve in close interaction with microbial partners, which is crucial for plant phenology and fitness. As seeds are the bottleneck of vertically transmitting plant microbiota, we proposed to unravel the seed microbiome of the under-domesticated and resilient crop fonio. Our study investigated the bacterial seed endophyte diversity across 126 sequenced fonio accessions from distinct locations in West Africa. We conducted a correlation study of the structures and functions of the seed-associated microbiomes with the native geo-climate and soil structure data. We also performed Genome-wide association studies (GWAS) to identify genetic loci associated with seed endophyte diversity.

**Result** We report that fonio millet has diverse heritable seed endophytic taxa. We analyzed the seed microbiomes of 126 fonio accessions and showed that despite the diversity of microbiomes from distinct geographical locations, all fonio genetic groups share a core microbiome. In addition, we observed that native soil composition, geo-climatic factors, and host genotype correlate with the seed microbiomes. GWAS analysis of genetic loci associated with endophyte seed bacterial diversity identified fonio SNPs associated with genes functioning in embryo development and stress/defense response.

**Conclusion** Analysis of the seed endophyte of the climate-smart crop fonio indicated that despite possessing a heritable core microbiome, native conditions may shape the overall fonio seed microbiomes in different populations. These distinct microbiomes could play important roles in the adaptation of fonio to different environmental conditions. Our study identified the seed microbiome as a potential target for enhancing crop resilience to climate stress in a sustainable way.

**Keywords** Seed microbiome, Core microbiome, Microbiome engineering, Plant stress adaptation

\*Correspondence:

Maged M. Saad  
maged.saad@kaust.edu.sa  
Simon G. Krattinger  
simon.krattinger@kaust.edu.sa  
Heribert Hirt  
heribert.hirt@kaust.edu.sa

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Fonio (*Digitaria exilis*) is one of the oldest, indigenous, semi-domesticated, and fastest-maturing cereals in West Africa [23, 27]. Fonio can grow in nutrient-poor soil and is very drought tolerant, which makes it suitable for dry-land agriculture [7]. Nutritionally, fonio is comparable to other cereals [60]. It could complement the staple crops amid climate change and desertification challenges. Due to some wild characteristics, including seed shattering and lodging, fonio yields are much lower compared to the major cereal crops rice, maize, and wheat [8]. Fonio is a tetraploid species ( $2n=4\times=36$ ) with a high inbreeding rate. Population genomic analyses revealed that the genetic diversity in fonio is relatively low [6]. Recently, a fonio high-quality, chromosome-scale reference assembly and whole-genome sequencing data of a fonio diversity panel were established and analyzed [5]. The diversity panel included 183 wild and semi-domesticated accessions covering a wide geographic and climatic range. The fonio population structure was shaped by climatic, geographic, and ethnolinguistic factors, and the 183 fonio accessions were divided into six distinct genetic clusters/groups and one admixed group.

Besides harnessing genetic information and modification, breeders can improve plants by exploring and maneuvering their association with microbes [31]. For instance, plants grown in arid conditions associate and interact with bacteria that help plants combat abiotic challenges [17]. Such plant–microbe interaction symbiotically attains a high degree of adaptation to the extreme environment. In addition, the microbiome plays a crucial role in plant growth and development, nutrient uptake, and protection against biotic stress. Arguably, plants are holobionts meaning that a community of microbes coevolves with their host contributing to host phenology and fitness [38]. Therefore, microbes perform crucial functions throughout plant life history [49].

The selection and assembly of the plant microbiome is a complex, multistep, successional process. Members of horizontally (from the surrounding niche) or vertically (from the mother plant) acquired microbiota can contribute to the ultimate composition of plant microbiota [49, 52]. The embryo surface, storage tissue, and seed coat represent distinct seed microhabitats. Seeds harbor fewer microbial taxa than other plant tissues, and the number of seed microbes varies among plant species [1]. The vertical transmission of potentially beneficial microbial communities to fortify the progeny of a plant species opens a new technological application in crop breeding [9]. Seed microbe engineering could be used for targeted metabolites or antimicrobial compound production to improve plant biomass and yield under biotic or abiotic stresses [50].

A core microbiota is associated with a host in broad-ranging environmental conditions [55]. Compared to adult plants (with rhizosphere), the core microbes are likely to be more reproducibly found in seeds and seedlings [54]. For example, the core endophytic microbiota of maize is conserved in seeds across boundaries of evolution, ethnography, and ecology [29]. Also, maize seeds constitute a significant source of inoculum for juvenile root endophytes [28]. It has been recently proposed that one should describe the potential functional relevance of core microbiota rather than their taxonomy [59]. The dynamics of vertically transmitted core seed microbiota were also shown in six rice varieties [59], but the study of their functions is yet unexplored.

Here we investigated the seed-associated microbiome of an underutilized, resilient orphan crop. We report that 126 fonio accessions have diverse seed endophytic microbiomes. The six and admixed fonio genetic groups were found to share a heritable core seed microbiome. Moreover, we observed that native soil composition, geoclimatic factors, and host genotype correlate with the seed microbiomes of different fonio accessions. Finally, using a genome-wide association study (GWAS) for endophyte seed bacterial diversity, we found that distinct fonio genetic loci are associated with genes with roles in embryo development and stress/defense response. Our study provides a proof of concept to understand the composition and diversity of the seed-associated microbiomes of a cereal from distinct geographical locations. Furthermore, our study highlights the potential of the seed microbiome as a target for microbiome engineering in future crop breeding programs.

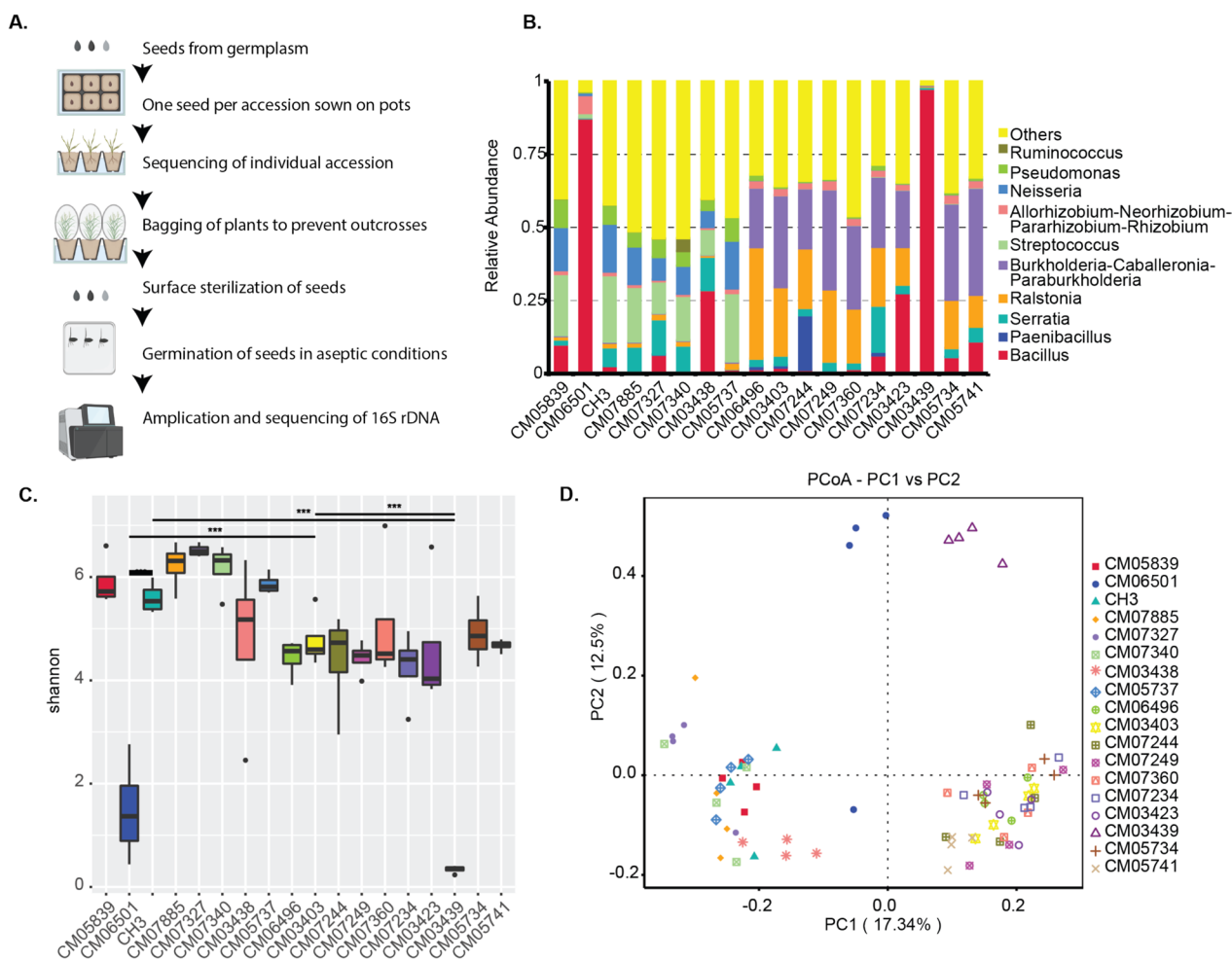
## Materials and methods

### Germplasm

Seeds from a diversity panel described by Abrouk et al. [5] were used. We germinated a single seed per accession simultaneously in the same soil in the greenhouse (Fig. 1A). The resulting seeds from the propagated single plants of 126 accessions (118 accessions with 3 replicates and 8 accessions with 2 replicates) were then used in the pilot and main studies. For vertical transmission, we used the seeds from the original and the propagated plants from three accessions.

### Extraction of DNA from seedlings and amplification of bacterial 16S rDNA

We surface sterilized the seeds using 70% ethanol + 0.05% sodium dodecyl sulfate for 10 min on a shaker, washed 2 times in 96% ethanol, and let to dry. We germinated surface sterilized seeds on 0.5 MS media (MS basal salts, Sigma). The seedlings were harvested with liquid nitrogen in sterile condition using



**Fig. 1** Diversity of seed-associated microbiomes in 18 fonio accessions. **A** Scheme showing sample collection and processing in the study. We germinated a single seed per accession and used the progeny seed for sterilization, germination, 16S rDNA sequencing and analysis. (Illustration created by Biorender.com) Histograms of the relative abundances of the top 10 bacterial communities at genus level (**B**) and Boxplot of Shannon index showing diversity in the 18 fonio accessions (**C**). Group 1 (CM05839, CM06501, CM06496), group 2 (CM07885, CH3, CM03403), group 3 (CM07249, CM07244, CM07327) and group 4 (CM07340, CM07234, CM07360), and group 5 (CM03438, CM03423, CM03439) and group 6 (CM05737, CM05741, CM05734). Kruskal–Wallis test, \* indicates  $p < 0.05$ , \*\* indicates  $p < 0.01$  and, and \*\*\* indicates  $p < 0.001$ . The box boundaries indicate the first and third quartile. Lines extending from the boxes (whiskers) indicate variability outside the lower and upper quartiles. The lines in the middle of the boxes represent the median values of Shannon index for accession. Outliers are plotted as individual points. **D** Principal component analysis (PCoA) shows the grouping patterns of the accession based on unweighted UniFrac distance

sterilized equipment including tubes with closed lid and pulverized. We performed crude extraction of genomic DNA by adding an extraction buffer (100 mM Tris pH=8, 50 mM EDTA, 500 mM NaCl, 1.5% SDS). We mixed the lysate with 1:1 vol of isopropanol and centrifuged at 14,000 rpm for 10 min. We washed the DNA pellets with 70% ethanol, dried and dissolved them in 1X TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). We checked the purity of the DNA using Nanodrop (Thermo) and qubit (Thermo).

**Amplicon generation and illumina sequencing**

We amplified 16S rDNA using V5-V7 (799F/1193R) primers with the barcode (799F, 5'-AACMGGATTA GATACCCCKG-3', 1193R, 5'-ACGTCATCCCCACCT TCC-3'). We performed the PCR reaction with Phusion® High-Fidelity PCR Master Mix (NEB) (including a negative control). We checked the PCR products on 2% agarose gel, and 400–450 bp amplicons were mixed in an equal-density ratio and then purified the PCR product with a Qiagen Gel Extraction Kit (Qiagen). We generated

the sequencing libraries using NEB Next<sup>®</sup> Ultra DNA Library Prep Kit for Illumina and added the index codes, following the manufacturer's recommendations. We assessed the library quality on the Qubit<sup>®</sup> 2.0 Fluorometer (Thermo Scientific) and Agilent Bioanalyzer 2100 system. At last, we sequenced the library on an Illumina platform and generated 250 bp paired-end reads.

### **Bioinformatic analysis**

#### ***Paire-end reads assembly and quality control***

Paired-end reads were assigned to samples based on their unique barcodes and truncated by cutting off the barcode and primer sequences. We used FLASH (v1.2.11, <http://ccb.jhu.edu/software/FLASH/>) to merge the reads to get raw tags after removing the barcode and primer sequence. We used fastp software to obtain high-quality clean tags. We then blast clean tags to the database using Vsearch software. We removed the chimera to obtain final effective tags.

#### ***Cluster/denoise and species annotation of amplicon sequence variants (ASVs)***

We used DADA2 or deblur module in QIIME2 software to denoise and obtain the final ASVs. Then, we used the Classify-sklearn moduler in QIIME2 software to compare ASVs with the database and to obtain the species annotation of each ASV. The absolute abundance of individual microbial taxa within a sample was determined by quantifying the number of reads or sequences associated with each ASV. Greater numbers of assigned sequences to an ASV corresponded to increased absolute abundance within the sample.

#### ***Alpha diversity***

We apply alpha diversity in analyzing the complexity of species diversity with six indices, including observed species, chao1, Shannon, Simpson, ACE, and goods-coverage (<http://www.mothur.org>). All these indices in our samples were calculated with QIIME (Version 1.7.0) and displayed with R software (Version 2.15.3).

#### ***Beta diversity***

We used beta diversity analysis to evaluate differences of samples in species complexity. We calculated beta diversity on both weighted and unweighted UniFrac distances by QIIME software (Version 1.7.0). We performed PCoA analysis to get principal coordinates and visualize complex multidimensional data using the WGCNA package, stat packages, and ggplot2 package in R software (Version 2.15.3). We obtained a distance matrix of weighted or unweighted UniFrac among samples obtained before transforming to a new set of orthogonal axes.

#### ***Functional prediction of core ASVs***

We predicted the core microbiome's function using PICRUSt2 (Phylogenetic Investigation of Communities by Reconstruction of Unobserved Stats 2) software using the default "max parsimony" method for hidden-state prediction and a Nearest Sequenced Taxon Index (NSTI) value of 2.0. The Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologs (KO) abundances predictions were obtained, and its description were obtained from the KofamKOALA-KEGG Orthology Search database.

#### ***Phylogeny of core ASVs***

The phylogenetic analysis of 191 core ASVs was done using the PHYLIP program at default mode. The tree was visualized using Tree Viewer and each of the ASVs' absolute abundance values was highlighted.

#### ***Genome-wide association study of fonio microbiome***

A genome-wide SNP map of fonio used in this study was generated previously based on the whole-genome sequencing of the fonio diversity panel [5]. To find the significant genetic variants in the host that are associated with the microbiome, we considered the first axis of the UniFrac weighted PCoA of one replicate as phenotypes (i.e.,  $\beta$ -diversity) and SNP data from 99 fonio accessions as genotypes. The 99 accessions were selected based on replicates clustering together on the PCoA analysis. When all three replicates from a particular accession did not exhibit coherent clustering patterns, indicating potential inconsistencies, we excluded that specific sample from the GWAS analysis to ensure the reliability and coherence of our association analysis. We then filtered SNPs with 5% Minor allele frequency (MAF), 10% missing rate using VCFtools (v0.1.17) [15] and Linkage disequilibrium (LD) with  $R^2 \geq 0.3$  (using PLINK software v1.90 [45] resulting in 137,476 SNPs. We tested the association of SNPs with microbiome variability by the general linear model (GLM) using three principal components as covariates to control for population structure. We identified the significant associations above the threshold of Bonferroni  $p$ -value =  $1e-05$ . We then considered a genomic window of 50 kb upstream and 50 kb downstream for each position as a candidate region. To test the relationship between microbiome variability and environmental factor, Pearson's correlation test was performed using MVapp (<https://mvapp.kaust.edu.sa/>).

## **Results**

### **Fonio accessions have diverse seed endophytic microbiomes**

The fonio genetic population structure was previously defined by six distinct and one admixed genetic group [5].

Fonio accessions were grown for one generation in the greenhouse and in the same soil to minimize any experimental bias (Fig. 1A). For a pilot study, we randomly selected three accessions from the six genetic and one admixed group to understand the composition of the seed endophytic microbiome. These accessions belong to group 1 (CM06501, CM05839, CM06496), group 2 (CM07885, CH3, CM03403), group3 (CM07249, CM07244, CM07327), group 4 (CM07340, CM07234, CM07360), group 5 (CM03438, CM03423, CM03439), and group 6 (CM05737, CM05741, CM05734). We aseptically germinated surface sterilized seeds of the 18 accessions and amplified 16S rDNA from gDNA (4 replicates). Using the ASV (Amplicon Sequence Variant) approach, we obtained a total of 11.4 million reads that could be assembled into 13,970 ASVs of bacterial taxa (Table S1). The rarefaction curves showed that the sequence library size was sufficient to cover the microbial diversity in each sample (Fig. S1A). The dominant genera were *Bacillus*, *Paenibacillus*, *Serratia*, *Ralstonia*, *Streptococcus*, *Burkholderia*, *Neisseria*, and *Pseudomonas* (Fig. 1B, Table S2). Next, we calculated the alpha diversity of the observed species using the Shannon index. While two accessions CM03439 and CM06501 have lower alpha diversity with the predominance of *Bacillus*, others have high alpha diversity (Fig. 1C, Fig. S1B Table S3). To further gain insights into the microbiota structures, we used PCoA plots based on unweighted UniFrac distance to visualize the separation of the microbiota structures in these 18 accessions, each with four replicates (Fig. 1D). We observed that CM03439 and CM06501 with lower diversity clearly differentiated, while accessions with moderate and high diversity are partially differentiated. Overall, this pilot study showed that the seed bacterial endophytic microbiomes of different fonio accessions are potentially composed of unique and shared taxa.

### Diverse genetic fonio groups share a core microbiome

The fonio genetic population structure correlated with geo-climatic factors and was defined by six distinct and one admixed genetic group [5]. We used 126 fonio accessions representing these six genetic and one admixed groups for an in-depth analysis of the structure and function of the seed-associated microbiomes (Fig. S2A,

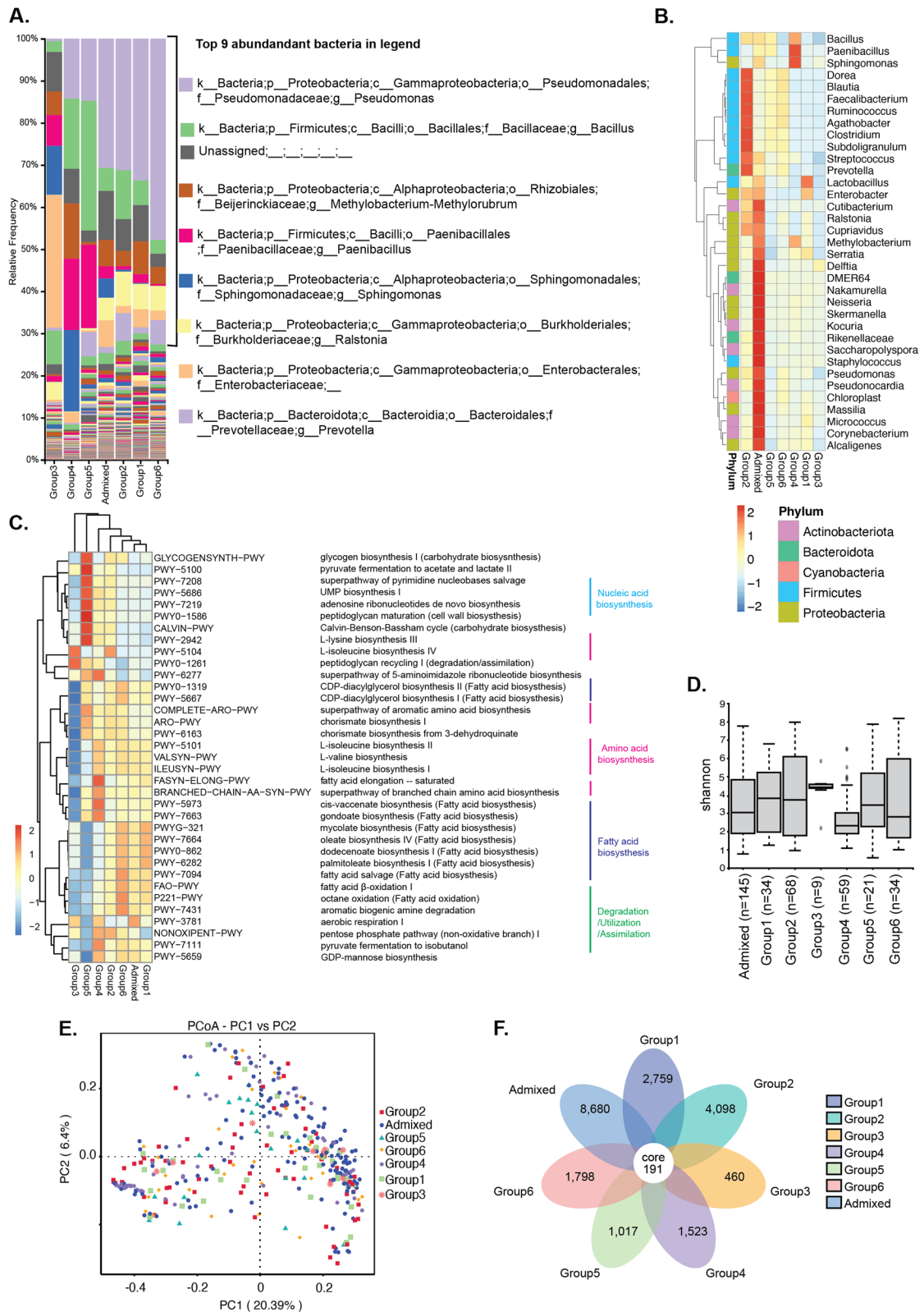
Table S4). Using the ASV approach, 73.8 million reads were assembled into 30,356 ASVs (unfiltered) of bacterial taxa (Table S5). The rarefaction curves (Fig. S2B) showed that the sequence library size was sufficient to cover the microbial diversity in each group. Overall, the dominant phyla across the fonio genetic groups were Proteobacteria and Firmicutes (Fig. S2C), whereas the dominant genera were *Pseudomonas* and *Bacillus* (Fig. 2A). Interestingly, the genera *Lactobacillus* and *Enterobacter* are overrepresented in fonio group 1, whereas group 2 exhibits a significant overrepresentation of the Firmicutes phylum (Fig. 2B). While fonio group 4 stands out for its specific overrepresentation of the genera *Bacillus*, *Paenibacillus*, and *Sphingomonas* the admixed group showed a pronounced overrepresentation of the phyla Proteobacteria and Actinobacteriota (Fig. 2B).

To gain insights into the metabolic potential of the microbiome, we conducted a Picrust2 analysis, from which we inferred the top 35 pathways (MetaCyc) based on 16S rRNA gene sequences (Fig. 2C). Among these pathways, we observed both anabolic processes, such as the biosynthesis of fatty acids, nucleic acids, and amino acids, as well as catabolic pathways involved in degradation/utilization. While the functional roles did not result in clear clusters among the groups in the PCA plot (Fig. S2D), we did identify instances of over/under representation of specific functions among different groups. These findings suggest that the functions performed by seed-associated endophytes could potentially contribute to plant growth and development through nutrient availability and help to adapt phenology and fitness to the respective environmental conditions as the prominence of these pathways varied among the fonio genetic groups.

To assess the species richness and diversity (alpha diversity), we analyzed the Shannon indexes among the fonio genetic groups revealing variation across the genetic groups (Fig. 2D). For example, group 6, with a lower sample size ( $n=34$ ), is more diverse than group 2 ( $n=68$ ) or the admixed group ( $n=145$ ) with a higher sample size. We used PCoA plots based on unweighted UniFrac distance to visualize the microbiota structures. We did not observe any differentiation of fonio

(See figure on next page.)

**Fig. 2** Core and diverse ASVs across the fonio genetic population. **A** Histograms of the relative abundances of the top bacterial communities at genus level in the fonio genetic group (admixed, group 1, group 2, group 3, group 4, group 5, and group 6). Heat map showing over/underrepresented bacteria at genus level (**B**) and functional predictions across fonio accession groups (**C**). **D** The box plots show the Shannon diversity index across the genetic fonio groups where “n” is the number of samples. The box boundaries indicate the first and third quartile. Lines extending from the boxes (whiskers) indicate variability outside the lower and upper quartiles. The lines in the middle of the boxes represent the median values of Shannon index for each group. Outliers are plotted as individual points. **E** PCoA of 126 fonio accessions (color and shape code is based on the corresponding groups) based on unweighted UniFrac distance. **F** The flower diagram shows unique and core taxa with the admixed group



**Fig. 2** (See legend on previous page.)

accessions that belong to the same genetic group, indicating that the microbiome composition is complex and potentially shares the taxa (Fig. 2E). Therefore, we took a reductionist approach [20] to understand the microbiome composition of each fonio genetic group. To visualize the abundance of unique and shared microbial core taxa in the six genetic and admixed groups, we generated a floral diagram (Fig. 2F). Here, the sample size in each genetic group is proportionate with their unique taxa. For example, group 2, with the highest number of samples, had the most abundant unique taxa. Despite coming from distinct geographical locations, all genetically distinct fonio groups shared 191 core bacterial taxa. Also, the admixed group, which shares genetic diversity with other groups, retains all the core microbiota but contains >8600 unique bacterial taxa. A side-by-side comparison of Fig. S2E and F shows that the number of unique microbial taxa in each fonio genetic group is lower in Fig. 2F. This indicates that the admixed fonio group shares some microbial taxa with each non-admixed fonio group, the core taxa remain the same. Among the core, the majority of phyla are Proteobacteria (31%), Firmicutes (25%), Actinobacteriota (21%), Bacteroidota (8%), and unassigned/partially assigned/others (15%) (Table S6). Although the absolute abundance of each core taxa varied among the fonio genetic groups, the most abundant genera are *Pseudomonas* sp. (ASV0) (Fig. S3, Table S7). We further predicted the function of the core taxa in metabolism, signaling, and genetic information processing using KEGG ontology (KO) (Fig. S2F). The most prevalent KO included the terms peptidase and inhibitors, secretion systems, two core component systems, and transcription factors. We hypothesize that these functional traits possess plant-beneficial properties that may help the endophytes to establish the symbiosis with fonio.

As seeds represent a bottleneck in inheriting microbiomes [2], we checked for the vertical transmission of the respective seed microbiome structures in three accessions (Fig. 3A). For this, we analyzed the presence of common ASVs in the parent and progeny seeds of CM05839 of group 1, CM07327 of group 3, and CM05757 of group 6. We observed that in accession CM05839, 62% ( $337/543 \times 100$ ) of progeny ASVs were acquired from parents, and 38% ( $206/543 \times 100$ ) of progeny ASVs were acquired from soil. (Fig. 3B). Likewise in accession CM07327 and CM05757, 74% ( $358/489 \times 100$ ) and 52% ( $208/400 \times 100$ ) of progeny ASVs were acquired from parent and 26% ( $131/489 \times 100$ ) and 48% ( $192/400 \times 100$ ) from soil, respectively (Fig. 3C and D, Table S8). Altogether, these data show that although a significant proportion of the seed microbiomes can be transferred vertically from seed to seed, the overall microbiome

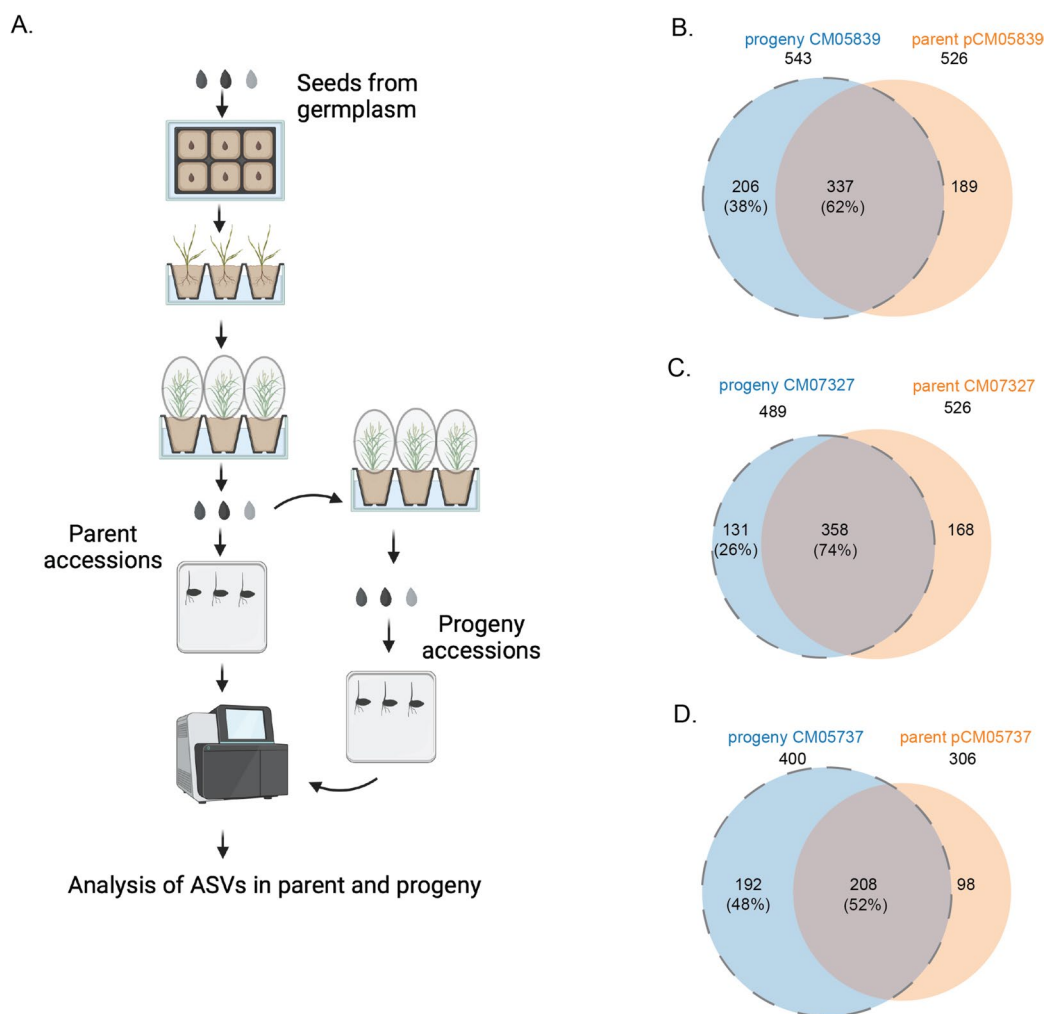
composition remains dynamic as a notable portion of ASVs can originate from the soil environment.

#### **Soil composition, geo-climatic factors, and host genotype correlate with the fonio seed microbiome**

As the fonio accessions are derived from distinct geo-climatic conditions but retain unique/hub microbiota (Fig. S4), we hypothesized that the environmental factors might influence the assembly, presence, and shape of the seed microbiomes. To test this hypothesis, we obtained surface soil data (pH, density, gravel, sand content) from their collection sites (Table S9). We observed a correlation (Pearson's correlation;  $p < 0.01$ ) between the microbial diversity (the first principal component of PCoA) and soil variables, including pH, gravel content, soil density, gravel, and carbon content (Fig. 4, Table S10). Because the fonio genetic diversity correlates with geographical (latitude, longitude, altitude) and climate factors (mean temperature and precipitation) [5], we asked whether the seed microbiomes show similar correlations. We found a significant correlation with longitude, altitude, and temperature but not with precipitation and latitude. However, the fonio genotypes also correlated with soil factors such as density, gravel, sand, and carbon content. In conclusion, we observed a positive correlation between the fonio genetic population structure and the diversity in their associated microbiomes. Hence, soil composition, geo-climatic factors, and the host genotype correlate with the seed microbiomes of different accessions (Fig. 6).

#### **Genome-wide association study (GWAS) unveils fonio genetic loci associated with endophyte microbiome structure**

The correlation between the genetic diversity of seed microbiomes with the genetic diversity of fonio (Fig. 4) is an ideal starting point to find genetic loci in fonio that are associated with the seed microbiome structure [18]. We performed GWAS to find associations between host genetic loci and seed microbiome diversity. Our GWAS with seed microbiome  $\beta$ -diversity revealed significant associations for 17 fonio SNPs on chromosomes 2B, 4A, 4B, 5A, 5B, 7A, 7B, 8B, and 9B (Fig. 5A, Fig. S5). Out of these, 16 SNPs were intergenic. As done previously, we considered 50 kb downstream and upstream of the SNP locus to obtain a list of candidate genes associated with the genetic locus [5], revealing 117 genes that might influence microbiome  $\beta$ -diversity. One prominent candidate interval on chromosome 7B contained ten genes (Fig. 5A, Table 1, Table S11), including a homolog of the *Arabidopsis* *AtVps11* (vacuolar protein-sorting-associated protein 11) gene. *Vps11* plays a role in vacuole biogenesis during embryogenesis [53]. Dexi7B01G0005290 is a homolog of *AtCLPBI* (HSP101). *AtCLPBI* plays a role in releasing ribosomal RNAs from stress granules

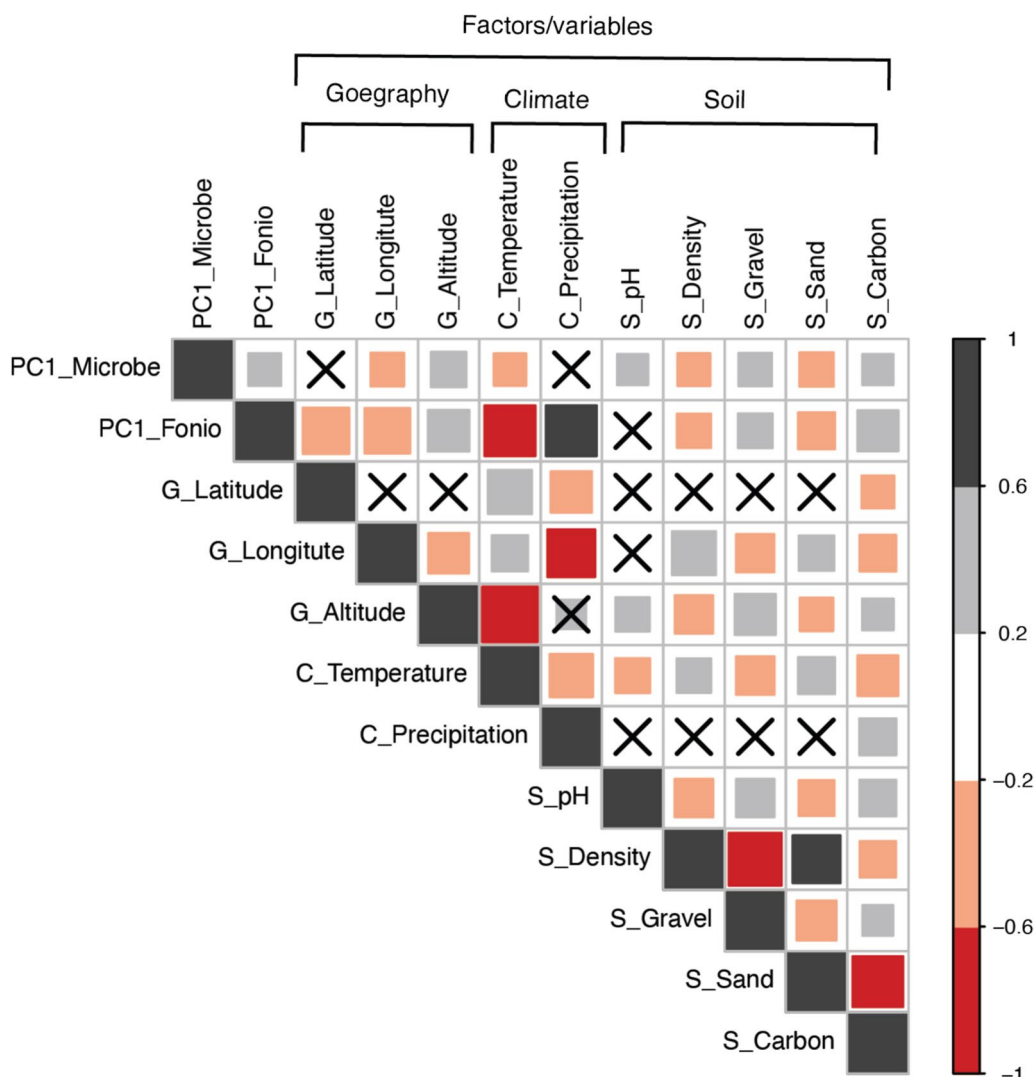


**Fig. 3** Heritable microbial taxa in 3 selected fonio accessions. **A** Scheme showing experimental set up to study vertical and horizontal transmission of ASVs parent and progeny (illustration created by Biorender.com). **B–D** Venn diagram of vertically transmitted ASVs in 3 fonio accessions: CM05839, CM07327, and CM05737. The accessions (parent accessions) were grown for one more generation to get the progeny seeds. These seeds were surface sterilized, germinated and 16S amplicon sequencing was performed

for heat stress recovery and plays a role during seed germination (M. [34, 48]. Dexi7B01G0005300 shares similarities with the F-box protein family which regulates diverse cellular processes, such as cell cycle transition, signal transduction, and transcriptional regulation [37]. Besides, we found some genes in other chromosomes with a putative involvement in embryonic development, such as *AtXTH28* (homologous to Dexi2B01G0015860), *AtKCR1* (resembling to Dexi4A01G0012020), *AtABC17* (akin to Dexi5A01G0007910), *AtPDIL1-1* (similar to Dexi5B01G0013640), *AtNAGLU* (homologous to Dexi7A01G0021150), and *AtMYB98* (sharing similarity with Dexi9B01G0004890). Another set of genes is potentially involved in plant–microbe interaction. For instance, *WAK2* (homologous to Dexi7A01G0021190)

is a wall-associated receptor kinase-2 involved in pathogen response and cell expansion during development. A dominant allele of *WAK2* activates defense responses, such as the upregulation of numerous genes involved in pathogen resistance and cell wall biogenesis, and ROS accumulation [36]. Another gene involved in ROS regulation is *StRBOHC* (homologous to Dexi8B01G0010430). The related members, *AtRBOHD* and *AtRBOHF* are key regulators of immunity [46]. *AtFMO1* (homologous to Dexi8B01G0010390) is a flavin-dependent monooxygenase, which enhances bacterial and fungal resistance [41, 35]. We did not perform gene ontology (GO) analysis for the biological function as many candidate genes encode for proteins with unknown functions (Table S11). Altogether, we found





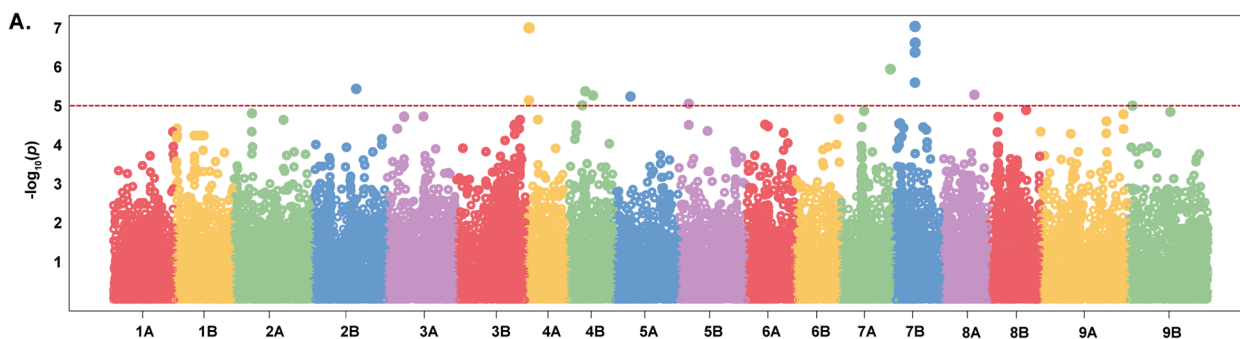
**Fig. 4** Correlation of seed microbiome diversity with environmental variables. The figure shows the Pearson correlation coefficients between selected traits. The color and size of the square reflect the strength of the correlation. The non-significant correlations, with a  $p$ -value above 0.01, are indicated with a cross in the individual cells. The correlation coefficients are calculated using raw data, with a total number of 126 fonio accessions. PC1\_microbes are the first PC axis from the PCoA analysis based on weighted UniFrac distances, which represents the microbiome variability. PC1\_Fonio is the first PC axis from the PCA analysis based on SNP frequencies taken from Abrouk et al. that represents the fonio genetic diversity. G\_Latitude, G\_Longitude, and G\_Altitude are the geographical factors (G). C\_Temperature (annual mean temperature), and C\_Precipitation (annual mean precipitation) are the climate conditions (C). S\_pH: the topsoil pH in water. S\_Density: the topsoil bulk density. S\_Gravel: the topsoil gravel content are soil variable (S)

fonio SNPs to be associated with roles in development, defense response, and cell-wall integrity (Table 1).

**Discussion**

Seeds harbor diverse bacterial endophytes, which can be composed of up to 2 billion bacterial cells with a complexity of up to 9000 species [9]. Many studies show that domestication decreased seed microbiome diversity [3, 33]. Fonio, being semi-domesticated, shows high species

richness as compared to other crops such as barley, rice, and wheat [4, 14]. A meta-study on 50 plant species worldwide revealed that the seed-core bacterial taxa may include *Pseudomonas*, *Rhizobium*, *Sphingomonas*, *Methylobacterium*, and *Paenibacillus* [51]. We identified these taxa among the core 191 taxa (Table S7) but did not find these taxa (absolute abundance > 1) in all accessions (Table S12). This variation may be attributed to cases where certain bacteria could only be partially assigned (Table S5).



**Table 1.** List of selected genes with annotated functions in other plant species

Fonio gene ID	Closest homologue in other plant species <sup>a</sup>	Annotated function	Functional category
Dexi2B01G0015860	XTH28: Xyloglucan endotransglucosylase/hydrolase protein 28( <i>At</i> )	Development, cell wall biogenesis	Development
Dexi4A01G0012020	KCR1: Very-long-chain 3-oxoacyl-CoA reductase 1 ( <i>At</i> )	Embryo development ending in seed dormancy	
Dexi4A01G0013860	CDKF-1: Cyclin-dependent kinase F-1 ( <i>Os</i> )	Maintenance of root meristem identity	
Dexi4B01G0002960	At4g22758: Uncharacterized protein At4g22758 ( <i>At</i> )	Regulation of the developmental process	
Dexi5A01G0007910	ABC17: Protein ABC17%2C chloroplastic ( <i>At</i> )	Embryo development ending in seed dormancy	
Dexi5B01G0013640	PDIL1-1: Protein disulfide isomerase-like 1-1 ( <i>At</i> )	Embryo development ending in seed dormancy	
Dexi7A01G0021150	NAGLU: Alpha-N-acetylglucosaminidase ( <i>At</i> )	Embryo development	
Dexi7B01G0005260	VPS11: Vacuolar protein-sorting-associated protein 11 ( <i>At</i> )	Development, vacuole organization	
Dexi9B01G0004890	MYB98: Transcription factor MYB98 ( <i>At</i> )	Embryo sac development	
Dexi4B01G0002990	CAF1-10: Probable CCR4-associated factor 1 homologue 10 ( <i>At</i> )	Defense response to bacterium	
Dexi7A01G0021190	WAK2: Wall-associated receptor kinase 2 ( <i>At</i> )	Pathogen response during development	
Dexi7B01G0005300	At1g67340: F-box protein At1g67340 ( <i>At</i> )	Defense response	
Dexi8B01G0010390	MO1: Monoxygenase 1 ( <i>At</i> )	Defense response	
Dexi8B01G0010430	RBOHC: Respiratory burst oxidase homologue protein C ( <i>St</i> )	Defense response/cell differentiation	
Dexi2B01G0015830	ANN5: Annexin D5 ( <i>At</i> )	Response to heat, light, and salt stress	Others
Dexi4A01G0011970	At5g55950: Nucleotide-sugar uncharacterized transporter 2 ( <i>At</i> )	Cell division	
Dexi4B01G0003040	Early nodulin-93 ( <i>Gm</i> )	Nodule formation in soybean	
Dexi5A01G0007950	WIP1: Bowman-Birk type wound-induced proteinase inhibitor WIP1( <i>Zm</i> )	Anatomical structure development	
Dexi7B01G0005270	At5g41620: USO1-like protein ( <i>At</i> )	Response to wounding and temperature	
Dexi7B01G0005290	CLPB1: Chaperone protein ClpB1 ( <i>At</i> )	Cellular response to hypoxia and heat	
Dexi9B01G0004870	DRIP1: E3 ubiquitin protein ligase DRIP1 ( <i>At</i> )	Response to water deprivation	

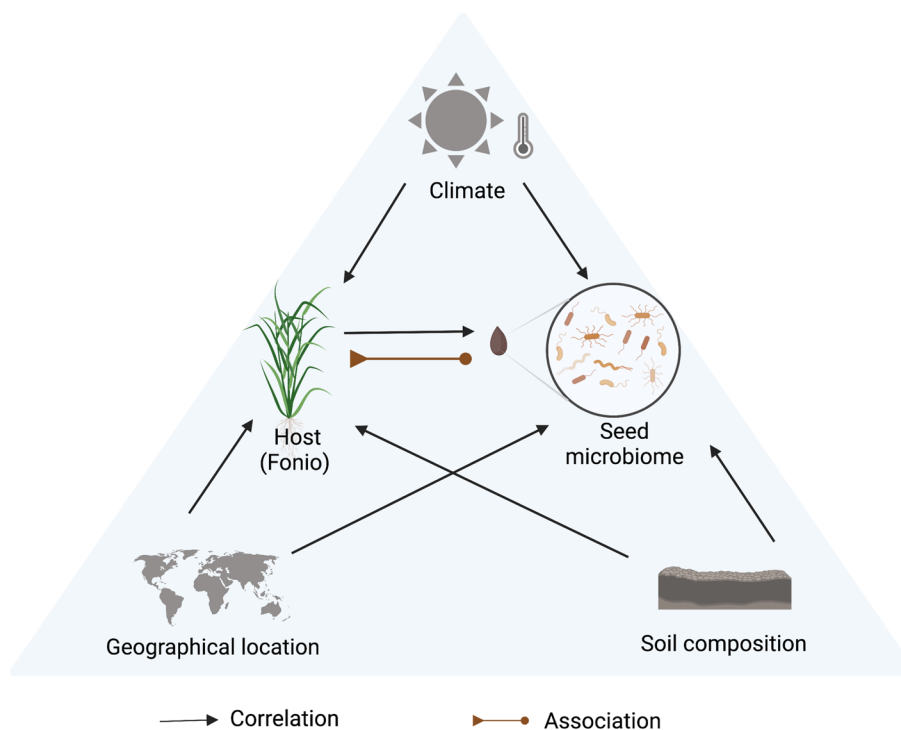
<sup>a</sup>At=*Arabidopsis thaliana*, Os=*Oryza sativa*, St=*Solanum tuberosum*, Zm=*Zea mays*, Gm=*Glycine max*

**Fig. 5** GWAS with the first axis of uniface weighted PCoA using a general linear model (GLM) and enrichment analysis. A The Manhattan plots display the association  $p$ -value for each SNPs in fonio. The red horizontal lines represent the  $p = 1e - 5$  significance threshold. Table 1 List of the fonio gene ID, closest homolog in other species, annotated function with categories

Currently, the role of endophytic bacteria in seeds is linked to plant growth and protection, yet the specific mechanisms remain to be elucidated in detail [30]. We have attempted to elucidate the potential functions of the seed-associated microbiomes in fonio. Our functional analysis revealed the prominence of 35 pathways within distinct groups which included both anabolic and catabolic processes. Notably, we also observed some of these pathways are over/underrepresented by certain groups indicating the endophytes may fine-tune their functions based on the local environment. These results imply that seed-associated endophytes generally support plant growth promotion through nutrient availability and assimilation, with potential adjustments influenced by soil and geo-climatic factors. Therefore, to identify the core endophytic functions across the fonio accessions, we performed the functional profiling of the core microbiome 191 strains, which showed the terms peptidase and inhibitors, secretion systems, and two core component systems. Previously, seed endophytes have been shown

to produce biocontrol inhibitors to protect plants from pathogens [39]. Notably, most of the seed endophytic isolates exhibited the production of extracellular enzymes [32] that may assist microorganisms in biofilm formation, combat nematodes, aid in plant nutrient acquisition and in the establishment of microbial communities within plants [21] (Chen et al., 2007) [42]. Also, two core component systems are essential for host-associated bacteria to establish colonization [58]. It is important to note that we have not validated our analysis through alternative methods such as metagenomics or transcriptomics. Together, our analysis is in line with previous studies and suggests the functions of core taxa can be crucial for establishing the endophytic community in the fonio population.

Several studies showed that environmental factors drive rhizosphere and phyllosphere microbiome assembly [16]. Also, host genotypes strongly and weakly shape the leaf and root microbiome, respectively [56]. However, the seed is a bottleneck in the generational continuity of the plant microbiome and may be crucial for plant



**Fig. 6** A model proposing climate, geographical location and soil composition may shape host genotype and seed-associated endophytes. A conceptual model illustrating the potential influences of climate, geographical location, soil composition, and the genetic makeup of the host plant to the composition of seed-associated endophytes ( $p > 0.01$ ). Host genetic loci is also associated ( $p = 1e-5$ ) with seed microbiome diversity

adaptation. We show in this study that environmental factors and soil properties correlate with the seed microbiome of fonio (Fig. 6), indicating that environmental factors may affect microbial composition and transgenerational transmission.

Previous GWAS showed that host genetic differences influence the composition and diversity of rhizosphere-associated bacteria in Arabidopsis, maize, and Sorghum [19, 40, 43]. Notably, the rhizosphere/leaf/root microbiome is complex and can vary according to environmental conditions [10]. GWAS was also performed with the leaf-associated bacteria of Arabidopsis, maize, and rice [24, 47, 57]. However, to our knowledge, GWAS of seed microbiome traits have not been conducted yet.

The seed-associated endophytic microbes are “selected” in continuity to the next generation; their use may be beneficial. However, seed-to-seed endophyte variability in common beans has also been reported [13]. The variation among the replicates (a pool of six seeds) in seven barley accessions has also been shown (Bziuk et al., 2021b). In our study, among 126 accessions of fonio, 24 accessions showed variation among replicates. Therefore, we used 99 accessions for GWAS analysis and acquired similar peaks with replicates. We found that genetic differences among fonio accessions correlate with

the diversity of the seed microbiome. Fonio SNPs associated within and surrounding genes with characterized roles in embryo development, stress/defense response, cell-wall integrity, and nucleic acid metabolism could directly or indirectly affect microbiome composition and diversity. The function of annotated candidate genes highlighted by our study aligns with previous studies. For example, the plant loci responsible for defense affect microbial community variation in the phyllosphere [25] and *RBOHD*, one of the candidate genes in our study, is required for phyllosphere microbiota homeostasis in Arabidopsis [44]. Another GWAS study with the Arabidopsis root microbiome identified host SNPs associated with genes that play roles in immunity, cell wall integrity, and development [11]. One of the genes that shape the bacterial richness and diversity in Arabidopsis is *FMO1* and is also among the candidate genes in our study [12]. We also identified a correlation of microbiome diversity with genes involved in embryonic development, a subject that should receive further attention. Overall, our data suggest that plant genetic loci regulate microbiome composition in a spatially and temporally dependent manner and future experiments with genetic mutants should help to unravel the underlying mechanisms. Taken together, although seed microbiome research is still in its infancy,

our work shows that it has great potential in advancing our understanding of the plant-microbiome-environment interaction and in seed microbiome engineering of crops.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-023-01725-5>.

**Additional file 1: Fig S1.** Microbial diversity in 18 fonio accessions (A) Rarefaction plot. (B) Observed features (ASVs). **Fig. S2.** Microbial diversity in 126 fonio accessions across six genetic group and one admixed group (A) Number and distribution of fonio accessions in group (B) Rarefaction plot (C) Histograms of the relative abundances of the top bacterial communities at Phylum (D) Principal components analysis of PICRUSt functional predictions on 126 fonio accessions. Color code is based on the corresponding groups). (E) The flower diagram shows unique and core taxa without admixed group. (F) Functional analysis of the core 191 ASVs for key KEGG pathways. The KO encoded for each subfunction is shown as a bar plot. **Fig. S3.** Phylogenetic tree and absolute abundance of 191 ASVs across the groups represented as stacked bar plot. Each group highlighted with distinct colors. **Fig. S4.** Geographical distribution of fonio accessions showing (A) topsoil sand fraction, (B) topsoil pH, (C) topsoil gravel content, (D) topsoil bulk density, (E) precipitation and (F) temperature. **Fig. S5.** (A) Quantile-Quantile (QQ) plots: the quantile distribution of observed  $p$ -values (on the y-axis) versus the quantile distribution of expected  $p$ -values. (B) The Manhattan plots display the association  $p$ -value for each SNPs in fonio. The red horizontal lines represent the  $P=1e-5$  significance threshold.

### Additional file 2.

## Acknowledgements

We would like to thank Dr. Abdul Aziz Eida for the useful discussion.

## Authors' contributions

HH conceived and directed the research. HIA and SGK provided the seeds and soil-geo-climatic data. NT, with the help of AHS, performed all the lab experiments. MMS received the sequencing data. HIA and SP performed the bioinformatic analysis. NT interpreted the results and wrote the manuscript with substantial inputs from other authors. NT, HIA, MMS, SGK and HH coordinated the research. All authors have read and approved the manuscript.

## Funding

The work was supported by a grant to HH from the King Abdullah University of Science and Technology BAS/1/1062-01-01.

## Availability of data and materials

The raw sequencing data used for bacterial composition are available on NCBI under Submission ID: SUB12628648 and Bioproject ID: PRJNA927337. This study project has been deposited at DDBJ/ENA/GenBank under the accession KHVQ00000000. The version described in this paper is the first version, KHVQ01000000.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Plant Science Program, Biological and Environmental Science and Engineering Division (BESE), King Abdullah University of Science and Technology

(KAUST), 23955-6900 Thuwal, Saudi Arabia. <sup>2</sup>Center for Desert Agriculture, Biological and Environmental Science and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), 23955-6900 Thuwal, Saudi Arabia.

Received: 5 April 2023 Accepted: 18 November 2023

Published online: 17 January 2024

## References

- Abdelfattah, A., Ayco, @, Tack, J. M., Lobato, C., Wassermann, B., & Berg, G. (2022). From seed to seed: the role of microbial inheritance in the assembly of the plant microbiome. *Trends in Microbiology*, 0(0). <https://doi.org/10.1016/J.TIM.2022.10.009>.
- Abdelfattah, A., Tack, A. J. M., Lobato, C., Wassermann, B., Berg, G. From seed to seed: the role of microbial inheritance in the assembly of the plant microbiome. *Trends Microbiol.* 2022. <https://doi.org/10.1016/J.TIM.2022.10.009>.
- Abdullaeva, Y., Ambika Manirajan, B., Honermeier, B., Schnell, S., Cardinale, M. Domestication affects the composition, diversity, and co-occurrence of the cereal seed microbiota. *J Adv Res.* 2020;31:75–86. <https://doi.org/10.1016/J.JARE.2020.12.008>.
- Abdullaeva, Y., Ratering, S., Ambika Manirajan, B., Rosado-Porto, D., Schnell, S., Cardinale, M. Domestication impacts the wheat-associated microbiota and the rhizosphere colonization by seed- and soil-originated microbiomes, across different fields. *Front Plant Sci.* 2022;12:3213. <https://doi.org/10.3389/FPLS.2021.806915/BIBTEX>.
- Abrouk, M., Ahmed, H. I., Cubry, P., Šimoníková, D., Cauet, S., Pailles, Y., Bettgenhaeuser, J., Gapa, L., Scarcelli, N., Couderc, M., Zekraoui, L., Kathiresan, N., Čížková, J., Hřibová, E., Doležel, J., Arribat, S., Bergès, H., Wieringa, J. J., Gueye, M., ... Krattinger, S. G. (2020). Fonio millet genome unlocks African orphan crop diversity for agriculture in a changing climate. *Nature Communications*, 11(1). <https://doi.org/10.1038/S41467-020-18329-4>.
- Adoukonou-Sagbadja, H., Schubert, V., Dansi, A., Jovtchev, G., Meister, A., Pistrick, K., Akpagana, K., & Friedt, W. (2007). Flow cytometric analysis reveals different nuclear DNA contents in cultivated Fonio (*Digitaria* spp.) and some wild relatives from West-Africa. *Plant Systematics and Evolution*, 267(1–4), 163–176. <https://doi.org/10.1007/S00606-007-0552-Z/METRICS>.
- Adoukonou-Sagbadja, H., Wagner, C., Ordon, F., & Friedt, W. (2010). Reproductive system and molecular phylogenetic relationships of fonio millets (*Digitaria* spp., Poaceae) with some polyploid wild relatives. *Tropical Plant Biology*, 3(4), 240–251. <https://doi.org/10.1007/S12042-010-9063-0/TABLES/3>.
- Ayenan, M. A. T., Sodedji, K. A. F., Nwankwo, C. I., Olodo, K. F., & Alladassi, M. E. B. (2018). Harnessing genetic resources and progress in plant genomics for fonio (*Digitaria* spp.) improvement. *Genetic Resources and Crop Evolution*, 65(2), 373–386. <https://doi.org/10.1007/S10722-017-0565-6/FIGURES/4>.
- Berg, G., Raaijmakers, J. M. Saving seed microbiomes. *ISME J.* 2018;12(5):1167. <https://doi.org/10.1038/S41396-017-0028-2>.
- Bergelson, J., Mittelstrass, J., & Horton, M. W. (2019a). Characterizing both bacteria and fungi improves understanding of the Arabidopsis root microbiome. *Scientific Reports* 2019 9:1, 9(1), 1–11. <https://doi.org/10.1038/s41598-018-37208-z>.
- Bergelson, J., Mittelstrass, J., & Horton, M. W. (2019b). Characterizing both bacteria and fungi improves understanding of the Arabidopsis root microbiome. *Scientific Reports* 2019 9:1, 9(1), 1–11. <https://doi.org/10.1038/s41598-018-37208-z>.
- Bergelson, J., Mittelstrass, J., & Horton, M. W. (2019c). Characterizing both bacteria and fungi improves understanding of the Arabidopsis root microbiome. *Scientific Reports* 2019 9:1, 9(1), 1–11. <https://doi.org/10.1038/s41598-018-37208-z>.
- Bintarti, A. F., Sulesky-Grieb, A., Stopnisek, N., Shade, A. Endophytic microbiome variation among single plant seeds. *Phytobiomes Journal*. 2022;6(1):45–55. <https://doi.org/10.1094/PBIOMES-04-21-0030-R/ASSET/IMAGES/LARGE/PBIOMES-04-21-0030-RF5.JPEG>.
- Bziuk, N., Maccario, L., Straube, B., Wehner, G., Sørensen, S. J., Schikora, A., Smalla, K. The treasure inside barley seeds: microbial diversity and

- plant beneficial bacteria. *Environmental Microbiomes*. 2021;16(1):1–21. <https://doi.org/10.1186/S40793-021-00389-8/FIGURES/4>.
15. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156. <https://doi.org/10.1093/BIOINFORMATICS/BTR330>.
  16. Dastogeer KMG, Tumpa FH, Sultana A, Akter MA, Chakraborty A. Plant microbiome—an account of the factors that shape community composition and diversity. *Current Plant Biology*. 2020;23: 100161. <https://doi.org/10.1016/J.CPB.2020.100161>.
  17. de Zélicourt, A., Synek, L., Saad, M. M., Alzubaidy, H., Jalal, R., Xie, Y., Andrés-Barrao, C., Rolli, E., Guerard, F., Mariappan, K. G., Daur, I., Colcombet, J., Benhamed, M., Depaepe, T., van der Straeten, D., & Hirt, H. (2018). Ethylene induced plant stress tolerance by *Enterobacter* sp. SA187 is mediated by 2-keto-4-methylthiobutyric acid production. *PLoS Genetics*, 14(3). <https://doi.org/10.1371/JOURNAL.PGEN.1007273>.
  18. Deng, S., Caddell, D. F., Xu, G., Dahlen, L., Washington, L., Yang, J., & Coleman-Derr, D. (2021a). Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *The ISME Journal* 2021 15:11, 15(11), 3181–3194. <https://doi.org/10.1038/s41396-021-00993-z>.
  19. Deng, S., Caddell, D. F., Xu, G., Dahlen, L., Washington, L., Yang, J., & Coleman-Derr, D. (2021b). Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *The ISME Journal* 2021 15:11, 15(11), 3181–3194. <https://doi.org/10.1038/s41396-021-00993-z>.
  20. Fitzpatrick, C. R., Salas-González, I., Conway, J. M., Finkel, O. M., Gilbert, S., Russ, D., Teixeira, P. J. P. L., & Dangl, J. L. (2020). The plant microbiome: from ecology to reductionism and beyond. 74, 81–100. <https://doi.org/10.1146/ANNUREV-MICRO-022620-014327>.
  21. Flemming H-C, Wingender J. The biofilm matrix. *Nat Rev Microbiol*. 2010;8(9):623–33. <https://doi.org/10.1038/nrmicro2415>.
  22. García-López R, Cornejo-Granados F, Lopez-Zavala AA, Cota-Huizar A, Sotelo-Mundo RR, Gómez-Gil B, Ochoa-Leyva A. Otus and asvs produce comparable taxonomic and diversity from shrimp microbiota 16s profiles using tailored abundance filters. *Genes*. 2021;12(4):564. <https://doi.org/10.3390/GENES12040564/S1>.
  23. Ghanem ME. Africa's forgotten crops could offset growing food insecurity. *Nature*. 2022;605(7908):30. <https://doi.org/10.1038/D41586-022-01185-1>.
  24. Horton, M. W., Bodenhausen, N., Beilsmith, K., Meng, D., Muegge, B. D., Subramanian, S., Vetter, M. M., Vilhjálmsson, B. J., Nordborg, M., Gordon, J. I., & Bergelson, J. (2014a). Genome-wide association study of Arabidopsis thaliana leaf microbial community. *Nature Communications* 2014 5:1, 5(1), 1–7. <https://doi.org/10.1038/ncomms6320>.
  25. Horton, M. W., Bodenhausen, N., Beilsmith, K., Meng, D., Muegge, B. D., Subramanian, S., Vetter, M. M., Vilhjálmsson, B. J., Nordborg, M., Gordon, J. I., & Bergelson, J. (2014b). Genome-wide association study of Arabidopsis thaliana leaf microbial community. *Nature Communications* 2014 5:1, 5(1), 1–7. <https://doi.org/10.1038/ncomms6320>.
  26. Jeske, J. T., & Gallert, C. (2022). Microbiome analysis via OTU and ASV-based pipelines—a comparative interpretation of ecological data in WWTP systems. *Bioengineering*, 9(4). <https://doi.org/10.3390/BIOENGINERING9040146/S1>.
  27. Cruz JF, Beavogui F. Fonio, an African Cereal. France: CIRAD; 2016.
  28. Johnston-Monje D, Mousa WK, Lazarovits G, Raizada MN. Impact of swapping soils on the endophytic bacterial communities of pre-domesticated, ancient and modern maize. *BMC Plant Biol*. 2014;14(1):1–19. <https://doi.org/10.1186/S12870-014-0233-3/FIGURES/6>.
  29. Johnston-Monje D, Raizada MN. Conservation and diversity of seed associated endophytes in zeas across boundaries of evolution, ethnography and ecology. *PLoS ONE*. 2011;6(6):e20396. <https://doi.org/10.1371/JOURNAL.PONE.0020396>.
  30. Jonkers, W., Gundel, P. E., Verma, S. K., & White, J. F. (2022). Editorial: seed microbiome research. *Frontiers in Microbiology*, 13. <https://doi.org/10.3389/fmicb.2022.943329>.
  31. Kaul S, Choudhary M, Gupta S, Dhar MK. Engineering host microbiome for crop improvement and sustainable agriculture. *Front Microbiol*. 2021;12:1125. <https://doi.org/10.3389/FMICB.2021.635917/BIBTEX>.
  32. Khalaf EM, Raizada MN. Taxonomic and functional diversity of cultured seed associated microbes of the cucurbit family. *BMC Microbiol*. 2016;16(1):131. <https://doi.org/10.1186/s12866-016-0743-2>.
  33. Kim H, Lee KK, Jeon J, Harris WA, Lee YH. Domestication of *Oryza* species eco-evolutionarily shapes bacterial and fungal communities in rice seed. *Microbiome*. 2020;8(1):1–17. <https://doi.org/10.1186/S40168-020-00805-0/FIGURES/6>.
  34. Kim, M., Swenson, J., McLoughlin, F., & Vierling, E. (2022). Mutation of the polyadenylation complex subunit CstF77 reveals that mRNA 3' end formation and HSP101 levels are critical for a robust heat stress response. *The Plant Cell*, 18. <https://doi.org/10.1093/PLCELL/KOAC351>.
  35. Koch M, Vorwerk S, Masur C, Sharifi-Sirchi G, Olivieri N, Schlaich NL. A role for a flavin-containing mono-oxygenase in resistance against microbial pathogens in Arabidopsis. *Plant J*. 2006;47(4):629–39. <https://doi.org/10.1111/J.1365-3113X.2006.02813.X>.
  36. Kohorn BD, Kohorn SL, Todorova T, Baptiste G, Stansky K, McCullough M. A dominant allele of arabidopsis pectin-binding wall-associated kinase induces a stress response suppressed by MPK6 but not MPK3 mutations. *Mol Plant*. 2012;5(4):841–51. <https://doi.org/10.1093/MP/SSR096>.
  37. Kuroda H, Takahashi N, Shimada H, Seki M, Shinozaki K, Matsui M. Classification and expression analysis of arabidopsis F-box-containing protein genes. *Plant Cell Physiol*. 2002;43(10):1073–85. <https://doi.org/10.1093/PCP/PCF151>.
  38. L'Hoir M, Duponnois R. Combining the seed endophytic bacteria and the back to the future approaches for plant holobiont breeding. *Frontiers Agronomy*. 2021;3:68. <https://doi.org/10.3389/FAGRO.2021.724450/BIBTEX>.
  39. Matsumoto H, Fan X, Wang Y, Kusstatscher P, Duan J, Wu S, Chen S, Qiao K, Wang Y, Ma B, Zhu G, Hashidoko Y, Berg G, Cernava T, Wang M. Bacterial seed endophyte shapes disease resistance in rice. *Nature Plants*. 2021;7(1):60–72. <https://doi.org/10.1038/s41477-020-00826-5>.
  40. Micallef SA, Shiaris MP, Colón-Carmona A. Influence of Arabidopsis thaliana accessions on rhizobacterial communities and natural variation in root exudates. *J Exp Bot*. 2009;60(6):1729–42. <https://doi.org/10.1093/JXB/ERP053>.
  41. Mishina TE, Zeier J. The Arabidopsis flavin-dependent monooxygenase FMO1 is an essential component of biologically induced systemic acquired resistance. *Plant Physiol*. 2006;141(4):1666. <https://doi.org/10.1104/PP.106.081257>.
  42. Paungfoo-Lonhienne C, Lonhienne TGA, Rentsch D, Robinson N, Christie M, Webb RI, Gamage HK, Carroll BJ, Schenk PM, Schmidt S. Plants can use protein as a nitrogen source without assistance from other organisms. *Proc Natl Acad Sci*. 2008;105(11):4524–9. <https://doi.org/10.1073/pnas.0712078105>.
  43. Peiffer JA, Spor A, Koren O, Jin Z, Tringe SG, Dangl JL, Buckler ES, Ley RE. Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc Natl Acad Sci USA*. 2013;110(16):6548–53. [https://doi.org/10.1073/PNAS.1302837110/SUPPL\\_FILE/PNAS.2013028371.PDF](https://doi.org/10.1073/PNAS.1302837110/SUPPL_FILE/PNAS.2013028371.PDF).
  44. Pfeilmeier, S., Petti, G. C., Bortfeld-Miller, M., Daniel, B., Field, C. M., Sunagawa, S., & Vorholt, J. A. (2021). The plant NADPH oxidase RBOHD is required for microbiota homeostasis in leaves. *Nature Microbiology* 2021 6:7, 6(7), 852–864. <https://doi.org/10.1038/s41564-021-00929-5>.
  45. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559. <https://doi.org/10.1086/519795>.
  46. Rawat, A. A., Hartmann, M., Harzen, A., Lukan, R., Stolze, S. C., Forzani, C., Abts, L., Reißweber, S., Rayapuram, N., Nakagami, H., Urgen Zeier, J., Hirt, H., & Zeier, U. (2022). <https://doi.org/10.1111/NPH.18592>.
  47. Roman-Reyna, V., Pinili, D., Borja, F. N., Quibod, I. L., Groen, S. C., Mulyaning-sih, E. S., Rachmat, A., Slamet-Loedin, I. H., Alexandrov, N., Mauleon, R., & Oliva, R. (2019). The rice leaf microbiome has a conserved community structure controlled by complex host-microbe interactions. *BioRxiv*, 615278. <https://doi.org/10.1101/615278>.
  48. Sajeev N, Baral A, America AHP, Willems LAJ, Merret R, Bentsink L. The mRNA-binding proteome of a critical phase transition during Arabidopsis seed germination. *New Phytol*. 2022;233(1):251–64. <https://doi.org/10.1111/NPH.17800>.
  49. Shade A, Jacques MA, Barret M. Ecological patterns of seed microbiome diversity, transmission, and assembly. *Curr Opin Microbiol*. 2017;37:15–22. <https://doi.org/10.1016/J.MIB.2017.03.010>.

50. Shahzad R, Khan AL, Bilal S, Asaf S, Lee IJ. What is there in seeds? Vertically transmitted endophytic resources for sustainable improvement in plant growth. *Front Plant Sci.* 2018;9:24. <https://doi.org/10.3389/FPLS.2018.00024/BIBTEX>.
51. Simonin M, Briand M, Chesneau G, Rochefort A, Marais C, Sarniguet A, Barret M. Seed microbiota revealed by a large-scale meta-analysis including 50 plant species. *New Phytol.* 2022;234(4):1448–63. <https://doi.org/10.1111/NPH.18037>.
52. Soucy S. M., Huang, J., & Gogarten, J. P. (2015). Horizontal gene transfer: building the web of life. *Nature Reviews Genetics* 2015 16:8, 16(8), 472–482. <https://doi.org/10.1038/nrg3962>.
53. Tan X, Wei J, Li B, Wang M, Bao Y. AtVps11 is essential for vacuole biogenesis in embryo and participates in pollen tube growth in *Arabidopsis*. *Biochem Biophys Res Commun.* 2017;491(3):794–9. <https://doi.org/10.1016/J.BBRC.2017.07.059>.
54. Toju, H., Peay, K. G., Yamamichi, M., Narisawa, K., Hiruma, K., Naito, K., Fukuda, S., Ushio, M., Nakaoka, S., Onoda, Y., Yoshida, K., Schlaeppli, K., Bai, Y., Sugiura, R., Ichihashi, Y., Minamisawa, K., & Kiers, E. T. (2018). Core microbiomes for sustainable agroecosystems. *Nature Plants* 2018 4:5, 4(5), 247–257. <https://doi.org/10.1038/s41477-018-0139-4>.
55. Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T., & Singh, B. K. (2020). Plant–microbiome interactions: from community assembly to plant health. *Nature Reviews Microbiology* 2020 18:11, 18(11), 607–621. <https://doi.org/10.1038/s41579-020-0412-1>.
56. Wagner, M. R., Lundberg, D. S., del Rio, T. G., Tringe, S. G., Dangl, J. L., & Mitchell-Olds, T. (2016). Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nature Communications*, 7. <https://doi.org/10.1038/NCOMMS12151>.
57. Wallace JG, Kremling KA, Kovar LL, Buckler ES. Quantitative genetics of the maize leaf microbiome. *Phytobiomes Journal.* 2018;2(4):208–24. <https://doi.org/10.1094/PBIOMES-02-18-0008-R/ASSET/IMAGES/LARGE/PBIOMES-02-18-0008-RT3.JPEG>.
58. Wiesmann CL, Zhang Y, Alford M, Hamilton CD, Dosanjh M, Thoms D, Dostert M, Wilson A, Pletzer D, Hancock REW, Haney CH. The ColR/S two-component system is a conserved determinant of host association across *Pseudomonas* species. *ISME J.* 2023;17(2):286–96. <https://doi.org/10.1038/s41396-022-01343-3>.
59. Zhang X, Ma YN, Wang X, Liao K, He S, Zhao X, Guo H, Zhao D, Wei HL. Dynamics of rice microbiomes reveal core vertically transmitted seed endophytes. *Microbiome.* 2022;10(1):216. <https://doi.org/10.1186/S40168-022-01422-9/FIGURES/7>.
60. Zhu F. Fonio grains: Physicochemical properties, nutritional potential, and food applications. *Comprehensive Reviews in Food Science and Food Safety.* 2020;19(6):3365–89. <https://doi.org/10.1111/1541-4337.12608>.
61. Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. <https://doi.org/10.1038/nbt1325>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.