

# Native edaphoclimatic regions shape soil communities of crop wild progenitors

María José Fernández-Alonso<sup>1,2,\*</sup>, Miguel de Celis<sup>3</sup>, Ignacio Belda<sup>4</sup>, Javier Palomino<sup>1</sup>, Carlos García<sup>5</sup>, Juan Gaitán<sup>6</sup>, Jun-Tao Wang<sup>7,8,9</sup>, Luis Abdala-Roberts<sup>10</sup>, Fernando D. Alfaro<sup>11</sup>, Diego F. Angulo-Pérez<sup>12</sup>, Manoj-Kumar Arthikala<sup>13</sup>, Danteswari Chalasani<sup>14</sup>, Jason Corwin<sup>15</sup>, Duan Gui-Lan<sup>16</sup>, Antonio Hernandez-Lopez<sup>13</sup>, Kalpana Nanjareddy<sup>13</sup>, Siddaiah Chandra Nayaka<sup>17</sup>, Babak Pasari<sup>18</sup>, Thanuku Samuel Sampath Kumar Patro<sup>19</sup>, Appa Rao Podile<sup>14</sup>, Teresa Quijano-Medina<sup>10</sup>, Daniela S. Rivera<sup>11</sup>, Pullabhotla Venkata Subba Rama Narshima Sarma<sup>14</sup>, Salar Shaa<sup>20</sup>, Pankaj Trivedi<sup>15</sup>, Qingwen Yang<sup>21</sup>, Yue Yin<sup>16,22</sup>, Eli Zaady<sup>23</sup>, Yong-Guan Zhu<sup>16</sup>, Brajesh K. Singh<sup>7,8</sup>, Manuel Delgado-Baquerizo<sup>22,\*</sup>, Pablo García-Palacios<sup>3,24,\*</sup>, Ruben Milla<sup>1,25,\*</sup>

<sup>1</sup>Area of Biodiversity and Conservation, Department of Biology and Geology, Physics and Inorganic Chemistry, Rey Juan Carlos University, C/ Tulipán s/n, 28933 Móstoles, Spain

<sup>2</sup>Departamento de Geología y Geoquímica, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain

<sup>3</sup>Departamento de Suelo, Planta y Calidad Ambiental, Instituto de Ciencias Agrarias, Consejo Superior de Investigaciones Científicas, 28006 Madrid, Spain

<sup>4</sup>Department of Genetics, Physiology and Microbiology, Faculty of Biology, Complutense University of Madrid, 28040 Madrid, Spain

<sup>5</sup>Department of Soil and Water Conservation and Organic Waste Management, CEBAS-CSIC, 30100 Murcia, Spain

<sup>6</sup>Instituto de Suelos - INTA Castelar, CONICET, Universidad Nacional de Luján, 6700 Luján, Buenos Aires, Argentina

<sup>7</sup>Hawkesbury Institute for the Environment, Western Sydney University, Penrith, NSW 2751, Australia

<sup>8</sup>Global Centre for Land-Based Innovation, Western Sydney University, Penrith, NSW 2751, Australia

<sup>9</sup>School of Science, Western Sydney University, Penrith, NSW 2751, Australia

<sup>10</sup>Departamento de Ecología Tropical, Campus de Ciencias Biológicas y Agropecuarias, Universidad Autónoma de Yucatán, 97000 Mérida, Yucatán, Mexico

<sup>11</sup>GEMA Center for Genomics, Ecology and Environment, Universidad Mayor, Santiago 8580745, Chile

<sup>12</sup>Unidad de Recursos Naturales, Centro de Investigación Científica de Yucatán, 97205 Mérida, Yucatán, Mexico

<sup>13</sup>Ciencias Agrogenómicas, Escuela Nacional de Estudios Superiores Unidad León, Universidad Nacional Autónoma de México, León 37689, Guanajuato, Mexico

<sup>14</sup>Department of Plant Sciences, School of Life Science, University of Hyderabad, Hyderabad, 500046, Telangana, India

<sup>15</sup>Microbiome Network and Department of Agricultural Biology, Colorado State University, Fort Collins, CO 80523-1177, United States

<sup>16</sup>State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

<sup>17</sup>Department of Studies in Biotechnology, University of Mysore, Manasagangotri, Mysuru 570006, Karnataka, India

<sup>18</sup>Department of Agronomy and Plant Breeding, Islamic Azad University, Sanandaj Branch, Sanandaj 6616935391, Iran

<sup>19</sup>Agricultural Research Station (ACRIP center-small millets), Acharya NG Ranga Agricultural University, Vizianagaram 535001, Andhra Pradesh, India

<sup>20</sup>Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), 06466 Seeland, Germany

<sup>21</sup>National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

<sup>22</sup>Laboratorio de Biodiversidad y Funcionamiento Ecosistémico, Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), CSIC, 41012 Sevilla, Spain

<sup>23</sup>Katif Research & Development Center, Sdot Negev, Netivot 8771002, Israel

<sup>24</sup>Department of Plant and Microbial Biology, University of Zurich, CH-8008 Zurich, Switzerland

<sup>25</sup>Global Change Research Institute, Rey Juan Carlos University, 28933 Móstoles, Spain

\*Corresponding authors: María José Fernández Alonso, Departamento de Geología y Geoquímica, Universidad Autónoma de Madrid, Avd. Francisco Tomás y Valiente 7, 28049 Madrid, Spain. E-mail: mariajose.fernandez@uam.es; Manuel Delgado-Baquerizo, Laboratorio de Biodiversidad y Funcionamiento Ecosistémico, Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), CSIC, Av. Reina Mercedes 10, 41012 Sevilla, Spain. E-mail: m.delgadobaquerizo@gmail.com; Pablo García-Palacios, Departamento de Suelo, Planta y Calidad Ambiental, Instituto de Ciencias Agrarias, Consejo Superior de Investigaciones Científicas, C. de Serrano 115b, 28006 Madrid, Spain. E-mail: pablo.garcia@ica.csic.es; Rubén Milla, Area of Biodiversity and Conservation, Department of Biology and Geology, Physics and Inorganic Chemistry, Universidad Rey Juan Carlos, C. Tulipán s/n, 28933 Móstoles, Spain. E-mail: ruben.milla@urjc.es

## Abstract

Unveiling the soil biological communities ecologically associated with crop wild progenitors (CWPs) in their habitats of origin is essential for advancing productive and sustainable agriculture. A field survey was conducted to investigate the edaphoclimatic conditions and soil bacterial, fungal, protist, and invertebrate communities of 125 populations of direct progenitors of major crops for world agriculture. The wild populations clustered into four ecoregions shaped by two edaphoclimatic dimensions: one summarizing variations in soil sand contents and nutrients concentrations, and the other featuring changes in aridity, soil pH, and carbon storage potential. We identified a common soil core community across CWPs that varied significantly along deserts to tropical seasonal forests and savannas. The assembly of the soil core community was driven by varying environmental preferences amongst soil biodiversity kingdoms, reflecting potential shifts in their functional profiles. The tropical ecoregion exhibited higher proportion of acidophilic bacteria, fungal, and protist parasites, whilst desert ecosystems harboured greater abundances of saprophytic fungi and heterotrophic protists. Moreover, CWPs displayed unique microhabitats that incorporate variability into the soil community assembly. Our work reveals the biogeography of soil communities associated with CWPs, the first step towards the development of microbial rewilding initiatives.

**Keywords:** centres of origin; crop wild progenitors; ecoregions; edaphoclimate; rewilding; soil core microbiome; soil core microfauna

Received: 25 November 2024. Revised: 20 June 2025. Accepted: 12 August 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of the International Society for Microbial Ecology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## Introduction

The global food system relies on a small set of crops that trace their origins to specific geographic regions around the world [1]. In these native habitats, the crop wild progenitors (CWPs) and their associated soil microbiomes have historically co-evolved, establishing interactions that influence plant performance and ecosystem functioning [2]. Hereby, the soil microbiome associated with the wild progenitors of major crops is an untapped resource to meet global agricultural challenges [3, 4]. However, today, we lack the most basic understanding on the soil microbiome at the crop centres of origin, especially when considering multiple domestication regions and crops simultaneously [5]. The processes of plant domestication and farming have altered host-microbial relationships in modern crops [6], undermining the functionality, resilience, and sustainability of croplands [7]. Reinstating part of the ancient biotas of CWPs that support beneficial interactions in modern crops constitutes the foundations of the microbiome rewilding framework [8]. For this to happen, assessing the environmental preferences of soil biodiversity at the habitats of origin of major crops is a first step.

Soils are the most biodiverse habitat on Earth [9], and are the principal inoculum source of microbes and invertebrates colonizing the rhizosphere and root tissues [10, 11]. Soil and climatic factors are of paramount importance in determining soil community assembly and further host plant control on microbiome recruitment [12]. However, most studies addressing the microbiome of CWPs do not focus on native soils under the environmental conditions in which wild populations thrive [6, 13]. Notably, one study revealed that certain taxa present in native soils of a CWP were absent in cropland soils, with the latter exhibiting increased diversity but a simpler network structure, suggesting a homogenized and less specialized soil microbiome [14]. Microbial taxa associated with CWPs may display traits linked to local conditions and stressors in their habitats of origin that are absent in modern agricultural soils [8, 14]. Hence, to take advantage of the microbiomes of CWPs to improve current crops, we must understand their distribution and ecology in order to develop a baseline for more applied research. However, a comprehensive biogeographical investigation of soil biodiversity patterns across multiple kingdoms is absent. This approach can identify key members of the soil microbiota of CWPs to be reinstated using microbial inoculants and synthetic communities that help current crops cope with changing environmental conditions [7].

The joint characterization of the edaphoclimatic context and the associated soil biota amongst CWPs allows the establishment of regions with interrelatedness ecological components. It is particularly meaningful to identify sets of soil biotic taxa that most effectively colonize the native environments and thereby are dominant, and also being associated with multiple CWPs, suggesting host-microbe relationships with important functional roles [15]. This ecoregion perspective can assist in identifying which components of the soil core microbiome of CWPs could be most effectively and successfully transferred to modern agroecosystems [16, 17]. Additionally, environmental characteristics at the sites of origin of CWPs may provide underlying factors modulating the evolution and breeding of modern crop phenotypes and, consequently, agroecosystem functioning [18]. Yet, the ecological legacy of modern crops from their sites of origin is typically overlooked in agricultural studies [19]. Investigating in which ecoregions CWPs inhabit and how they relate with the associated soil core community may offer a novel framework for better explaining the effects of domestication on plant phenotypes, and for microbial rewilding studies in agriculture.

Here, we investigated the biogeography and edaphoclimatic factors associated with the soil core microbiome and micro-fauna at the habitats of origin of major crops. We performed a global standardized field survey at 125 wild populations across the distribution ranges of the recognized CWPs of 10 major crops (maize, rice, wheat, potato, bean, little millet, barley, sunflower, soya, and cotton; Fig. 1), which account for 40% of the total area annually harvested worldwide and 61% of the global production of primary crops in the period 2012–21 [20, 21]. Our survey compiles information on climate, potential ecosystem carbon sinks, soil parameters, and the core taxa and co-occurrence networks of soil bacteria, fungi, protists, and invertebrates from six domestication centres (North America, Mesoamerica, Central Andes, Fertile Crescent, Indomalaya region, Yangtze and Yellow River valleys). We addressed the following questions: (i) what is the edaphoclimatic niche of CWPs in their native habitats?, (ii) is there a common soil core community (microbes and microfauna) embracing different CWPs and ecoregions?, and (iii) which edaphoclimatic factors determine the soil biodiversity of CWPs?

## Materials and methods

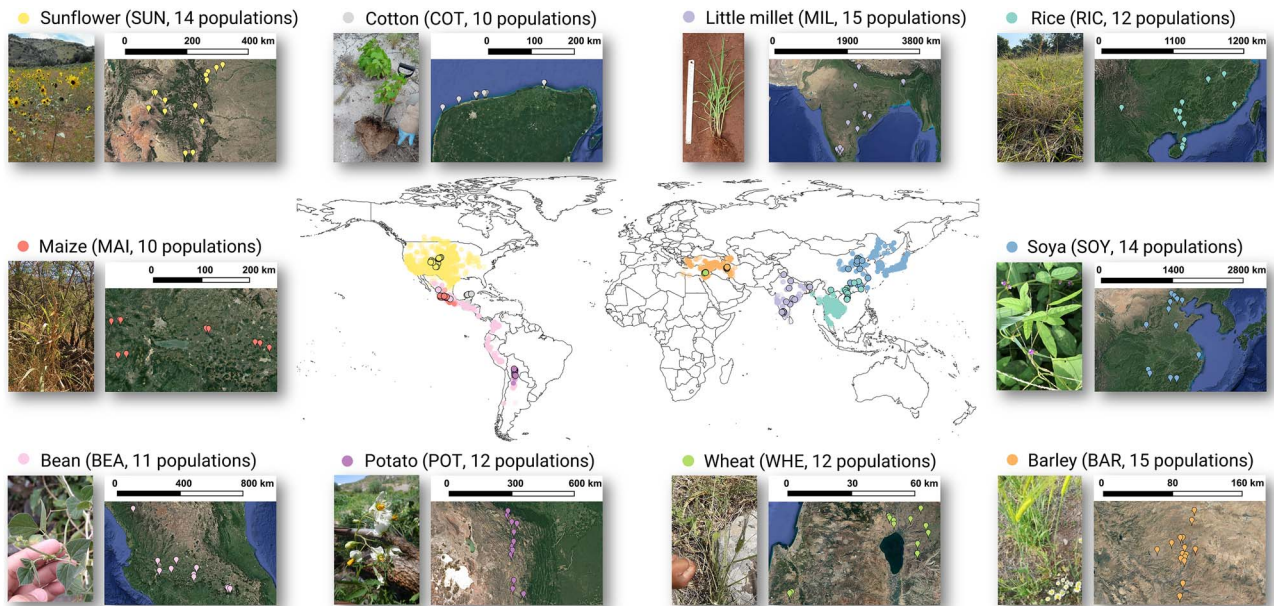
To investigate the soil biogeography at the habitats of origin of major crops, we sampled 125 wild populations of 10 species of CWPs that encompassed broad geographic and environmental gradients. For each population, we gathered climate, potential ecosystem carbon sinks (i.e., energy availability to trophic networks), soil fertility, and physiochemical data. We measured soil bacteria, fungi, protist, and invertebrate communities through amplicon sequencing. We used principal component and module analyses to characterize the sites of origin of CWPs by distinct edaphoclimatic features (i.e., ecoregions). Then, we identified the core members of ancestral agriculture using abundance (10% more abundant phylotypes) and prevalence (phylotypes present in the populations of at least four CWPs) criteria. We calculated co-occurrence networks to identify the main modules of the soil biotic community and assessed the relative abundance patterns across ecoregions. Finally, we used linear mixed effects models to identify the environmental drivers responsible for clustering patterns of the soil core members of ancestral agriculture.

### Crop wild progenitors

We selected 10 major crops of global relevance (maize, rice, wheat, potato, bean, little millet, barley, sunflower, soya, and cotton), including a woody crop, a tuberous crop, and several herbaceous crops harvested for their seeds. We used the Crop Origins and Phylo Food database to identify crops' geographic origin and their most likely wild progenitor ([https://rubenmilla.github.io/Crop\\_Origins\\_Phylo/](https://rubenmilla.github.io/Crop_Origins_Phylo/), [22]). For each CWP, we selected 10–15 truly wild populations in habitats across its native range of distribution, maximizing the environmental variation amongst populations (altitudinal/latitudinal or other edaphoclimatic gradients). In each population, we surveyed a 20 m × 20 m plot. The mean separation distance between native populations within each species was 364 km (ranging from 1 to 2265 km). In total, we sampled 125 populations of CWPs across disparate regions of the world (Fig. 1).

### Soil sampling

In 2020 and 2021, we collected native soil samples from each wild population, focusing on soil communities associated with CWPs. This soil compartment is widely used in ecological and microbiome studies [23], harbours the highest biodiversity and



**Figure 1.** Global distribution of the wild progenitors of the 10 selected crops. Distributions are denoted as borderless points, whilst the specific populations surveyed in this study are bordered points. Global occurrences were retrieved from the global biodiversity information facility database [20]. The common names of modern crops designate the wild populations surveyed in this project. The abbreviations of the crop names, and the total number of surveyed populations are given in brackets. The wild progenitors of the crops are: *Gossypium hirsutum* L., cotton; *Glycine max* subsp. *soja* (Siebold & Zucc) H. Ohashi, soja; *Helianthus annuus* L., sunflower; *Hordeum vulgare* subsp. *spontaneum* (K. Koch) Asch. & Graebn., barley; *Oryza rufipogon* Griff., rice; *Panicum sumatrense* Roth, little millet; *Phaseolus vulgaris* L., common bean; *Solanum berthaultii* Hawkes, potato; *Triticum dicoccoides* (Körnicker) G. Schweinfurth, wheat; *Zea mays* subsp. *parviglumis* Iltis & Doebley, maize. Images of wild populations taken during field campaigns. Landsat images from Google earth pro (<https://earth.google.com/web>). Sunflower image retrieved from [58].

stands as the foremost source of plant microbes [24]. We selected five plants each spaced at least 5 m apart during the flowering stage, which is a period that facilitates the identification of CWPs, guarantees their active growing, and is characterized by key plant–soil interactions [24]. We uprooted the plants along with an entire soil monolith within a 10-cm radius and to a depth of 20-cm. We collected the soil that was not intimately attached to the plant roots by shaking off any soil not tightly adhering to them [23]. We combined soil samples from each plot to create a composite and homogeneous sample for each population. We immediately sieved the soil samples at 2 mm and stored them at 4°C until shipment to the Universidad Rey Juan Carlos (URJC, Madrid, Spain) for processing and analysis. Once in Spain, we froze a subset of each soil sample at –20°C until DNA extraction. We air dried and stored the remaining of each sample for physiochemical analyses. Little millet samples could not leave the country due to legal constraints and they were processed at the University of Hyderabad (Telangana, India) and extracted for soil DNA. Then, soil physiochemical properties were analysed at the International Crops Research institute for the Semi-arid tropics (ICRISAT, Hyderabad, India), following the same procedures applied to the rest of the CWPs.

### Soil physiochemical analyses

We analysed electrical conductivity and pH in deionized water extracts (suspension ratio of 1:5 and 1:2.5 weight:volume, respectively). We estimated soil texture (percentage of sand) using the Bouyoucos hydrometer method [25]. We determined the total soil carbon and nitrogen concentrations by the Dumas dry combustion procedure using an elemental analyser (C/N Flash EA 112 Series-Leco Truspec). We determined the soil organic carbon (SOC) by dry combustion in 2 N HCl pre-treated samples, except for soil samples of little millet that were analysed by the

Walkley–Black method. We converted SOC data obtained with the Walkley–Black to dry combustion SOC using an established conversion factor of 1.15 [26]. We extracted total concentrations of soil micronutrients (Ca, Na, K, B, Mg, Mn, Cu, Fe, and Zn) using wet acid digestion and determined them using inductively coupled plasma optical emission spectrometry (ICAP 6500 DUO ICP-OES spectrometer, Thermo-Scientific) according to ISO 11885:1996. We analysed mineral nitrogen and available phosphorus (in the form of  $\text{NH}_4^+$ ,  $\text{NO}_3^-$  and  $\text{PO}_4^-$ ) in soil water extracts by ionic exchange chromatography with a cation-exchange column (Metrosep A Supp 5–250/4.0, Herisau).

### Amplicon sequencing

We extracted soil DNA from 0.5 g of a sample using the DNeasy PowerSoil Kit (Qiagen) according to the manufacturer's instructions. We checked the quality and quantity of DNA using a NanoDrop 2000 (Thermo Fisher Scientific, USA) and Qubit Fluorometer (Thermo Fisher Scientific, USA). We determine the biotic community profiling through amplicon sequencing on the Illumina MiSeq platform of the Next-Generation Sequencing Facility at the Western Sydney University (Richmond, NSW, Australia). For the library preparation, we used the primer set 799F/1193R to amplify the bacterial 16S rRNA gene (V5–V7 region), which effectively reduces the amplification of host DNA (e.g. chloroplasts and other organellar sequences) and improves bacterial community detection [27, 28]. The eukaryotic 18S rRNA gene Euk1391f/EukBr was sequenced to study both protist and invertebrates [29]. We used the primer sets FITS7/ITS4 to amplify the nuclear ribosomal internal transcriber spacer 2 (ITS2 region), which is a common maker for fungal communities [30].

We trimmed primers from paired raw reads and performed an end-trim (base quality <26) to increase the rate of successful merge. We merged the resultant paired reads using USEARCH

[31] and FLASH [32] and discarded sequences of low quality (maximum expected error of 1.0). We denoised the sequences (error-correction) and clustered them into zOTUs (100% sequence similarity) using UNOISE3 [33]. We annotated representative sequences against the UNITE database (2022.05) using QIIME2 [34] for ITS data, and the SILVA database for 16S and 18S rDNA data, respectively. We rarefied at 12000, 8000, 5000, and 250 reads per sample for bacteria, fungi, protists, and invertebrates, respectively (Fig. S1). We explored the potential lifestyle of the soil fungal community using the FungalTraits v1.2 database based on the fungal ITS data [35].

## Environmental data

We gathered climate and ecosystem productivity data, for the coordinates of each population, from global databases. We obtained climatic variables [mean annual precipitation, mean annual temperature (MAT), precipitation seasonality (PSEA), temperature seasonality, and mean diurnal range temperature (MDR)] from the WorldClim dataset version 2 [36], and the UNEP aridity index (AI = precipitation/potential evapotranspiration, where the higher the index the lower the aridity) from the Global Aridity Index and Potential Evapotranspiration Climate Database version 2 [37]. We used the Normalized Difference Vegetation Index (NDVI) as a proxy for net aboveground primary productivity, based on values from the MODIS satellite imagery MOD13Q1. We followed a standard procedure for NDVI calculation at a global scale. We acquired 23 NDVI values per year with a pixel size of 250 × 250 m and used them to calculate annual NDVI data for each year in the period from 2000 to 2021. To do so, we averaged the product values between the date of the minimum NDVI ( $n$ ) and the date  $n - 1$  of the following year at each site. MODIS data are geometrically and atmospherically corrected and include a reliability index of data quality based on the environmental conditions in which the data were recorded.

## Data analyses

We carried out exploratory data analysis to inspect the structure of missing values (4% of the environmental database). We imputed the multivariate missing data using the *mice* package and the predictive mean matching model with 999 iterations [38]. Data imputation did not alter the distribution of original data as shown in density plots (Fig. S2).

## Edaphoclimatic distribution of CWPs

To outline the bioclimatic characterization of the populations of the CWPs surveyed, we used Whittaker's biomes classification [39], the aridity index and NDVI (an indirect measure of vegetation cover and primary productivity). To evaluate the edaphoclimatic differences amongst the populations surveyed, we ran a Principal Component Analysis (PCA) on 18 centred and scaled local soil and climate variables. The variables with the highest square cosine were used to interpret the PCA axes (*FactoMineR* package [40], Fig. S3). To describe ecoregions with similar edaphoclimatic conditions, we performed a module analysis based on the first 5 PCA axes, which captured 74% of the total variance. We multiplied the PCA axes by the square root of their eigenvalues to weight the classification outcome according to their importance. We calculated Euclidean distances to overcome the multicollinearity of the edaphoclimatic variables and exploit their orthogonal properties. We used the K-means algorithm and the Calinski–Harabasz index to select the optimal number of groups (“*cascadeKM* function”, *vegan* package [41], Fig. S4). Clustering results were represented using elliptic shapes based on multivariate t-distribution. To test

the differences in the dimensions of the edaphoclimatic space by CWPs, we used the Wilcoxon signed-rank test and linear models with the applied Bonferroni-estimated marginal mean comparisons as post-hoc analyses.

## Assignment of soil core community

Given our global focus, we identified the soil core microbiome and microfauna of ancestral agriculture based on both relevance and prevalence criteria [15]. First, we kept only the 10% most common phylotypes in terms of average relative abundance. Then, within that subset of commonest phylotypes, we selected phylotypes present in at least four CWPs (i.e., 40% of the plant species surveyed). We conducted these analyses independently for bacteria, fungi, protists, and invertebrates. The resulting abundance tables contain phylotypes that (i) make a substantial contribution to ecosystem functionality, (ii) are prevalent and thereby adapted to contrasting climates and soils, thus likely providing considerable versatility for their agricultural application, and (iii) reduce potential spurious associations of rare taxa, since zero-inflated databases may confound ecological interpretations in downstream co-occurrence network analyses [42–44]. To enable the identification of overlap and uniqueness patterns of phylotypes in the soil core community across CWPs, we used Upset plots (*ComplexHeatmap* package [45]).

## Structure of the community: co-occurrence network analyses

To explore the structure of the soil core community of ancestral agriculture, we conducted co-occurrence network analyses. We calculated potential co-occurrences between core phylotypes using an improved version of the function originally described in de Celis et al. [46], which applies a probabilistic model to presence-absence matrices [47]. This approach yielded a list of phylotypes pairs that co-occurred more frequently than expected by chance. We built independent networks for bacteria, fungi, protists, and invertebrates, where nodes were phylotypes and edges were significant pairwise co-presence between nodes ( $P$ -value  $< .05$ ). We determined modularity and clusters (modules of dense connections that are sparsely connected to other areas in the network) using a community finding algorithm via short random walks (“*module\_walktrap* function” from *igraph* package [48]). The modules can indicate soil microorganisms that potentially interact or share environmental preferences, although other mechanisms responsible for community assemblies such as dispersion, diversification, and drift could also underlie the network structure. To investigate whether biogeographical patterns exist in the overall network structures, we computed the proportion of nodes of a given module that were present in wild populations of CWP in each ecoregion (module completeness). Given the unequal number of phylotypes across biotic modules (Table S1), we calculated the standardized relative abundance [ $z$ -score,  $z_{ij}$  in Equation (1)] of phylotypes and averaged them per module and wild population. This standardization allows to obtain a balanced contribution of the phylotypes (i.e. nodes) to the relative abundances of the modules in the networks for downstream analyses.

$$z_{ij} = \frac{(x_{ij} - \bar{x}_i)}{\sigma_i} \quad (1)$$

Where  $z_{ij}$  represents the number of reads for the phylotype  $j$  in the wild population  $i$ ,  $\bar{x}_i$  is the average number of reads for wild population  $i$ , and  $\sigma_i$  is the standard deviation of the reads for the wild population  $i$ . To test for the differences in the relative

abundance of each module across CWP, we fit linear models and applied Bonferroni-estimated marginal mean comparisons as *post hoc* analyses.

### Environmental factors determining modules of soil core members

To test if and how the modules in the resulting networks were influenced by environmental factors, we used linear mixed models (LMMs), four groups of predictors (climate: MAT, MDR, PSEA, AI; soil properties: pH, conductivity, sand; soil micronutrient concentrations:  $\text{NH}_4^+$ ,  $\text{NO}_3^-$ ,  $\text{PO}_4^-$ , Ca, B, K, Mg, Zn, Fe, Mn, Mo; and potential ecosystem carbon sink: mean NDVI, SOC), and the relative abundance (z-score) of the modules as the dependent variable. We retained the above-mentioned subset of variables after checking for multicollinearity patterns in our dataset of environmental predictors by means of the variance inflation factor and using a threshold level of five (*regclass* package, [49]). To deal with the hierarchical design of the survey, we used the identity of the CWPs as a categorical random intercept in the model. The random factor accounts for local drivers affecting microbial community assembly that were not measured and that may be directly attributed to the identity of the CWP species. We optimized the structure of the fixed components in the LMMs through a backward selection procedure of the predictor variables according to the Akaike Information Criterion (*mass* package [50]). Model coefficients were estimated using the restricted maximum likelihood criterion (*lme4* package [51]) and 95% bootstrap confidence intervals were calculated based on 1000 simulations (*jtools* package [52]). To interpret the outcome of the LMMs, we standardized the predictors and decomposed the marginal coefficient of determination ( $\text{Rm}^2$ ) into unique and shared variance by each group of fixed predictors (*glmm.hp* package [53]). We checked LMM assumptions visually and using the Anderson–Darling normality test. All statistical analyses and visualizations using R version 4.2.4.

## Results and discussion

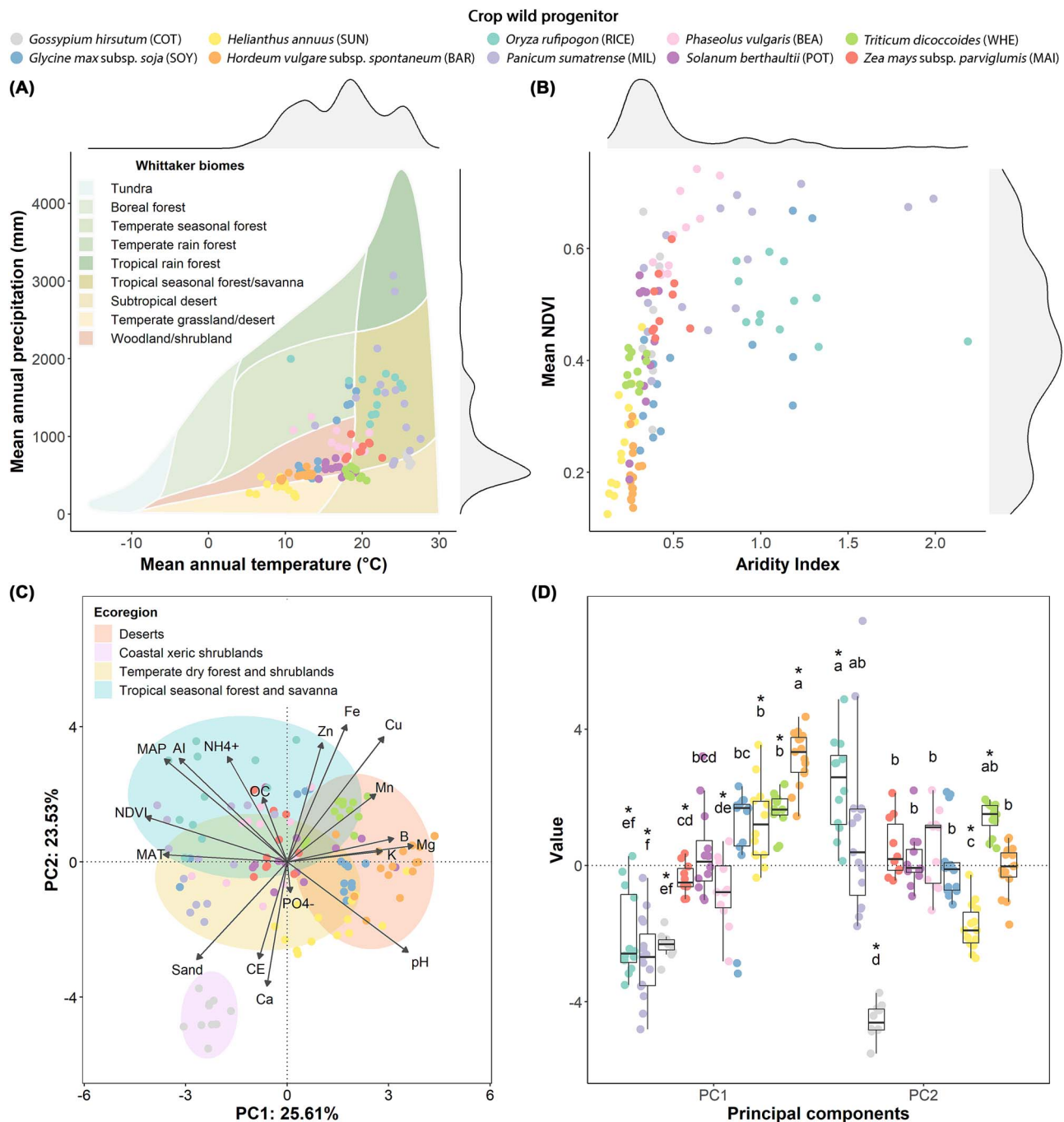
Our standardized field survey provides the first joint assessment of the soil core community—including bacterial, fungal, protist, and invertebrate phylotypes—and edaphoclimatic components of the most relevant CWPs investigated across their original distribution worldwide. The wild populations of our CWPs clustered into four ecoregions that differed widely in their edaphoclimatic attributes, such as water and energy supply, soil pH, texture, and fertility. Environmental variation amongst ecoregions suggests that the ecophysiological traits of CWPs are likely to differ amongst the more oligotrophic coastal shrubland ecoregion, deserts, and tropical seasonal forests and savannas. We identified the soil core community common to all CWPs, which were dominated by the major soil phyla commonly found in studies surveying soil ecosystems worldwide. Despite this commonality, we also found distinct ecological groups of soil core phylotypes that showed variation in their prevalence across ecoregions, driven by varying environmental preferences amongst kingdoms. These biogeographical patterns of the soil core microbes and microfauna point to changes in their life history strategies, showing increased proportional abundances of acidophilic bacteria, as well as fungal and protist parasites in the tropical seasonal forest and savanna ecoregion. Wild populations displayed specific microhabitats within ecoregions that could play a selective role in the assembly of the soil core microbiome. These findings improve our understanding of the ecology of CWPs and their associated soil communities, which holds significant implications for developing strategies for rewilding in agriculture.

## Ecoregions of crop wild progenitors

Our study provides a biogeographical characterization of the evolutionary origins of domesticated crops based on 125 populations of 10 CWPs (Fig. 1 and Table S2), distributed across all Whittaker biomes except for cold and high-precipitation temperate systems (Fig. 2A and B). The ecological niche of CWPs distributes along two dimensions that jointly describe variations in climate and soil, which delineate four markedly distinct ecoregions (Fig. 2C, Figs S3 and S4). The oligotrophic region of coastal xeric shrublands [encompassing wild cotton populations from the Yucatán Peninsula, Mexico [54]] separates from the remaining more fertile ecoregions, since it is characterized by the warmest temperatures, the lowest micronutrient concentrations, and the sandiest and saltiest soils (Tables S3 and S4). In these stressful environments, plant species most likely require physiological configurations with lower specific leaf area and higher tissue densities [55, 56]. On the opposite dimension, the ordination shows a gradient related to water and energy limitation, distinguishing amongst desert, temperate and tropical ecoregions (Fig. 2C and D). The desert ecoregion included arid and semi-arid grasslands with mild temperatures and alkaline soils, where the low rainfall constraint primary productivity and contributes to higher accumulation of total soil micronutrients [encompassed the wild barley and wheat from the Fertile Crescent (Israel and Iran [57]), and some populations of sunflower from North America (USA [58])]. The tropical seasonal forest and savanna ecoregion included warm and high-moisture environments, such as forests and wetlands, with mature acidic soils due to the intense weathering, and accumulation of SOC and inorganic nitrogen [included wild progenitors of little millet from the Indomalaya region (India [59]) and rice from the Yangtze and Yellow River valleys (China [60])]. The temperate dry forests and shrubland ecoregion was climatically intermediate, and presented higher soil concentrations of total phosphorus, but overall low fertility [agglutinated populations of maize and beans from Mesoamerica (Mexico [61, 62]), potato from the Central Andes (Bolivia [63]), and soya from China [64]]. The favourable climate conditions for plant growth in the tropical ecoregion are likely to yield larger size traits in CWPs and faster nutrient and carbon processing compared to progenitors in ecoregions with water or energy limitations such as deserts [55]. Whilst other factors such as environmental heterogeneity, phenotypic plasticity, intraspecific variability, or growth habits, may also influence changes in CWPs ecological performance [18, 55, 56], our biogeographical assessment of the evolutionary origins of major crops suggests different plant–soil functionalities amongst ecoregions. These ecological insights are useful to inform crop breeding and should be considered in research aimed at improving crops productivity and performance [18, 19].

## The soil core community of crop wild progenitors

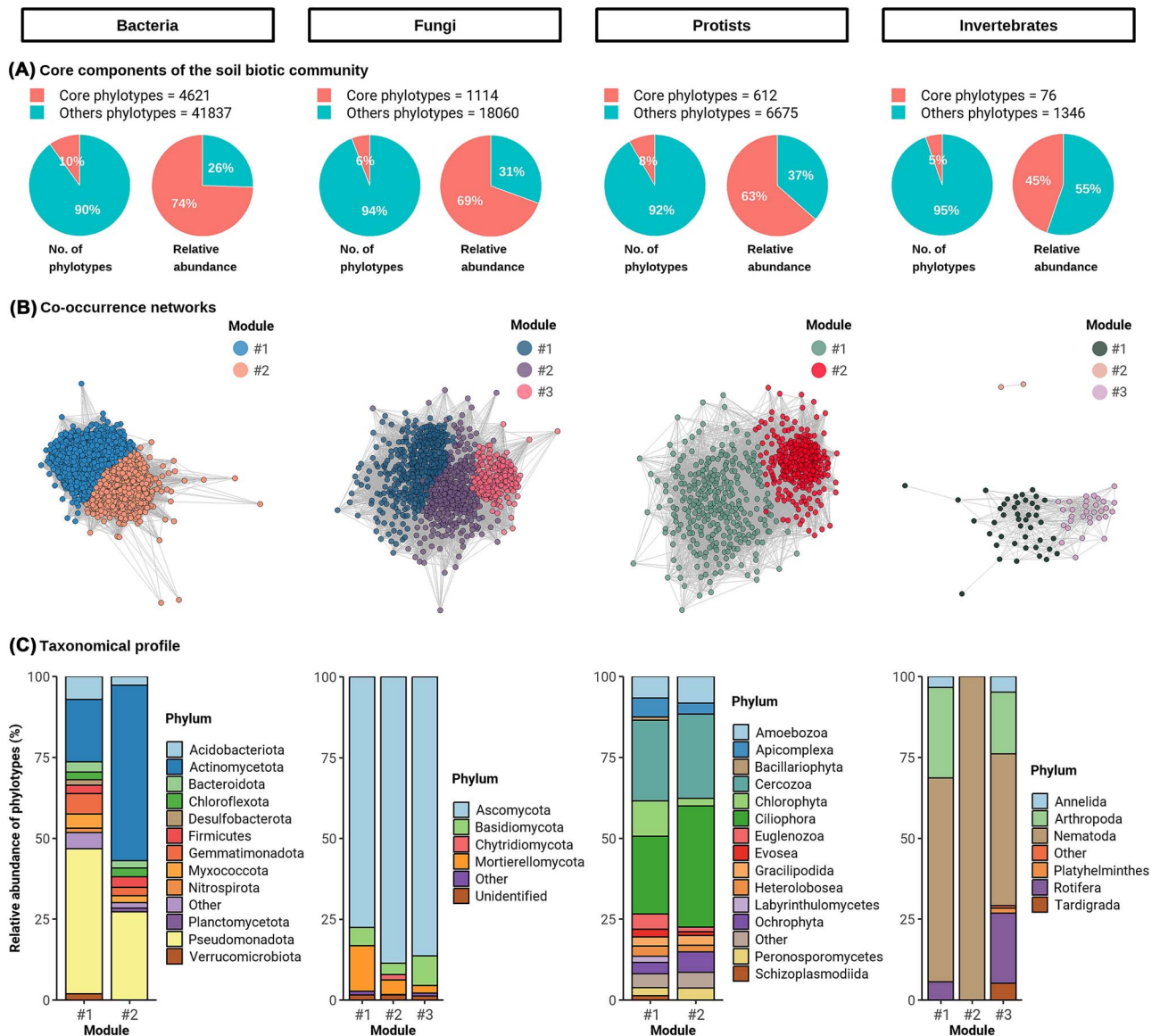
We identified the core phylotypes (i.e., zOTUs) for the four kingdoms comprising the soil biotic community and assessed their organization into modules in co-occurrence networks (Fig. 3B and Table S1). We conducted the analyses on abundant and prevalent taxa within the bulk soil community because these are preferentially filtered by the CWPs in the rhizosphere and contributed the most to increased soil multifunctionality compared to less abundant phylotypes [44]. These core phylotypes accounted for a small portion of the total richness (5–10% of the total number of phylotypes), although dominated in abundance (45%–74% of the total number of reads, Fig. 3A). Moreover, a majority of bacterial core phylotypes were consistently found in soils from all CWPs,



**Figure 2.** Edaphoclimatic characterization of the populations of crop wild progenitors (CWPs). Coloured circles are the surveyed populations of each CWP, and abbreviations in brackets indicate their corresponding modern crop. (A) Climate distribution of wild populations according to mean annual temperature and mean annual precipitation. The frequency distribution of the variables is depicted on the scatterplot axes. Coloured polygons depict the Whittaker's biomes [39]. (B) Ecosystem productivity of wild populations based on NDVI and aridity index. (C) Principal component analysis showing the edaphoclimatic space of ancestral agriculture. The first two principal components (PC1 and PC2) explained 49% of the variance. The arrows indicate the loadings of the edaphoclimatic variables, and the ellipses are a multivariate t-distribution coloured according to clustered ecoregions. (D) Boxplot of the loadings of wild populations within the edaphoclimatic space. Abbreviations of variables: (MAP: Mean annual precipitation, MAT: Mean annual temperature, AI: Aridity index, OC: Organic carbon, CE: Electrical conductivity, NDVI: Normalized difference vegetation index). Asterisks denote median significantly different from zero (Wilcoxon signed-rank test,  $P$ -value  $< .05$ ). Letters indicate significant differences amongst wild progenitors in each principal component ( $P$ -value adjusted with Bonferroni correction to account for multi-group comparisons,  $P$ -value  $< .05$ ).

suggesting they possess generalist functions that support their survival across diverse environments (Fig. S5). Soil core phylotypes of fungi, protists, and invertebrates were associated with smaller sets of CWPs, which may indicate a greater differentiation and dependence on local conditions (Figs S6–S8). Although the centres of origin of CWPs may represent a biogeographically biased

sample of soils worldwide, our results regarding core phylotypes aligns well with findings from global surveys (Fig. 3C), where Pseudomonadota, Actinomycetota, Ascomycota, Cercozoa, Ciliophora, and Nematoda are dominant phyla [11, 65–68]. Moreover, the soil core community included dominant genera of ecological and agricultural relevance, such as bacteria involved in phytohormones

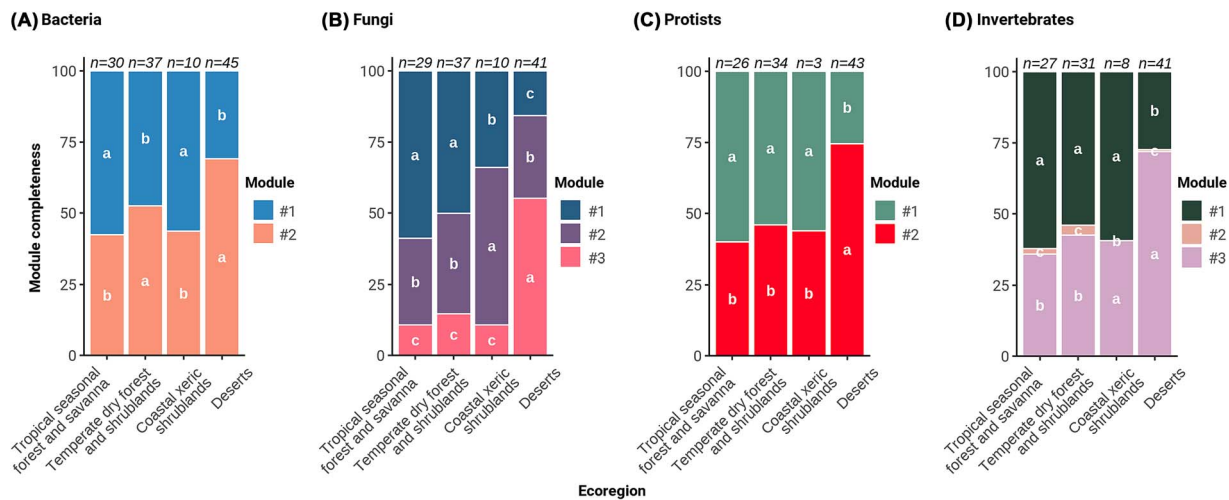


**Figure 3.** Core members of the soil biotic communities of ancestral agriculture (10% most abundant phylotypes present in soil samples from at least four crop wild progenitors). Plots in columns show data for bacteria, fungi, protists, and invertebrates. (A) Pie charts show the proportion of core phylotypes and their relative abundance compared to the remaining phylotypes. (B) Co-occurrence networks show significant relationships (edges with significance level  $P < .05$ ) between core phylotypes (nodes), which are coloured by main modules (highly connected groups of nodes). (C) Barplots show the relative abundance of phylotypes in each module, based on the number of reads of core phylotypes at the phylum level. Relative abundances  $<1\%$  were classified as “Other”.

regulation (e.g. *Sphingomonas* and *Arthrobacter*), plant pathogenic fungi and oomycetes (e.g. *Alternaria*, *Fusarium*, *Pythium*), or plant feeding and parasitic nematodes (e.g. *Tylenchida*) (Figs S9–S12) [17, 35]. Our results indicate that these taxa, which likely compromise the health and productivity of CWPs, were unevenly distributed across the crop’s centres of origin.

We identified the biogeographical patterns in soil biotic community by examining the proportion of nodes belonging to a given co-occurrence module across the edaphoclimatic ecoregions (i.e., module completeness, Fig. 4). Our results showed the existence of core phylotypes associated with deserts, since accounted for 55%–72% of total phylotypes in the ecoregion, and a second major module in tropical areas accounting for 34%–62%. These patterns may reflect interactions between soil taxa and plants, or the convergence of environmental preferences aligned with the edaphoclimatic gradient characterized by water and energy

limitation [68, 69] (Fig. 2). In the desert ecoregion, the soil core community presented higher proportions of saprophytic fungi (Fig. S13), predatory and decomposer soil protists such as Amoebozoa and Ciliophora, as well as soil invertebrates acting as detritivores and predators like Annelida, Rotifera, and Tardigrada (Figs 3C and 4). Interestingly, the wild progenitors of wheat and barley stand out for being associated with specific core phylotypes and shaping a distinct soil community from that of the other CWPs, suggesting a higher degree of adaptation to the arid sites (Figs S5–S8 and S14). In these environments, the scarce availability of key soil resources such as organic carbon and nitrogen might convert CWP rhizosphere into fertility islands, where plant-derived organic inputs concentrate, leading to increased colonization by saprophytic life forms compared to bare soil areas [70]. Knowing the patterns of the soil core community in native regions is essential, as it has been shown to enhance



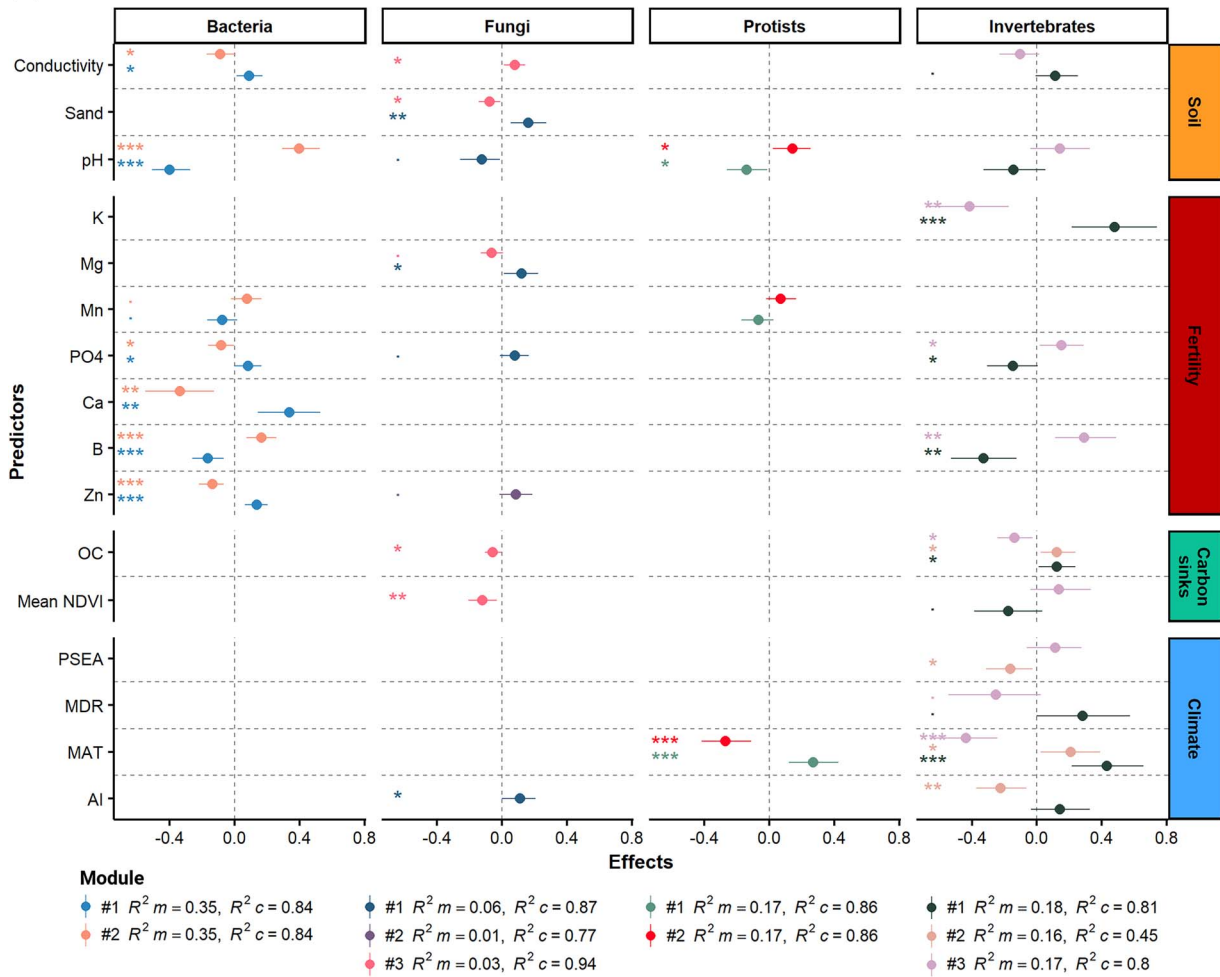
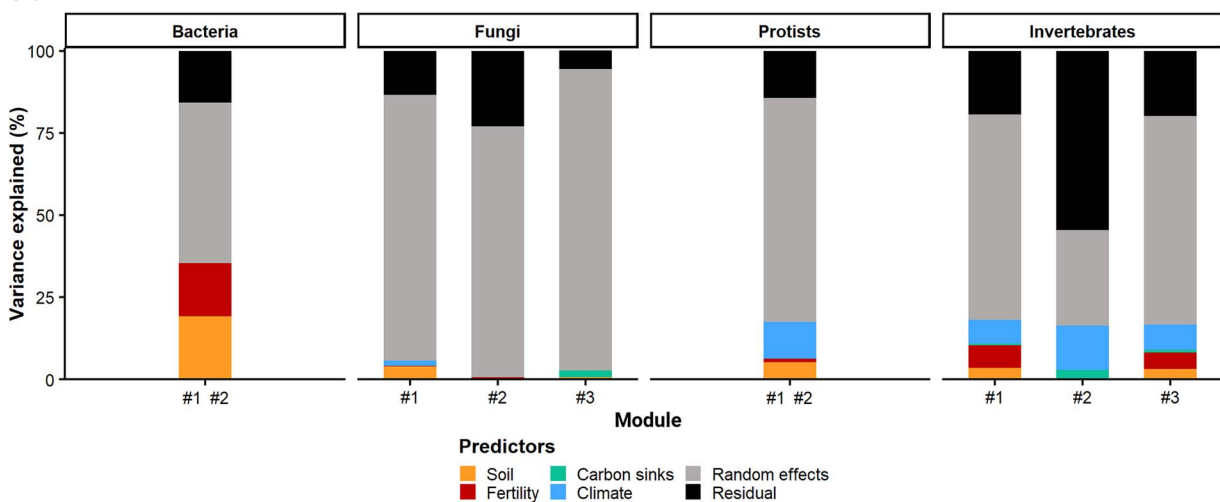
**Figure 4.** Module completeness of soil core members across the edaphoclimatic ecoregions. Colours indicate the averaged proportion of nodes belonging to a given module (i.e., module completeness) of (A) bacterial, (B) fungal, (C) protist, and (D) invertebrate co-occurrence networks.  $n$  = number of wild populations surveyed. Letters indicate significant differences amongst modules within each ecoregion and kingdom ( $\alpha = 0.05$ ).

both plant growth and health, as well as the effectiveness of synthetic communities [71, 72]. Phylotypes in the desert ecoregion where CWP thrive (Figs 3B and C and 4), may help to address one of the most important challenges in dryland agriculture, namely climate change and land degradation [73]. Further research should explore the functional properties of these soil core phylotypes and how they influence plant evolutionary responses that ameliorate plant drought stress and enhance nutrient availability in the rhizosphere, thereby promoting plant growth [24]. In contrast, the soil core community in tropical and temperate ecoregions exhibited greater abundances of Acidobacteriota, fungal parasites, invertebrate parasitic protists like Labyrinthulomycetes, Apicomplexa, and kinetoplastids from the Euglenozoa phylum, as well as invertebrates such as Nematodes and Arthropoda (Figs 3C and 4 and Figs S9–S13). Our results showing increased abundances of parasites in soil communities from tropical compared to dryland ecoregions, can be attributed to several factors such as the higher temperature and moisture levels, a wider variety of host species, and the longer transmission seasons (Table S3) [68]. The wild progenitors of little millet and rice harbour specific core phylotypes that are predictably characteristic of this humid environment, with more anoxic and reducing conditions (e.g. optimal for Desulfobacterota). Wild cotton, originating from marine environments, also shapes a singular core community of soil protists (Fig. 1 and Figs S9–S14). In contrast, the wild progenitors of bean and soya, which are found in regions with intermediate conditions, function as a hub for the soil biotic communities, sharing core phylotypes with all CWPs (Figs S5–S8, S13, and S14). This evaluation of the soil core community of CWPs suggests differences in potential agricultural applications. The ecological module from the desert ecoregions is likely important for stimulating nutrient mineralization and enhancing plant nutrient availability, whereas those from tropical ecoregions may help control the pathogenic organisms contributing to plant health.

### Edaphoclimatic factors shaping modules of soil core biodiversity members

We investigated the factors that determine variations in the abundances of the ecological modules within the soil core community of CWPs (Fig. 5). Soil pH was the main driver of the bacterial modules, along with soil salinity and several micronutrients.

This suggests that certain bacterial taxa abundant in the tropical and coastal shrubland ecoregions are selected for tolerance to acidic soils, as is the case of Acidobacteriota, and Pseudomonadota of the Xanthomonadaceae and Burkholderiaceae families, and salinity as Gammaproteobacteria of the family Alteromonadaceae and Bacteroidota of the families Flavobacteriaceae and Rhodothermaceae [67, 74]. Fungal modules were mainly influenced by sand content, zinc concentration, primary productivity, and aridity. Our results confirm that soil-dwelling fungi are significantly affected by vegetation and climate factors [65, 66]. These different environmental preferences amongst fungal functional groups [65]. As we said, there was less diversity of fungal parasites, but higher of saprophytic fungi, in areas with higher aridity and lower productivity (Fig. S13). Soil protists are distributed as a function of mean annual temperature and, to a lesser extent, soil pH. This indicates differences in the dominant mode of energy acquisition by soil protist communities [68], with increasing temperatures and lower pH favouring greater proportions of parasites as is the case of Labyrinthulomycetes, Apicomplexa, and kinetoplastids, in detriment of other free living heterotrophs such as Amoebozoa and Cercozoa. Soil invertebrates were influenced by a higher number of environmental predictors, likely due to complex interactions amongst trophic levels [69]. Our results are consistent with those of Bastida et al. [11] indicating climate, including its seasonal and daily variability, and vegetation productivity as main drivers of soil invertebrate communities. Taken together, we found that the sets of environmental factors shaping the biological soil community of CWPs varied amongst kingdoms, highlighting the complexity of community-level assembly in native soil environments, and pointing to pH as a common predictor to all of them. Our results are meaningful to guide soil microbial inoculation since we decipher the environmental conditions that can promote particular modules of the ancestral core microbiome [16]. This knowledge will allow the use of specific microbial inoculants adapted to local conditions ensuring their environmental persistence [16, 17]. The current political and economic context seems adequate to harness the potential of these soil core community of ancestral agriculture, as the industry of agricultural microbial products is increasing at a global annual rate of 17% [3].

**(A) Linear mixed effects models****(B) Variance partitioning**

**Figure 5.** Influence of the environment on module assemblies of soil biotic communities of ancestral agriculture. (A) Environmental predictors for each module and microbial group (i.e., bacteria, fungi, protists, and invertebrates). Coloured point ranges indicate the main fixed effects ( $\pm 95\%$  confidence intervals). Significance levels: (\*\*\*)  $p < 0.001$ , (\*\*)  $p < 0.01$ , (\*)  $p < 0.05$ , (.)  $p < 0.1$ .  $R^2 m$  and  $R^2 c$ —marginal and conditional  $r$ -squared in linear mixed models, respectively. (B) Variance partitioning by groups of environmental predictors. Total variance ( $R^2$ ) partitioned into residual variance, random effects variance (crop wild progenitor identity), and the fraction attributable to unique and shared variance by each group of potential predictors (soil properties: pH, conductivity, sand; soil fertility:  $\text{NH}_4^+$ ,  $\text{NO}_3^-$ ,  $\text{PO}_4^-$ , Ca, B, K, Mg, Zn, Fe, Mn, Mo; potential carbon sinks: Mean NDVI—normalized difference vegetation index, SOC—soil organic carbon; climate: MAT—mean annual temperature, MDR—mean diurnal range temperature, AI—aridity index). A single stacked bar of explained variance is shown when there are only two microbial modules in the co-occurrence network because their relative abundances are complementary.

Interestingly, the identity of CWPs was the major source of variation amongst ecological modules of the soil core community (i.e., random effects in the models in Fig. 5). This result may suggest that interactions between plant hosts and the soil microbiome shape module abundances at the local scale [75] (Fig. S14), where the active selection of distinct soil microbiomes takes place [76]. These selection processes could potentially be triggered by the plant phenotypes, or these host phenotypes could be the consequence of plant-independent ecological interactions within the native soil community that do not occur in agricultural systems [8]. Hence, distinct wild populations within ecoregions are potential reservoirs of the co-evolutionary relationships between CWPs and soil organisms [2, 8].

## Conclusions

We found that the populations of CWPs were distributed across widely diverse edaphoclimatic conditions that clustered into four ecoregions. The coastal xeric shrubland ecoregion showed saline sandy soils with low fertility, whilst desert, tropical seasonal forest, and savanna ecoregions featured more fertile soils, although they differed significantly in aridity, soil pH, carbon storage potential, and nutrients concentrations. The ecological modules of the soil core community exhibited distinct taxonomical profiles and abundance patterns across ecoregions, indicating shifts in the life-history strategies and functional traits of soil members. For instance, our analyses indicated that soils from tropical regions harboured greater proportions of acidophilic bacteria, as well as fungal and protist parasites, whereas deserts harboured higher abundances of saprophytic fungi and heterotrophic protists. Our results also suggest that CWPs host unique and diverse soil microbiomes, highlighting the wild populations as fundamental conservation units for the preservation of eco-evolutionary relationships between CWPs and their associated soil biota. This multi-kingdom characterization of soil biodiversity at the geographical regions where agriculture originated can serve as a foundation for targeted microbiome rewilding, and can further inform integrated soil conservation strategies, from plant genetic and soil diversities to their co-evolutionary interactions.

## Acknowledgements

Special thanks to José María Iriondo Alegría for his advice on methods, concepts, and sources of information on crop wild populations.

## Author contributions

María José Fernández-Alonso assisted by Miguel de Celis (Investigation, Data curation, Visualization, and Writing—original draft). Pablo García-Palacios, Ruben Milla and Manuel Delgado-Baquerizo with inputs from Brajesh K. Singh (Conceptualization, Methodology, Validation, Resources, Supervision, Project administration, Funding acquisition and Writing—review & editing). Javier Palomino, Carlos García, Luis Abdala-Roberts, Fernando D. Alfaro, Diego F. Angulo-Pérez, Manoj-Kumar Arthikala, Danteswari Chalasani, Jason Corwin, Duan Gui-Lan, Antonio Hernandez-Lopez, Kalpana Nanjareddy, Siddaiah Chandra Nayaka, Babak Pasari, Thanuku Samuel Sampath Kumar Patro, Appa Rao Podile, Teresa Quijano-Medina, Daniela S. Rivera, Pullabhotla Venkata Subba Rama Narshima Sarma, Salar Shaaf, Pankaj Trivedi, Qingwen Yang, Yue Yin, Eli Zaady, Yong-Guan Zhu (Investigation, Resources, Writing—review & editing). María José Fernández-Alonso, Miguel de Celis, Juan Gaitán and Jun-Tao Wang with

inputs from Ignacio Belda (Software, Validation, Formal analysis, and Writing—review & editing). María José Fernández-Alonso and Miguel de Celis contributed equally to this work.

## Supplementary material

Supplementary material is available at ISME Communications online.

## Conflicts of interest

The authors declare no conflicts of interest.

## Funding

This research was supported by a 2019 Leonardo Grant for Researchers and Cultural Creators, BBVA Foundation, granted to P.G.-P. The BBVA Foundation accepts no responsibility for the opinions, statements and contents included in the project and/or the results thereof, which are entirely the responsibility of the authors. The Spanish Ministry of Science and Innovation also supported this research (grant number PID2021-122296NB-100) to R.M. M.J.F.-A. was supported by a “María Zambrano” postdoctoral fellowship (Spanish Ministry of Universities) at the URJC. A.R.P acknowledges the DST, Govt. of India for the award of the JC Bose Fellowship (JCB/2017/000053) and the Ministry of Education, Govt. of India and Institution of Eminence Directorate (IOE), University of Hyderabad (UOH-IOE-RC3-21-065).

## Data availability

Extended data Amplicon sequencing data generated in this study is available through the European Nucleotide Archive (ENA) under the project accession number PRJEB89460. The specific run accession numbers are: ERR14986789-ERR14986913 for 16S rRNA amplicons, ERR14987213-ERR14987337 for ITS region amplicons, and ERR14987347-ERR14987471 for 18S rRNA gene amplicons. The data and scripts used to generate the analyses and figures presented in this study are openly available on Figshare: <https://doi.org/10.6084/m9.figshare.26990866.v1> [77].

## References

1. Milla R. Crop origins and Phylo food: a database and a phylogenetic tree to stimulate comparative analyses on the origins of food crops. *Glob Ecol Biogeogr* 2020;**29**:606–14. <https://doi.org/10.1111/geb.13057>
2. Pérez-Jaramillo JE, Carrión VJ, de Hollander M. et al. The wild side of plant microbiomes. *Microbiome* 2018;**6**:143. <https://doi.org/10.1186/s40168-018-0519-z>
3. Singh BK, Trivedi P, Egidio E. et al. Crop microbiome and sustainable agriculture. *Nat Rev Microbiol* 2020;**18**:601–2. <https://doi.org/10.1038/s41579-020-00446-y>
4. Fagorzi C, Passeri I, Cangioli L. et al. When biodiversity preservation meets biotechnology: the challenge of developing synthetic microbiota for resilient sustainable crop production. *J Sustain Agric Environ* 2023;**2**:5–15. <https://doi.org/10.1002/sae2.12038>
5. Pérez-Jaramillo JE, Mendes R, Raaijmakers JM. Impact of plant domestication on rhizosphere microbiome assembly and functions. *Plant Mol Biol* 2016;**90**:635–44. <https://doi.org/10.1007/s11103-015-0337-7>

6. Brisson VL, Schmidt JE, Northen TR. et al. Impacts of maize domestication and breeding on rhizosphere microbial community recruitment from a nutrient depleted agricultural soil. *Sci Rep* 2019;**9**:15611. <https://doi.org/10.1038/s41598-019-52148-y>
7. Nerva L, Sandrini M, Moffa L. et al. Breeding toward improved ecological plant–microbiome interactions. *Trends Plant Sci* 2022;**27**:1134–43. <https://doi.org/10.1016/j.tplants.2022.06.004>
8. Raaijmakers JM, Kiers ET. Rewilding plant microbiomes. *Science* 2022;**378**:599–600. <https://doi.org/10.1126/science.abn6350>
9. Anthony MA, Bender SF, van der Heijden MGA. Enumerating soil biodiversity. *Proc Natl Acad Sci* 2023;**120**:2017. <https://doi.org/10.1073/pnas.2304663120>
10. Guo J, Ling N, Li Y. et al. Seed-borne, endospheric and rhizospheric core microbiota as predictors of plant functional traits across rice cultivars are dominated by deterministic processes. *New Phytol* 2021;**230**:2047–60. <https://doi.org/10.1111/nph.17297>
11. Bastida F, Eldridge DJ, Abades S. et al. Climatic vulnerabilities and ecological preferences of soil invertebrates across biomes. *Mol Ecol* 2020;**29**:752–61. <https://doi.org/10.1111/mec.15299>
12. Escudero-Martinez C, Bulgarelli D. Engineering the crop microbiota through host genetics. *Annu Rev Phytopathol* 2023;**61**:257–77. <https://doi.org/10.1146/annurev-phyto-021621-121447>
13. Alegria Terrazas R, Balbirnie-Cumming K, Morris J. et al. A footprint of plant eco-geographic adaptation on the composition of the barley rhizosphere bacterial microbiota. *Sci Rep* 2020;**10**:12916. <https://doi.org/10.1038/s41598-020-69672-x>
14. Pérez-Jaramillo JE, de Hollander M, Ramírez CA. et al. Deciphering rhizosphere microbiome assembly of wild and modern common bean (*Phaseolus vulgaris*) in native and agricultural soils from Colombia. *Microbiome* 2019;**7**:114. <https://doi.org/10.1186/s40168-019-0727-1>
15. Custer GF, Gans M, van Diepen LTA. et al. Comparative analysis of core microbiome assignments: implications for ecological synthesis. *MSystems* 2023;**8**:e01066-22. <https://doi.org/10.1128/msystems.01066-22>
16. French E, Kaplan I, Iyer-Pascuzzi A. et al. Emerging strategies for precision microbiome management in diverse agroecosystems. *Nat Plants* 2021;**7**:256–67. <https://doi.org/10.1038/s41477-020-00830-9>
17. Hartmann M, Six J. Soil structure and microbiome functions in agroecosystems. *Nat Rev Earth Environ* 2022;**4**:4–18. <https://doi.org/10.1038/s43017-022-00366-w>
18. Delgado-Baquerizo M, Reich PB, García-Palacios P. et al. Biogeographic bases for a shift in crop C : N : P stoichiometries during domestication. *Ecol Lett* 2016;**19**:564–75. <https://doi.org/10.1111/ele.12593>
19. Gómez-Fernández A, Osborne CP, Rees M. et al. Disparities among crop species in the evolution of growth rates: the role of distinct origins and domestication histories. *New Phytol* 2022;**233**:995–1010. <https://doi.org/10.1111/nph.17840>
20. GBIF.org. GBIF Home Page 2023. [www.gbif.org](http://www.gbif.org) (26 January 2023, date last accessed).
21. FAO. Global statistical yearbook. FAOSTAT. *Database* 2023; <https://www.fao.org/faostat/en/#data/QCL> (January 3, 2023, date last accessed).
22. Milla R. Crop origins and Phylo food. *GitHub repository* 2021. [https://rubenmilla.github.io/Crop\\_Origins\\_Phylo/](https://rubenmilla.github.io/Crop_Origins_Phylo/) (30 May 2023, date last accessed).
23. Ramirez KS, Snoek LB, Koorem K. et al. Range-expansion effects on the belowground plant microbiome. *Nat Ecol Evol* 2019;**3**:604–11. <https://doi.org/10.1038/s41559-019-0828-z>
24. Trivedi P, Leach JE, Tringe SG. et al. Plant–microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol* 2020;**18**:607–21. <https://doi.org/10.1038/s41579-020-0412-1>
25. Gee GW, Bauder JW. Particle-size analysis. In: Arnold K. (ed.), *Methods of Soil Analysis: Part 1 Physical and Mineralogical Methods*. 677 South Segoe Road, Madison, WI 53711, USA: American Society of Agronomy–Soil Science Society of America; 2018, 383–411. <https://doi.org/10.2136/sssabookser5.1.2ed.c15>
26. Shamrikova EV, Kondratenok BM, Tumanova EA. et al. Transferability between soil organic matter measurement methods for database harmonization. *Geoderma* 2022;**412**:115547. <https://doi.org/10.1016/j.geoderma.2021.115547>
27. Shi S, Kumar S, Young S. et al. Evaluation of 16S rRNA gene primer pairs for bacterial community profiling in an across soil and ryegrass plant study. *J Sustain Agric Environ* 2023;**2**:500–12. <https://doi.org/10.1002/sae2.12075>
28. Wang F, Men X, Zhang G. et al. Assessment of 16S rRNA gene primers for studying bacterial community structure and function of aging flue-cured tobaccos. *AMB Express* 2018;**8**:182. <https://doi.org/10.1186/s13568-018-0713-1>
29. Stoeck T, Bass D, Nebel M. et al. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* 2010;**19**:21–31. <https://doi.org/10.1111/j.1365-294X.2009.04480.x>
30. Ihrmark K, Bödeker ITM, Cruz-Martinez K. et al. New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol Ecol* 2012;**82**:666–77. <https://doi.org/10.1111/j.1574-6941.2012.01437.x>
31. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;**26**:2460–1. <https://doi.org/10.1093/bioinformatics/btq461>
32. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 2011;**27**:2957–63. <https://doi.org/10.1093/bioinformatics/btr507>
33. Edgar RC. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv* 2016;081257. <https://doi.org/10.1101/081257v1>
34. Bolyen E, Rideout JR, Dillon MR. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;**37**:852–7. <https://doi.org/10.1038/s41587-019-0209-9>
35. Pölme S, Abarenkov K, Henrik Nilsson R. et al. FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. *Fungal Divers* 2020;**105**:1–16. <https://doi.org/10.1007/s13225-020-00466-2>
36. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* 2017;**37**:4302–15. <https://doi.org/10.1002/joc.5086>
37. Trabucco A, Zomer R. Global aridity index and potential evapotranspiration (ET0) climate database v2. *figshare Dataset*. 2019. <https://doi.org/10.6084/m9.figshare.7504448.v2>
38. van BS, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011;**45**:1–67. <https://doi.org/10.18637/jss.v045.i03>
39. Stefan V, Levin S. Plotbiomes: plot Whittaker biomes with ggplot2. R package version 0.0.0.9001. 2022.
40. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw* 2008;**25**:1–18. <https://doi.org/10.18637/jss.v025.i01>
41. Oksanen J, Simpson G, Blanchet F. et al. Vegan: community ecology package. R package version 2.6-6.1. 2024. <https://CRAN.R-project.org/package=vegan>

42. Guseva K, Darcy S, Simon E. et al. From diversity to complexity: microbial networks in soils. *Soil Biol Biochem* 2022;**169**:108604. <https://doi.org/10.1016/j.soilbio.2022.108604>
43. Delgado-Baquerizo M. Simplifying the complexity of the soil microbiome to guide the development of next-generation SynComs. *J Sustain Agric Environ* 2022;**1**:9–15. <https://doi.org/10.1002/sae2.12012>
44. de Celis M, Fernández-Alonso MJ, Belda I. et al. The abundant fraction of soil microbiomes regulates the rhizosphere function in crop wild progenitors. *Ecol Lett* 2024;**27**:e14462. <https://doi.org/10.1111/ele.14462>
45. Gu Z. Complex heatmap visualization. *IMeta* 2022;**1**. <https://doi.org/10.1002/imt2.43>
46. de Celis M, Ruiz J, Vicente J. et al. Expectable diversity patterns in wine yeast communities. *FEMS Yeast Res* 2022;**22**:1–5. <https://doi.org/10.1093/femsyr/foac034>
47. Veech JA. A probabilistic model for analysing species co-occurrence. *Glob Ecol Biogeogr* 2013;**22**:252–60. <https://doi.org/10.1111/j.1466-8238.2012.00789.x>
48. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal* 2006.
49. Petrie A. Regclass: tools for an introductory class in regression and modeling. R package version 1.6. 2020. <https://CRAN.R-project.org/package=regclass>
50. Venables WN, Ripley BD. *Modern Applied Statistics with S. Fourth*. New York: Springer-Verlag, 2002.
51. Bates D, Mächler M, Bolker B. et al. Fitting linear mixed-effects models using lme4. *J Stat Softw* 2015;**67**:1–48. <https://doi.org/10.18637/jss.v067.i01>
52. Long JA. Jtools: analysis and presentation of social scientific data. R package version 2.2.0, 2022. <https://cran.r-project.org/package=jtools>
53. Lai J, Zou Y, Zhang S. et al. Glmm.Hp: an R package for computing individual effect of predictors in generalized linear mixed models. *J Plant Ecol* 2022;**15**:1302–7. <https://doi.org/10.1093/jpe/rtac096>
54. Vega M, Quintero-Corrales C, Mastretta-Yanes A. et al. Multiple domestication events explain the origin of *Gossypium hirsutum* landraces in Mexico. *Ecol Evol* 2023;**13**:1–19. <https://doi.org/10.1002/ece3.9838>
55. Joswig JS, Wirth C, Schuman MC. et al. Climatic and soil factors explain the two-dimensional spectrum of global plant trait variation. *Nat Ecol Evol* 2021;**6**:36–50. <https://doi.org/10.1038/s41559-021-01616-8>
56. Gómez-Fernández A, Aranda I, Milla R. Early human selection of crops' wild progenitors explains the acquisitive physiology of modern cultivars. *Nat Plants* 2024;**10**:25–36. <https://doi.org/10.1038/s41477-023-01588-6>
57. Brown TA, Jones MK, Powell W. et al. The complex origins of domesticated crops in the fertile crescent. *Trends Ecol Evol* 2009;**24**:103–9. <https://doi.org/10.1016/j.tree.2008.09.008>
58. Marek LF. Crop wild relatives of sunflower in North America. In: GREENE StephanieL, Williams KA, Khoury CK, Kantar MB, Marek LF (eds), *North American Crop Wild Relatives, Vol. 2*. Switzerland: Springer Nature, 2019, 453–84. [https://doi.org/10.1007/978-3-319-97121-6\\_10](https://doi.org/10.1007/978-3-319-97121-6_10)
59. Goron TL, Raizada MN. Genetic diversity and genomic resources available for the small millet crops to accelerate a new green revolution. *Front Plant Sci* 2015;**6**. <https://doi.org/10.3389/fpls.2015.00157>.
60. Gross BL, Zhao Z. Archaeological and genetic insights into the origins of domesticated rice. *Proc Natl Acad Sci* 2014;**111**:6190–7. <https://doi.org/10.1073/pnas.1308942110>
61. Matsuoka Y, Vigouroux Y, Goodman MM. et al. A single domestication for maize shown by multilocus microsatellite genotyping. *Proc Natl Acad Sci* 2002;**99**:6080–4. <https://doi.org/10.1073/pnas.052125199>
62. Kwak M, Kami JA, Gepts P. The putative mesoamerican domestication center of phaseolus vulgaris is located in the Lerma–Santiago Basin of Mexico. *Crop Sci* 2009;**49**:554–63. <https://doi.org/10.2135/cropsci2008.07.0421>
63. Spooner DM, McLean K, Ramsay G. et al. A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proc Natl Acad Sci USA* 2005;**102**:14694–9. <https://doi.org/10.1073/pnas.0507400102>
64. Zhou Z, Jiang Y, Wang Z. et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol* 2015;**33**:408–14. <https://doi.org/10.1038/nbt.3096>
65. Egidi E, Delgado-Baquerizo M, Plett JM. et al. A few Ascomycota taxa dominate soil fungal communities worldwide. *Nat Commun* 2019;**10**:2369. <https://doi.org/10.1038/s41467-019-10373-z>
66. Tedersoo L, Bahram M, Pöhlme S. et al. Global diversity and geography of soil fungi. *Science* 1979;**204**:346. <https://doi.org/10.1126/science.1256688>
67. Delgado-Baquerizo M, Oliverio AM, Brewer TE. et al. A global atlas of the dominant bacteria found in soil. *Science* 2018;**359**:320–5. <https://doi.org/10.1126/science.aap9516>
68. Oliverio AM, Geisen S, Delgado-Baquerizo M. et al. The global-scale distributions of soil protists and their contributions to belowground systems. *Sci Adv* 2020;**6**:1–11. <https://doi.org/10.1126/sciadv.aax8787>
69. Potapov A, Lindo Z, Buchkowski R. et al. Multiple dimensions of soil food-web research: history and prospects. *Eur J Soil Biol* 2023;**117**:103494. <https://doi.org/10.1016/j.ejsobi.2023.103494>
70. Berdugo M, Delgado-Baquerizo M, Soliveres S. et al. Global ecosystem thresholds driven by aridity downloaded from. *Science* 2020;**367**:787–90.
71. Zhou Y, Liu D, Li F. et al. Superiority of native soil core microbiomes in supporting plant growth. *Nat Commun* 2024;**15**:6599. <https://doi.org/10.1038/s41467-024-50685-3>
72. Wei Z, Gu Y, Friman V-P. et al. Initial soil microbiome composition and functioning predetermine future plant health. *Sci Adv* 2019;**5**:1–11. <https://doi.org/10.1126/sciadv.aaw0759>
73. Cherlet M, Hutchinson C, Reynolds J. et al. Editors. *World Atlas of Desertification*, 3rd edn. Luxembourg: Publication Office of the European Union, 2018.
74. Rath KM, Fierer N, Murphy DV. et al. Linking bacterial community composition to soil salinity along environmental gradients. *ISME J* 2019;**13**:836–46. <https://doi.org/10.1038/s41396-018-0313-8>
75. Miller ET, Svanbäck R, Bohannan BJM. Microbiomes as meta-communities: understanding host-associated microbes through metacommunity ecology. *Trends Ecol Evol* 2018;**33**:926–35. <https://doi.org/10.1016/j.tree.2018.09.002>
76. Edwards J, Johnson C, Santos-Medellín C. et al. Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc Natl Acad Sci* 2015;**112**:E911–20. <https://doi.org/10.1073/pnas.1414592112>
77. Microwild consortia. Biogeographic manuscript. *figshare*. Preprint. 2025. <https://doi.org/10.6084/m9.figshare.26990866.v1>