Genetic bases of plant-microorganisms interaction in rice

Lorenzo Sena

PhD in Agri-Food Sciences, Technologies and Biotechnologies

UniMORE (XXXVII Cycle)

Tutor: Prof. Nicola Pecchioni

Co-Tutor: Dr. Patrizia Vaccino

PhD Coordinator: Prof. Fabio Licciardello

Contents

List of Figures

List of Tables

Abstract

Rice stands as one of the primary sources of sustenance for humanity, ranking among the top three most cultivated cereals worldwide. It serves as a staple food for numerous communities, particularly in developing countries. However, rice cultivation faces several challenges, including the need to consistently increase yields while combating diseases that afflict the crop. Simultaneously, there is a growing emphasis on reducing the reliance on pesticides and other artificial substances in agricultural practices.

Rice biodiversity, cultivated and wild, represents a potentially huge reservoir of useful genetic variation for many traits, including the ability to be colonized by beneficial microorganisms. Therefore, the phenotypic and genomic knowledge of a large number of accessions gives the opportunity to identify genomic *loci* on which a selection pressure could be applied, to improve modern high yielding varieties by breeding programs.

The study of interactions between plants and naturally occurring soil microorganisms is gaining increasing attention due to its potential significance in advancing sustainable agriculture. Numerous studies have already been published on the interactions and effects of certain Plant Growth-Promoting Microorganisms (PGPMs), although much about these interactions remains poorly understood.

This study aims to delve deeper into the genetic foundations potentially involved in the interaction between PGPMs and rice plants. Specifically, a diazotrophic bacterium, *Kosakonia sacchari* RCA25, was used to inoculate approximately 200 genotypes of rice (*Oryza sativa* L. subsp. *Japonica*), whose genotyping by GBS (Genotyping By Sequencing) had been previously carried out. Genome-Wide Association Studies (GWAS) were conducted to identify genomic regions implicated in the plant-microorganism interaction.

The first phase of the experiment focused on the controlled inoculation of *K. sacchari* RCA25 into the rice plants, with 3 replicates for each genotype. Sterilized de-hulled seeds were prepared, and the bacterium was inoculated one week after germination. One week after the inoculation, the bacterial colonization rate was quantified, through CFUs counting, specifically at the root level. The mean colonization values from the replicates were utilized for subsequent statistical and genomic association analyses.

The results showed a high variability in the levels of colonization by the same microorganism across different rice accessions, although no significant associations were found between genotypes and bacterial infection rates. This variability suggests a complex interplay between small-effect genetic factors and variation in the phenotype of microbial colonization, highlighting the need for further investigation to better understand the mechanisms underlying these interactions and their potential applications in rice cultivation.

Introduction

2.1 Rice

Rice (*Oryza sativa* L.) is the third cereal in global production after maize and wheat, with a harvested area of approximately 170 million hectares and a global yield of around 800 million tons in 2023 (FAO, 2023), and has a vital importance for food security and agricultural economies (Ghazy *et al.*, 2024). As a staple food for more than half of the world's population, rice contributes over 20% of the calories consumed by humans worldwide (Gutaker *et al.*, 2020; He *et al.*, 2021). It is particularly crucial for the poorest communities, serving as their primary food source (Zeigler & Barclay, 2008). Beyond its role as a dietary staple, rice is also utilized in various industries (Ghazy *et al.*, 2024).

Rice is a semi-aquatic annual grass plant belonging to the *Poaceae* family. The genus *Oryza* comprises approximately 22 species, 20 of which are wild (Muthayya *et al.*, 2014). Among the cultivated species, *O. sativa*, or Asian rice, is the most widely grown, cultivated on all continents except Antarctica, while *O. glaberrima*, or African rice, is primarily cultivated in Africa and is gradually being replaced by *O. sativa* (Muthayya *et al.*, 2014). Rice is cultivated in diverse ecosystems, including lowland paddy fields, deepwater or seasonal flood conditions, and upland

rainfed areas (Gutaker *et al.*, 2020). Irrigated lowland systems, where rice is grown in bunded fields, account for nearly three-quarters of global rice production and can yield two to three crops per year (Muthayya *et al.*, 2014). Rainfed lowland farming, prevalent in South Asia, Southeast Asia, and Africa, produces 20% of the world's rice, while upland rice farming in dryland conditions contributes 4% (Muthayya *et al.*, 2014).

O. sativa is believed to have been domesticated from its wild ancestor *O. rufipogon* between 10,000 and 14,000 years ago (Zeigler & Barclay, 2008). While *O. rufipogon* is commonly accepted as the progenitor of *O. sativa*, *O. nivara*, often considered the annual form of *O. rufipogon*, has also been proposed as an ancestor (Choi *et al.*, 2017). *O. sativa* is genetically subdivided into major groups: *aus*, *indica*, *japonica* (sometimes further divided into temperate and tropical *japonica*), and aromatic rice (Choi *et al.*, 2017). Among these, *japonica* and *indica* are the most widely distributed globally (Gutaker *et al.*, 2020). *Japonica* rice, characterized by short to medium grains, is primarily grown in temperate regions of Japan and northern China, while *indica* rice, with long grains, is cultivated in tropical and subtropical Asia (Muthayya *et al.*, 2014). The domestication of *japonica* rice began around 9,000 years ago in the lower Yangtze Valley, while proto-*indica* rice cultivation started over 5,000 years ago in the lower Ganges Valley (Gutaker *et al.*, 2020). Recent studies suggest that domestication alleles originated in *japonica* rice in East Asia, with the spread of *japonica* varieties to South Asia around 4,000 years ago, leading to the introgression of these alleles into proto-*indica* or local *O. nivara* populations, resulting in the emergence of *indica* rice (Gutaker *et al.*, 2020).

The domestication of *O. glaberrima*, the other cultivated rice species, is thought to have derived from the wild ancestor *O. barthii* (Muthayya *et al.*, 2014).

Two primary hypotheses have been proposed regarding the domestication of rice. The single origin hypothesis suggests that a single domestication event led to the acquisition of desirable traits, followed by the incorporation of additional characteristics, particularly resistance traits, from wild varieties. In contrast, the multiple origin hypothesis posits that multiple, independent domestication events from different populations of wild rice varieties gave rise to the subgroups *indica* and *japonica* (Choi *et al.*, 2017; He *et al.*, 2021). Genetic studies further support the idea that each Asian rice variety group/subspecies (*aus*, *indica*, and *japonica*) had distinct subpopulations of wild rice (*O. nivara* or *O. rufipogon*) as its progenitor (Choi & Purugganan, 2018). Wild rice can be divided into three major subpopulations, designated as *Or-I*, *Or-II*, and *Or-III*, with *japonica* being most closely related to *Or-III*, and *aus* and *indica* most closely related to *Or-I* (Choi & Purugganan, 2018).

Rice is not only a critical food crop but also a culturally and economically significant plant with a rich history of domestication and diversification. Its cultivation spans a wide range of environments, making it a versatile and resilient crop.

Ex-situ genebanks play a crucial role in conserving genetic diversity for the development of new genomic knowledge, and the deployment of useful variation (Aubry, 2023; Manual of Seed Handling in Genebanks, 2006). To exploit their full potential, the conservation of diversity must be accompanied by the long-term characterization of genebank accessions. In the present PhD project, a series of historical data collected from the rice *ex-situ* collection of CREA in Vercelli has been analyzed as part of a parallel effort to understand the value of the conserved germplasm. The published paper is attached in the appendix (Sansoni *et al.*, 2024).

2.2 Plant Growth-Promoting Microorganisms

Plant Growth-Promoting Microorganisms (PGPMs) play a crucial role in enhancing plant growth and health through various direct and indirect mechanisms. These microorganisms contribute to plant development by fixing atmospheric nitrogen, solubilizing essential nutrient elements, and producing phytohormones (Siddika *et al.*, 2024). Additionally, PGPMs produce molecules like ExoPolySaccharides (EPS), siderophores, and antioxidants, which further support plant growth and stress tolerance (Siddika *et al.*, 2024).

Among the PGPMs, *Enterobacter* and *Kosakonia* are notable for their ability to colonize a wide range of plant hosts, with *Kosakonia* strains naturally colonizing plants across nearly all major clades of angiosperms (Becker *et al.*, 2018). The role of PGPMs extends to nutrient cycling and increasing nutrient availability in the soil, which is essential for plant growth (Becker *et al.*, 2018; Rodriguez & Durán, 2020). For example, diazotrophs, a group of nitrogen-fixing bacteria, can supplement biological nitrogen directly, promote nutrient uptake, and enhance phytohormone synthesis, thereby improving plant tolerance to biotic and abiotic stresses (Saini *et al.*, 2022). The interaction between plants and microorganisms is not limited to the rhizosphere but also includes endophytes, which reside within plant tissues without causing harm (Hardoim *et al.*, 2015). A thorough review of the available literature about crops and microorganisms interactions has been conducted during the present PhD project. In the paper attached, Sena *et al.* (2024) provide a detailed description of the categories of endophytic interacting microorganisms, as well as their potential roles and effects on enhancing plant tolerance to the impacts of climate change.

Materials and methods

3.1 Rice panel

The rice panel used in the experiment was defined by selecting about 200 genotypes from the Risinnova collection, an international collection composed by about 300 *O. sativa* L. spp. *japonica* varieties, previously constituted to obtain genomic insights to counteract biotic and abiotic stresses. The Risinnova collection is represented for about a half by Italian varieties, and for the other part by accessions from all over the world (**Figure 3.1**).

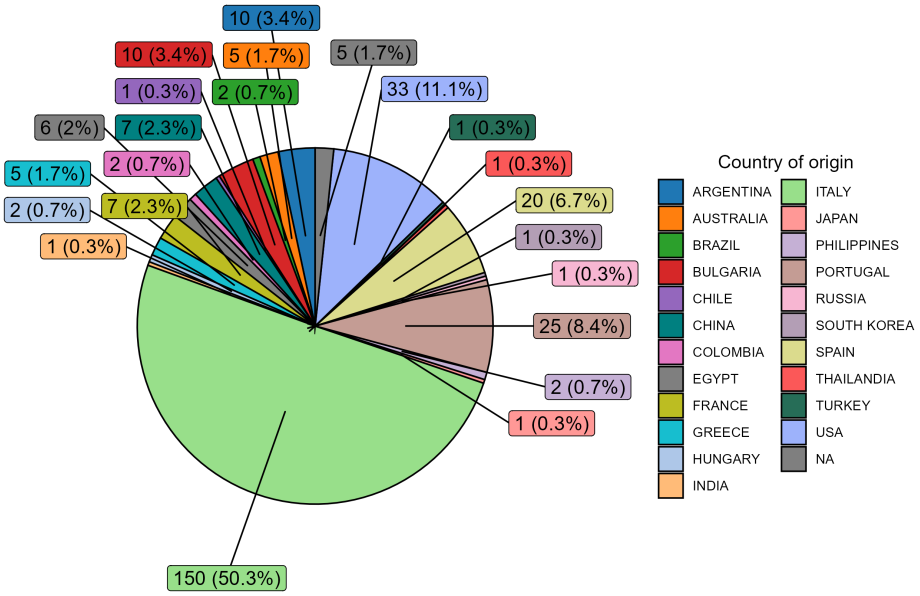


Figure 3.1: Pie chart showing the international representation of the rice varieties composing the Risinnova collection.

The genotypes were selected through a phylogenetic tree, to maximize the genetic variability between them. **Table 3.1** lists the genotypes selected for the association study, after some filtering for contrasting results of the inoculation phase.

Table 3.1: Genotypes selected for the association study after the inoculation results.

N°	GENOTYPE
1	A201
2	AGATA
3	AGOSTANO
4	ALAN
5	ALEXANDROS
6	ALLORIO
7	ALPE
8	ALPHA
9	AMERICANO
10	ANSEATICO
11	ANTARES
12	APOLLO
13	ARGO
14	ARIETE
15	ARTEMIDE
16	ASIA
17	AUGUSTO
18	BAHIA
19	BALDO
20	BALILLA
21	BELLEPATNA
22	BENGAL
23	BIANCA
24	BOMBILLA
25	BRAZOS
26	BURMA
27	CALENDAL
28	CALMOCHI_101

N°	GENOTYPE
29	CAMPINO
30	CARINA
31	CARNAROLI
32	CARNISE
33	CASTELMOCHI
34	CHIPKA
35	CIGALON
36	CLOT
37	COCODRIE
38	COLINA
39	CT36
40	CT58
41	DELFINO
42	DELLMONT
43	DIXIEBELLE
44	DUCATO
45	ERCOLE
46	ERMES
47	ESCARLATE
48	EUROPA
49	EUROSE
50	EUROSIS
51	FAMILIA_181
52	FAST
53	FIDJI
54	FLIPPER
55	FRANCES
56	FRANGRANCE

N°	GENOTYPE
57	GALILEO
58	GANGE
59	GARDESADRI
60	GIADA
61	GIANO
62	GIGANTE_VERCELLI
63	GIOVANNI_MARCHETTI
64	GITANO
65	GIZA_178
66	GLADIO
67	GOOLARAH
68	GRAAL
69	GRALDO
70	GREPPI
71	GRITNA
72	GZ6296
73	HANDAO_11
74	HANDAO_297
75	HARLEM
76	IAC32.52
77	IBO380.33
78	IBO400
79	IR64
80	ITALMOCHI
81	ITALPATNA_X_MILYANG
82	ITALPATNA48
83	JACINTO
84	KARNAK

N°	GENOTYPE
85	KING
86	KRYSTALLINO
87	L201
88	L202
89	L205
90	LADY_WRIGHT
91	LAMONE
92	LENCINO
93	LIDO
94	LOTO
95	LUNA
96	LUXOR
97	M006
98	M203
99	M204
100	MARATELLI
101	MARENY
102	MAYBELLE
103	MEDUSA
104	MERLE
105	MESANES2
106	MIARA
107	MOLO
108	MUSA
109	NANO
110	NILO
111	NOVARA
112	ONICE

N°	GENOTYPE
113	ORIGINARIO
114	OSCAR_X_SUWEON
115	OSTIGLIA
116	OTA
117	P6
118	PANDA
119	PEGONIL
120	PIERINA_MARCHETTI
121	PLOVDIV22
122	PLUS
123	PROMETEO_2
124	PUNTAL
125	REDI
126	REXMONT
127	RIBE
128	RINALDO_BERSANI
129	RINGO
130	RIZZOTTO_51.1
131	ROBBIOSEL1
132	RODEO
133	RONALDO
134	RONCAROLO
135	RONCOLO
136	ROXANI
137	RPC12
138	RUBI
139	RUBINO
140	RUSSO1

N°	GENOTYPE
141	S101
142	S102
143	SAEDINENIE
144	SAFARI
145	SAGRES
146	SAKHA_102
147	SAKHA_103
148	SALOIO
149	SALVO
150	SAMBA
151	SANDOCA
152	SANDORA
153	SANTANDREA
154	SANTERNO
155	SATURNO
156	SELENIO
157	SELN244A
158	SENATORE_NOVELLI
159	SEQUIAL
160	SESIAMOCHI
161	SETANTUNO
162	SILLA
163	SIRIO_CL
164	SLAVA
165	SMERALDO
166	SOURE
167	SPRINT
168	STRELLA

N°	GENOTYPE
169	SUPER
170	TAICHUNG_G65
171	TEJO
172	TEXMONT
173	THAIPERLA
174	ULISSE
175	UPLA63
176	UPLA64
177	UPLA66
178	UPLA75
179	UPLA77
180	VENERE
181	VENERIA
182	VIALE
183	VIALONE_190
184	VIALONE_NANO
185	VICTORIA
186	VOLANO
187	VULCANO
188	XIANGOU2
189	YRM6.2
190	ZENA
191	ZENITH
192	ZHENSHANG97

3.2 *Kosakonia sacchari* RCA25

Kosakonia sacchari is a Gram-negative, aerobic, not spore-forming, motile rod bacteria with peritrichous flagella. The name of the species was derived by its association with sugarcane (*Saccharum officinarum* L.) (Chen *et al.*, 2014). *K. sacchari* is a diazotrophic bacterium, able to fix atmospheric nitrogen reducing N_2 to NH_3 in low pO_2 conditions.

The *genus Kosakonia* is part of the family *Enterobacteriaceae*, which also includes well-known pathogens such as *Escherichia* and *Salmonella*. *K. sacchari* itself was reclassified as a member of the *genus Kosakonia*, from the previous classification of *Enterobacter sacchari* (Chen *et al.*, 2014). This reclassification was part of a broader taxonomic revision (Brady *et al.*, 2013; Giri, 2019). The first strain of *K. sacchari*, $SP1^T$, was isolated in 1994 from a sugarcane cultivar (GT11) in Nanning, Guangxi, China (Chen *et al.*, 2014). This bacterium has also been found in other plant-associated environments. For instance, an endophytic strain, BO-1, was isolated from the surface-sterilized stem of a sweet potato (*Ipomoea batatas*) (Shinjo *et al.*, 2016). Strain BO-1 is closely related to *K. sacchari* strain SP1T, with which it shares 98.67% Average Nucleotide Identity (ANI).

Both strains exhibit diazotrophic growth under nitrogen-deficient conditions and show positive reactions in acetylene reduction activity tests, confirming their nitrogen-fixing capabilities (Shinjo *et al.*, 2016). The genome of *K. sacchari* strain BO-1 has been sequenced, revealing 4,448 coding sequences, seven sets of rRNA genes, and 84 tRNA genes. Genome annotation confirmed the presence of the *nif* gene locus, which is responsible for nitrogen fixation (Shinjo *et al.*, 2016). Members of the *genus Kosakonia*, including *K. sacchari*, are known for their beneficial interactions with plants. *K. sacchari*, along with *K. arachidis*, *K. cowanii*, *K. oryzae*,

and *K. radicincitans*, has been isolated from various plant-related environments, such as rhizosphere soil, roots, and stems (Kämpfer *et al.*, 2016).

Kosakonia sacchari RCA25 strain was used during all the experiments. This strain can grow in a LB selective medium added with the antibiotic vancomycin. The RCA25 strain, and the relative inoculation protocol, were provided by the team lead by Prof. Roberto Defez, at CNR-IBBR (National Research Council - Institute of Biosciences and Bioresources) in Naples, Italy. The protocol was previously used in published works by the team, with little differences (Andreozzi *et al.*, 2019; Bianco *et al.*, 2021). This particular strain was chosen for the experiment because it has already been tested as an endophyte in rice plants, and it has also proven to be a rather common endophyte in relation to various plant species. Additionally, an inoculation protocol already tested on rice was available at the beginning of the experimentation, from CNR-IBBR.

3.3 Solutions and recipes used in the experiment

3.3.1 Solutions

Following is the list of solutions used:

- **Glycerol:** an agent used in the preparation of freezing stocks to preserve cells and prevent the formation of ice crystals which could damage the samples. Typically, as in this case, pure glycerol is diluted to a final concentration of 50% v/v with sterile deionized water.
- **70% ethanol**.
- **Sodium hypochlorite (NaClO):** approx. 5% (commercial bleach).

- **Tween-20:** a commercial name for Polysorbate-20, a detergent used (among other things) in the sterilization processes of biological material as it reduces surface tension and enhances the action of sterilizing agents.
- **Water-agar:** a gel solution for seed germination and early growth of seedlings.
- **PBS (Phosphate-Buffered Saline):** an isotonic saline solution, used to maintain the proper osmotic pressure inside the bacterial cells.
- **LB (Lysogeny Broth):** a basic liquid medium for the growth of the inoculum. For the aim of the described experiment, it is made selective by adding the antibiotic Vancomycin.
- **LB-agar:** a solid medium prepared by adding agar to the LB broth for the preparation of Petri dishes for CFU counting. As with the liquid LB described above, for this experiment the medium is made selective by adding Vancomycin.

3.3.2 Recipes

LB and LB-agar

SUBSTANCE	QUANTITY (g/L)
Tryptone	10
Yeast extract	5
NaCl	10

Adjust the pH to 7.4 with NaOH and bring to volume using deionized water. Autoclave. For LB-agar, add 15 g/L of Bacto-agar to the previous solution and autoclave it.

Water-agar

For the water-agar substrate, prepare an aqueous solution with a 0.8% agar concentration. The solution should then be autoclaved for sterilization.

PBS(Phosphate-Buffered Saline)

SUBSTANCE	QUANTITY (g/L)	QUANTITY (g/L) [10X]
NaCl	8	80
KCl	0.2	2
Na_2HPO_4	1.44	14.4
KH_2HPO_4	0.245	2.45

Adjust the pH to 7.4 with H_3PO_4 and bring to volume using deionized water. Autoclave.

Vancomycin 1000X stock

SUBSTANCE	QUANTITY (g)
Vancomycin chlorhydrate	0.02

- Add the appropriate amount of vancomycin powder to a small container or tube.
- Bring to the desired final volume with sterile deionized water.
- Mix the solution vigorously using a vortex mixer to ensure complete dissolution of vancomycin.
- Once the solution is prepared, it must be filtered using a syringe and a sterile 200 nm hydrophilic filter to remove any particulate matter.

- Transfer the filtered solution into 1.5 mL or 2 mL vials.
- Store the vials at -20 °C for long-term storage.

3.4 Inoculation protocol

3.4.1 Seeds sterilization

The inoculation protocol started with the sterilization of de-hulled seeds. 20 cleaned seeds were put inside 15 mL sterile plastic tubes, one for each tested genotype, and then kept in 70% ethanol (EtOH) for 8 minutes while gently stirring on an orbital shaker, to guarantee a proper action of the solution on the seeds surface. The EtOH solution was removed and seeds washed with deionized sterile water, for 2 minutes. After the removal of water, seeds were treated with 10 mL of 5% sodium hypochlorite (NaClO) solution added with 10 µL of Tween-20 and left in agitation for 25 minutes.

The solution was removed, and multiple washes of about 2 minutes each with deionized sterile water were performed. Seeds were considered sufficiently washed when no traces of NaClO smell were detectable.

3.4.2 Germination

Once the washing procedures were completed, the seeds were placed inside Magenta boxes prepared with a layer of about 1 cm of sterilized water-agar solution (0.8% agar), for germination. 5 to 6 seeds per box were put on the solidified water-agar medium, applying a gentle pressure, using flame-sterilized tweezers. For each genotype, 3 Magenta boxes were prepared.

Magenta boxes containing the seeds were left in a growth chamber, in darkness, at approximately 22°C for germination (**Figure 3.2**).

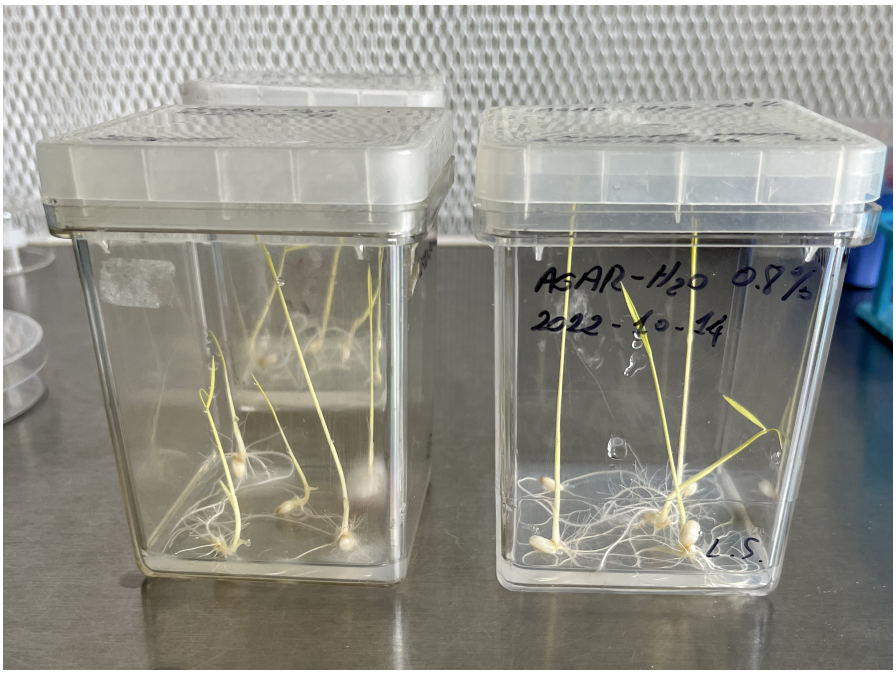


Figure 3.2: Magenta boxes used for the germination of sterilized de-hulled seeds. Boxes were autoclaved before the use, and they were prepared with a layer of water-agar medium for the germination of the seeds. In the photo, seedlings ready for the inoculation phase are shown.

3.4.3 Seedlings inoculation

For seedlings inoculation, 15 mL sterile plastic tubes filled with 10 mL of LB medium (see paragraph: LB and LB-agar, in Recipes) supplemented with 20 µg/mL vancomycin were used. For convenience, aliquots at a 1000-fold higher concentration (*i.e.*, 20 mg/mL) were prepared and stored at -20 °C (see paragraph: Vancomycin 1000X stock, in Recipes). To obtain the desired final concentration of 20 µg/mL, 10 µL of the 1000X vancomycin stock was added to the 10 mL of LB

medium.

The bacterial inoculum was thawed from the freezing stock at -80 °C and transferred to the tube containing the selective LB-vancomycin medium. The inoculum was then incubated at 30 °C for 24 hours in the dark under constant agitation. Once the inoculum reached an optical density at 600 nm (OD_{600}) of approximately 1, 1 mL of the bacterial solution was taken and centrifuged at 10000 RPM (Rounds Per Minute) for 5 minutes at room temperature. The pellet was gently resuspended in 1 mL of 1X PBS (Phosphate-Buffered Saline, see **PBS**, in **Recipes**).

The germinated seedlings were inoculated when the coleoptile and roots started to develop, typically after 5-7 days from the preparation of Magenta boxes. The seedlings were removed from the germination container, and were lightly cut at the root tip, approximately 3-4 mm from the tip, using flame-sterilized precision scissors and tweezers. The cutting aims to facilitate the penetration of the microorganisms inside the plants roots tissues. Plantlets were then placed in a 50 mL tube previously prepared with 30 mL of 1X PBS, 3 plants per tube as a single replicate.

From the tube containing the *K. sacchari* RCA25 strain (centrifuged and resuspended in PBS), 300 µL was taken and added to the 50 mL tube containing the plants. The tube was then placed in the dark with a gentle agitation for 4 hours at 30 °C.

After infection, the seedlings were transferred to labelled pots filled with a solid growth substrate, consisting of a 1:1 ratio of sand and vermiculite. Both the pots and the substrate were previously sterilized. Each group of 3 plants were sown in a single pot. The prepared pots were placed in a growth chamber under lights

with a light-dark ratio of 16/8 hours at approximately 28 °C. The pots for different treatments (*i.e.*, with RCA25 inoculum or negative controls) were kept separate to avoid possible contamination. Plants prepared in this way were left to grow for another week, watering regularly.

3.4.4 Plants collection and sterilization

The plantlets were removed from the pot, cleaned, and washed with tap water to remove any substrate residue. They were then put in 50 mL tubes (one per replicate), and surface sterilized using 30 mL of 70% EtOH for 1 minute, with the tube kept in agitation. After removing the EtOH, sterile deionized water was added in sufficient quantity to perform adequate washing, and the tubes were agitated for 2 minutes. The water was then removed, and a 5% NaClO solution was added. The hypochlorite washing was carried out for 5 minutes in agitation. Following the removal of the sodium hypochlorite solution, multiple washes with water were performed to ensure no traces of NaClO remained. Typically, six washes of 1 minute each were conducted, followed by a final wash of 2 minutes.

Sample checks were carried out on the water used, through plating, to ensure its sterility. Similarly, some tests were conducted to evaluate the effectiveness of the surface sterilization process of the plants before the collection of root tissue.

3.4.5 Roots collection

At the end of the washes, under horizontal laminar flow hood, the plants were taken with flame-sterilized tweezers, and the roots were cut off. The roots were conserved in 1,5 mL sterile tubes and weighted, while the remaining parts of the plant were discarded. All the roots from the 3 plants composing a replicate were conserved and weighted together. A mean value for the 1,5 mL tube weight was

calculated and then subtracted from each test.

The weighted roots were then transferred to a mortar, ensuring all equipment used (mortar and pestle, scissors, and tweezers) was sterilized in the autoclave or with a flame. 5 mL of 1X PBS was added to the mortar, and the roots were ground until well crushed. When the solution resulted homogeneous, an aliquot of 1 mL was put inside a 1,5 mL tube. This represented the 1:1 ratio solution, from which the following serial dilutions were prepared: 1:10, 1:100, 1:1000, and 1:10000.

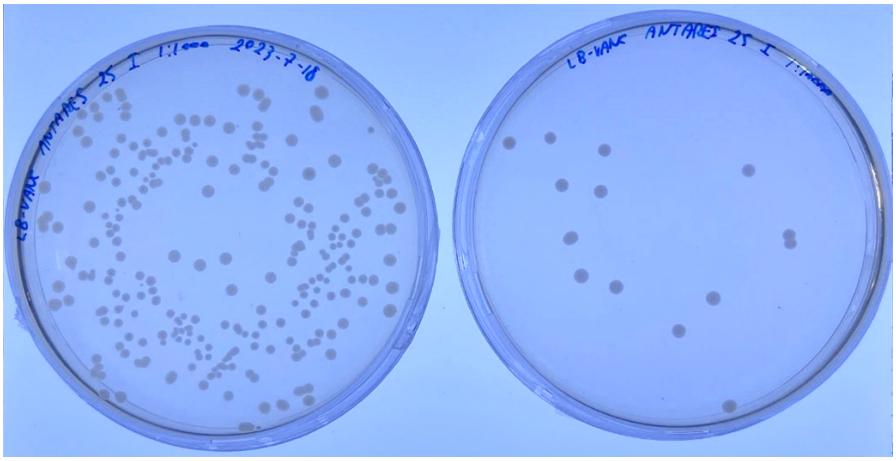
Alternatively, a TissueLyzer could be used to homogenize the root sample. In this case, 3-4 sterilized metal marbles were added to the tube containing the roots, followed by 1 mL of PBS. After a cycle in the TissueLyzer (1 minute at f = 30/s), the resulting solution was transferred to 15 mL tubes pre-filled with 4 mL of PBS.

3.4.6 Serial dilutions and plating

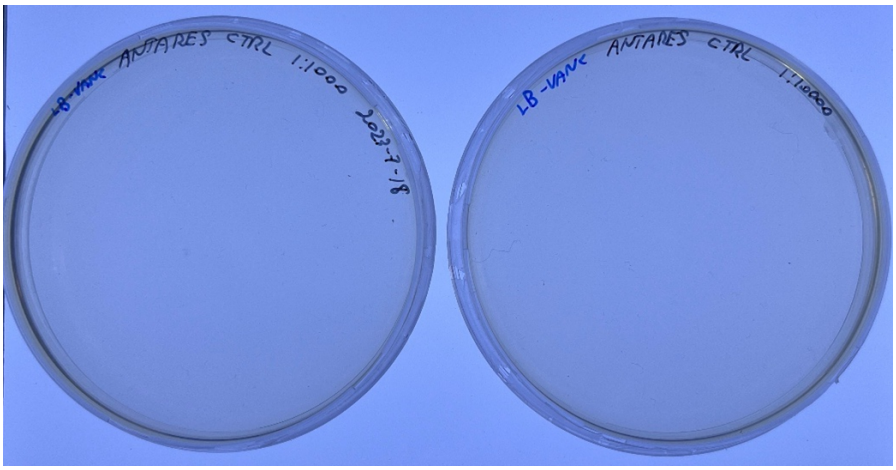
Serial dilutions were prepared using ultrapure sterilized water (MilliQ system). The prepared dilutions were 1:1, 1:10, 1:100, 1:1000, and 1:10000, with a final volume of 1 mL for each. 50 µL of the 1:1 solution was added to a previously prepared Petri dish containing a layer of LB-vancomycin medium and spread evenly to ensure optimal distribution over the entire surface of the agar. Metal inoculation loops, flame-sterilized, were used to perform the plating. The Petri dish was then closed, properly labelled with the necessary information, and set aside. This procedure was repeated for each dilution, and for the negative controls.

For each genotype, 3 replicates were prepared, and for each replicate 5 plates were done, one for each dilution, plus one for the negative control. Once all the plates were prepared, they were sealed with Parafilm and placed in a growth

chamber at 30 °C. After approximately 24 hours, the colony growth on the plates for the respective dilutions was checked (**Figure 3.3**).



(a)



(b)

Figure 3.3: Example of Petri plates prepared for the counting of the CFUs. In (**a**), the plates of the first replicate for the genotype "Antares", dilutions 1:1000 and 1:10000, are visible. In (**b**) the plates for the dilutions of the negative control. Plates are labeled with the medium (LB-vancomycin), the genotype, the strain (25, standing for RCA25), the number of the replicate, the dilution factor, and the date of the test.

3.4.7 CFUs counting

After 24 hours, plates were analysed and CFUs counted. For each test, one of the four dilutions (1:10, 1:100, 1:1000, 1:10000) was chosen, based on the number of CFUs present: the plates with a number of CFUs between 30 and 300 were selected for the definition of the colonization rates.

The main passages of the protocol described are represented in the scheme in **Figure 3.4**.

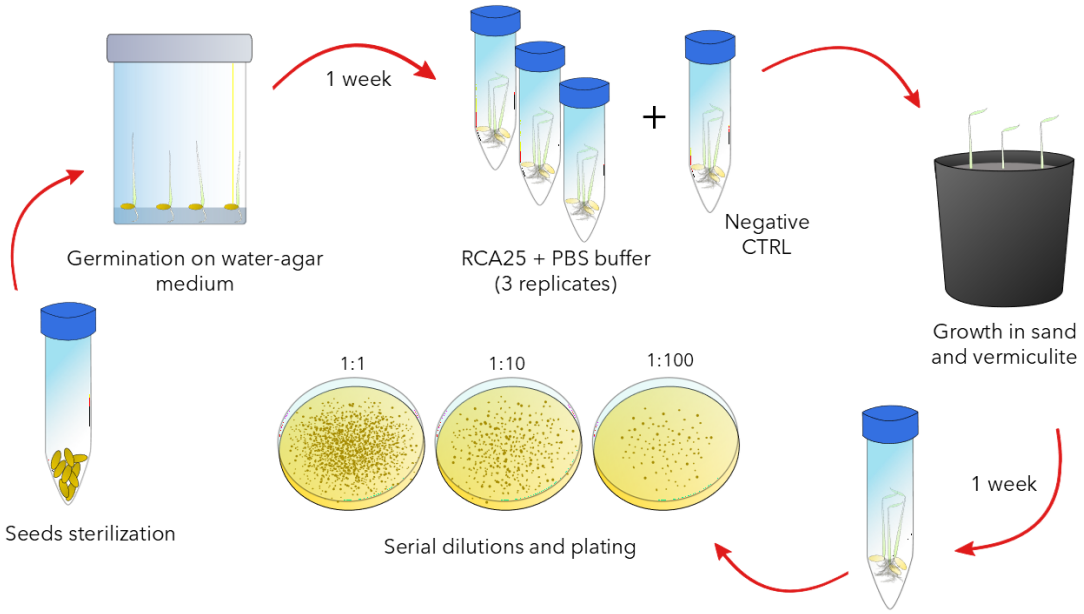


Figure 3.4: Schematic representation of the inoculation protocol, as described in the text.

Data analysis

4.1 Phenotypic data

Once all the data about the inoculation rates were recorded, the first step was to clean and organize them: missing data were removed, and some varieties which showed contrasting results were discarded from the dataset. The final number of colonizing bacteria for each test was defined considering the dilution factor and the total roots mass.

All the following *in silico* analyses were performed using the R programming language, jointly with RStudio software. The prepared phenotypic dataset was loaded on RStudio, for further organization and cleaning of the dataset.

4.2 Genotypic data

The initial step in processing the genotypic data involved filtering the raw dataset to remove missing data and rare variants. To achieve this, the following parameters were applied: genotypes with more than 10% of missing markers were removed. For the Minor Allele Frequency (MAF), a filter threshold of 5% was set. Additionally, individuals having more than 10% of missing data were re-

moved. All these filtering processes were carried out using Plink2 software (Chang *et al.*, 2015; Purcell *et al.*, 2007).

Once the filtered genotypic dataset was obtained, the following step consisted in pruning. Pruning involves selecting markers that are in Linkage Disequilibrium (LD) with each other and removing redundant markers, *i.e.* those that are in high LD with other markers. The pruning step was also performed using Plink2, selecting the following parameters: a sliding window of 50 nucleotides, a step size of 5, and an R2 threshold of 0.6. R2 is the criterion used to determine whether markers are in LD with each other.

The final step in processing the genotypic dataset was the imputation of missing data. This process, carried out using specific software and algorithms, allows for the estimation of missing markers based on the available ones, at the same position in other genotypes. Beagle software was used for imputation.

Subsequent data preparation and analysis steps were conducted using the R programming language in the R Studio software.

4.3 R script workflow

The pre-filtered and imputed genotypic data, along with the phenotypic data (*i.e.*, CFU values for the different replicates, adjusted based on root weight), were loaded into RStudio. Additionally, information regarding the geographical origins of the varieties included in the experiment was incorporated to generate descriptive plots of the collection used.

To load the files containing the data, the *readxl* (Wickham & Bryan, 2015) package was utilized, which is specifically designed for handling files in the proprietary Microsoft Excel .xlsx format.

The subsequent step involved further tidying and reorganizing the datasets to remove any remaining missing values (denoted as "NA" in R, meaning Not Available) and any unnecessary data that were not relevant to the analysis. This reorganization aimed to make the data more manageable and easier to work with in RStudio.

The values obtained from the different replicates were used to calculate the average infection rate for each accession by *K. sacchari* RCA25 (See Paragraph 2 "Loading and preparation of data" in the **R script** in **Supplementary material**).

Subsequent analyses involved the creation of graphical representations, including histograms and density plots, to visualize the distribution of the average CFU values calculated for each genotype. For these and subsequent plots, unless otherwise specified, the *ggplot2* package was employed (ggplot2: Elegant Graphics for Data Analysis (3e)). This package is widely used in R for creating complex and customizable visualizations, providing a powerful tool for data exploration and presentation (Paragraph 3.1 "Data distribution", in the **R script** in **Supplementary material**).

For statistical analysis and GWAS, the normal distribution of data is an important requirement. The distribution of data was analysed, using the Shapiro-Wilk test and graphical representation with histograms and QQ-plots (Quantile-Quantile plot, **Figure 5.1**). QQ-plot provides the distribution of the observed p-values in relation to the expected ones. The normal distribution of data as-

sumes that the former do not significantly deviate from the latter. Following these analyses, it was found that the data did not follow a normal distribution, so a square root transformation of the data was applied. The resulting values were found to have a normal distribution (**Figure 5.2**), and were used for subsequent analyses (Paragraphs 3.3 "QQ-plot of CFUs mean values", 3.4 "Data normalization", and 3.5 "Histograms and QQ-plots after transformation" of the **R script** in **Supplementary material**).

The boxplots graph (**Figure 5.3**) for the square root-transformed values of the different replicates produced for each genotype was very useful for visualizing the variability of the results among the different rice varieties tested and among the results of the individual replicates within the same genotype. Boxplots were also used to look for possible outliers (Paragraph 3.6 "Genotypes boxplots" in the **R script** in **Supplementary material**).

The data regarding the origin of the analysed genotypes, already available at CREA-CI institute in Vercelli, which is responsible for the maintenance of the accessions, were used to produce a pie chart representing the international distribution of the varieties considered (**Figure 3.1**). The pie chart shows the international representation of the collection, mainly composed of Italian varieties (See paragraph 3.7 "Origins pie chart" of the **R script** in **Supplementary material**).

The preliminary study of the data obtained in the laboratory continued with the calculation of common descriptive statistics parameters: range of values, mean, mode, median, standard deviation (sd), and Coefficient of Variation (CV) (**Table 4.1**). The Shapiro-Wilk test was also performed to evaluate the distribution of data, showing a not normally distributed dataset.

Table 4.1: Main descriptive statistics of the results obtained from the counting of CFUs.

PARAMETER	VALUE
Min	14493
Max	8131811
Mean	2216069
Median	1721052
1° quantile	812210
3° quantile	3229018
SD (Standard Deviation)	1793215

(For the part of R script regarding the descriptive statistical analysis of the phenotypic data, see Paragraph 4 "Statistical analysis", in **Supplementary material**).

Part of the descriptive study of genotypic data, as well as a fundamental component of the GWAS test, population structure allows for the classification of a collection of genotypes based on the sub-populations they represent. For the purposes of the GWAS test, taking population structure into account significantly reduces the number of false positives, *i.e.*, values erroneously resulting as significantly associated with the traits considered, as it corrects for the weight given by a shared genetic background due to common ancestors.

Regarding the population structure, two commonly used plots, Principal Components Analysis (PCA), and Neighbor-Joining (NJ) tree, help to visualize the distribution of genotypes based on the pedigree. The R packages used in this study to perform the GWAS analysis, *GAPIT* and *rMVP*, produce graphical out-

puts which comprehend these plots (**Figures 5.6**, **5.7**, **5.8**). However, they represent the entire collection analysed by these packages, which then select only the genotypes for which they find the relative phenotypic information to perform the association tests. For this reason, other plots for PCA and NJ tree were produced, in which the genotypes selected for the test were highlighted in red.

4.3.1 Principal-Components Analysis

The genotypic data was stored in .hmp (haplotype map) file, a standard format to manage this kind of data. This file was converted to .vcf (variant call format) file, and then to genlight object, a format required to perform the calculation of PCs, with the package *vcfR* (Knaus & Grünwald, 2017).

The calculation of the PCA was performed using the *adegenet* package (Jombart, 2008; Jombart & Ahmed, 2011). The resulting plot **(Figure 5.7)** was coloured to allow to visualize the distribution of the tested genotypes inside the entire Risinnova collection. See Paragraph 5.1 of the **R script** in **Supplementary material** for the code chunk relative to the PCA).

4.3.2 Neighbor-Joining Tree

As for the PCA, the genotypic data in genlight format was used to compute the NJ-tree. This was produced using the *ape* package (Paradis & Schliep, 2019). For the NJ-tree also, the tested varieties were highlighted in red in the relative plot **(Figure 5.9)** (See Paragraph 5.2 of the **R script** in **Supplementary material** for the code chunk relative to the NJ tree production).

4.3.3 Ancestry matrix

The population structure was analysed in depth with the *LEA* package (Frichot & François, 2015), that allows to infer the number of components of a population explaining most of the variance. For the population analysed in this experiment, a total of 30 clusters (Ks) were considered as sub-populations. The number of clusters was chosen based on the PCA results provided by *LEA* package. For each cluster, multiple runs are performed, and *LEA* provides functions to define which run is the best for a given K.

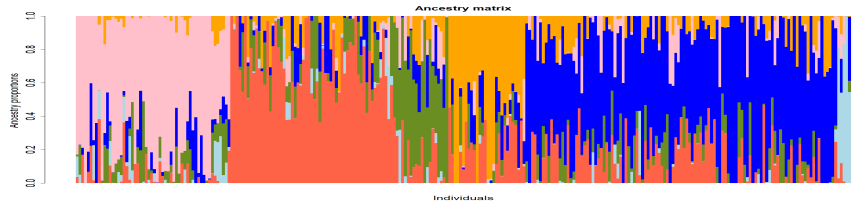
Based on the results obtained with *LEA*, four different K numbers were tested, to decide which was the best explaining the population structure:

NUMBER OF K	BEST RUN
6	4
9	4
16	4
19	4

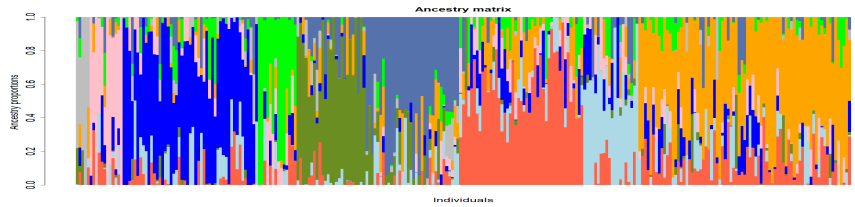
The results were represented as bar charts, with a column for each genotype and colours representing the clusters (*i.e.* the sub-populations). These plots are very helpful to determine the best number of sub-populations: selecting a low number of Ks will results in a simpler but less precise subdivision in clusters. On the other hand, a high number of Ks could give a better representation of the different sub-populations, but will increase the complexity of the plot, possibly resulting in a too strong subdivision of the genotypes.

Figure 4.1 shows the results of the 4 different hypothesized clustering. In **Figure 4.1a**, the population is divided in 6 clusters. In **4.1b**, 9 sub-populations

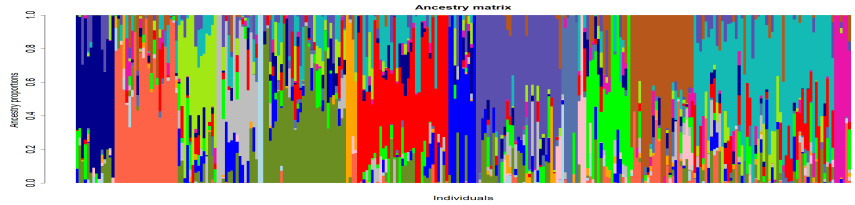
are represented. In **4.1c**, there are 16 sub-groups, and in **4.1d** the panel is divided in 19 sub-populations. From the different plots is clear that increasing the number of clusters can provide more precise results, but considering too many sub-populations can fragment too much the genotypes.



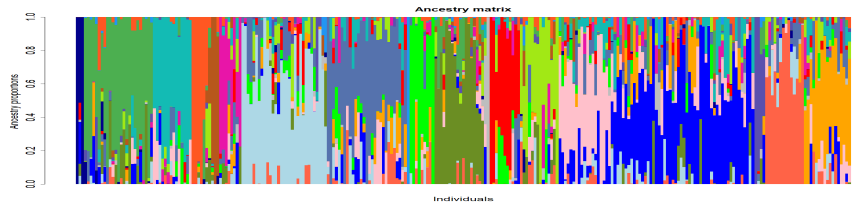
(a)



(b)



(c)



(d)

Figure 4.1: Plots representing different possible population structures for the genotypes of the Risinnova collection. Selecting different numbers of clusters, or K values (*i.e.*, sub-populations), yields different representations. Here, four possible clusterings were considered: (**a**) 6 Ks, (**b**) 9 Ks, (**c**) 16 Ks, and (**d**) 19 Ks. Based on the plots, K = 16 was chosen. Consequently, for performing GWAS tests, a number of sub-populations equal to 16 was selected.

4.3.4 GWAS

GWAS test was performed using two different R packages: *rMVP* (Yin *et al.*, 2021), and *GAPIT* v.3 (J. Wang & Zhang, 2021). While the resulting outputs are similar between the two packages, the preparation of the script is slightly different: *rMVP* requires a preparatory step, to produce specific genotypic, phenotypic, and map files to be used in the subsequent association analysis. *GAPIT* is more straightforward, because it is sufficient to load the previously prepared genotypic and phenotypic datasets. Code blocks relative to GWAS with *rMVP* and with *GAPIT* can be found at Paragraphs 6.1 and 6.2, respectively, inside the **R script** produced for the analysis, in **Supplementary material**.

4.3.5 LD-Decay

Once the results from the GWAS were obtained, markers showing interesting association levels were further analysed. To do so, LD-Decay was calculated, to have more precise information on the portion of the rice genome to study in proximity of the markers considered.

LD-Decay refers to the rate at which Linkage Disequilibrium (LD), the non-random association between alleles at different *loci*, decreases as the physical distance between those *loci* increases. It is typically measured using statistics such as r^2 or D' and visualized as a plot of LD (Y-axis) against physical distance (X-axis). LD-Decay is influenced by factors such as recombination rates, population history, and selection.

To calculate the LD-Decay, the script provided by mohsinali1990 on GitHub, was used.

The LD matrix was calculated using **TASSEL** software, version 5 (Bradbury *et al.*, 2007). The matrix was loaded in RStudio to produce the plot of the LD-Decay (**Figure 5.17**).

The code block relative to LD-Decay evaluation and plotting can be found at Paragraph 9 of the **R script** produced for the analysis, in **Supplementary material**.

Results

5.1 CFUs counting and data collection

The number of CFUs was manually counted after 24 hours from plating. For each genotype tested, 3 replicates were produced, and for each replicate, 3 seedlings were considered. The number of CFUs corrected for the dilution factor and the roots weight was considered.

The resulting values of mean inoculation rates for each genotype, after the cleaning of data and removal of bad quality results, are shown in **Table 5.1**.

Table 5.1: Genotypes selected for the association study after the inoculation results.

N°	GENOTYPE	MEAN CFUs
1	Slava	14493
2	Oscar X Suweon	17910
3	Handao_297	18100
4	Plus	22778
5	Roxani	24216
6	Salvo	35800
7	Carnaroli	39770
8	Giza 178	46814
9	Redi	75805
10	Xiangou2	116964
11	Silla	119580
12	Asia	141343
13	Bengal	152922
14	S101	156766
15	Upla64	160030
16	Goolarah	204842
17	Lido	216000
18	Medusa	223333
19	Ota	246142
20	CT36	260087
21	Ronaldo	264576
22	Plovdiv22	276596
23	Gange	299414
24	Chipka	320144
25	Fidji	363612
26	Vialone_Nano	374156
27	Frances	423395
28	Merle	439560
29	Soure	494444
30	Sakha_103	526188
31	Argo	532305
32	M204	546735
33	Zenith	565921
34	Giano	586740
35	Flipper	586819
36	Bellepatna	592203
37	Ibo380.33	596561
38	Texmont	598106
39	Calmochi_101	649107
40	Prometeo_2	664236
41	Ringo	665961
42	Safari	685304
43	Roncarolo	689557
44	Fast	716499
45	Loto	760923
46	Burma	774604
47	IR64	786765
48	Agostano	799273
49	Rubino	825147
50	Onice	829532
51	Maybelle	843614
52	A201	852201
53	Sirio_CL	874229
54	L205	878378
55	Sagres	880375
56	Giada	882284

N°	GENOTYPE	MEAN CFUs
57	M006	909259
58	Harlem	949441
59	Gigante Vercelli	967266
60	Upla63	978363
61	Greppi	994233
62	Zhenshang97	1017925
63	Familia_181	1047829
64	Upla77	1052676
65	Delfino	1066746
66	Seln244A	1076756
67	Krystallino	1091110
68	Rodeo	1118244
69	Puntal	1119875
70	Taichung_G65	1125000
71	Cocodrie	1141053
72	Jacinto	1161189
73	S102	1167803
74	Sandora	1178357
75	Nilo	1204545
76	Rizzotto_51.1	1280667
77	Lencino	1307117
78	Molo	1333768
79	Agata	1335526
80	Mareny	1363730
81	Nano	1379743
82	Americano	1405193
83	Eurose	1421944
84	Lady Wright	1441595

N°	GENOTYPE	MEAN CFUs
85	Dellmont	1449388
86	Thaiperla	1454545
87	Ulisse	1478416
88	Roncolo	1502661
89	Graldo	1506692
90	Frangrance	1520894
91	Pegonil	1537512
92	P6	1551926
93	Strella	1622631
94	Colina	1657112
95	Luxor	1667622
96	Novara	1721052
97	YRM6.2	1745833
98	Upla75	1772017
99	Samba	1833942
100	Allorio	1846223
101	L201	1934124
102	Musa	1999647
103	Maratelli	2085854
104	Santerno	2154507
105	Europa	2208837
106	Ercole	2216190
107	Upla66	2252439
108	Alexandros	2253182
109	Bianca	2257799
110	Viale	2267632
111	Robbiosel1	2294373
112	Saloio	2305493

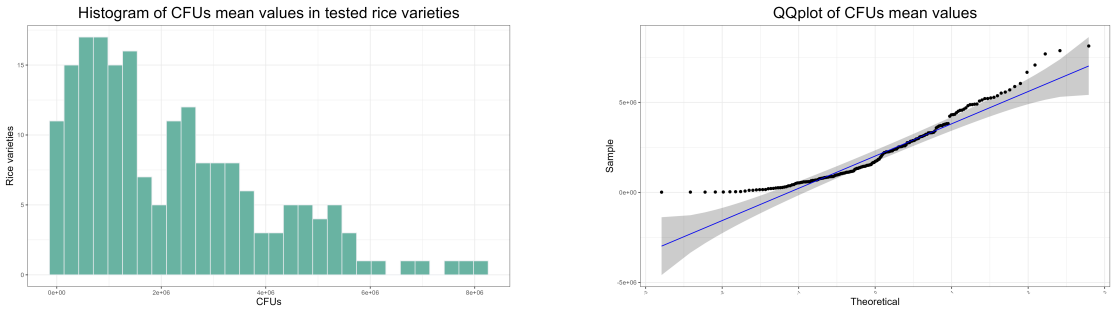
N°	GENOTYPE	MEAN CFUs
113	Italmochi	2312116
114	Italpatna x Milyang	2377507
115	Russo1	2391026
116	Cigalon	2408497
117	Bombilla	2412317
118	RPC12	2475450
119	Sakha_102	2522727
120	L202	2537656
121	Dixiebelle	2541379
122	Veneria	2549278
123	Sprint	2577187
124	Pierina Marchetti	2588119
125	GZ6296	2592186
126	Gitano	2651085
127	Anseatico	2757973
128	Clot	2760360
129	Panda	2770053
130	Saedinenie	2838542
131	Saturno	2844444
132	Ostiglia	2876545
133	Balilla	2883291
134	Ermes	2915444
135	Smeraldo	2966483
136	Italpatna48	2979902
137	Karnak	3004409
138	Vulcano	3076923
139	Campino	3099775
140	Miara	3111884
141	Alpha	3143756
142	Vialone_190	3174077
143	Volano	3225179
144	Augusto	3232856
145	Originario	3306704
146	Bahia	3311465
147	Giovanni Marchetti	3312871
148	Galileo	3322263
149	Rexmont	3340909
150	Gladio	3415181
151	CT58	3585029
152	Handao 11	3639540
153	Rubi	3687871
154	Alpe	3710113
155	Zena	3711678
156	Sesiamochi	3765873
157	Setantuno	3787879
158	Mesanes2	3816790
159	Sequial	3823390
160	Castelmochi	4231373
161	Lamone	4310185
162	Selenio	4322414
163	Ribe	4340281
164	Victoria	4427971
165	Eurosis	4507466
166	Gardesadri	4559962
167	Santandrea	4568125
168	Graal	4623727

N°	GENOTYPE	MEAN CFUs
169	Luna	4694813
170	King	4816071
171	Venere	4873485
172	Super	4879077
173	Ducato	4898873
174	Ariete	4908333
175	Calendal	5074155
176	Baldo	5133778
177	Artemide	5210398
178	Iac32.52	5214912
179	Rinaldo Bersani	5248433
180	Tejo	5283419
181	Antares	5370297
182	Carnise	5516116
183	Senatore Novelli	5572917
184	Alan	5695765
185	Carina	5880488
186	Apollo	6050962
187	Ibo400	6671846
188	Gritna	7075381
189	Sandoca	7691822
190	M203	7873864
191	Brazos	8131811

5.2 Data quality control and management

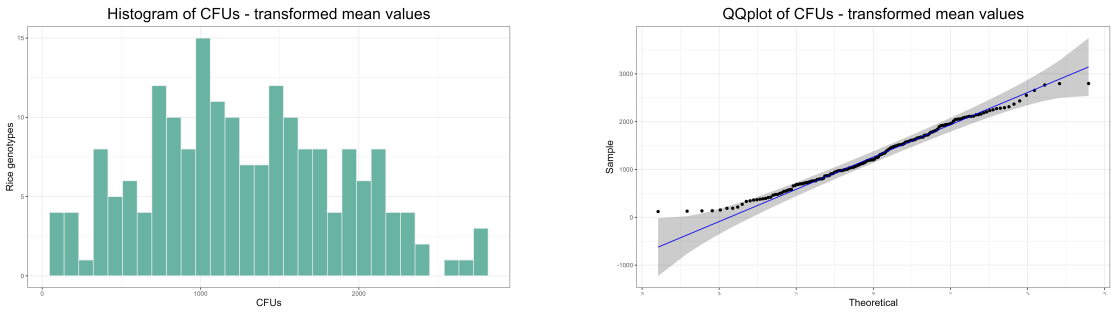
5.2.1 Distribution

For each genotype tested, mean values of the replicates were calculated. The resulting values were represented with a histogram and a QQ-plot to analyse the distribution of the data, before the normalization (**Figure 5.1**), and after the square root transformation step (**Figure 5.2**).



(a) **(b)**

Figure 5.1: (**a**) Histogram of the distribution of rice genotypes based on the resulting mean inoculation rate. (**b**) The distribution of the data represented through a QQ-plot. As can be seen, the distribution is not normal. This condition was further assessed through a Shapiro-Wilk test.



(a) **(b)**

Figure 5.2: The same plots represented in **Figure 5.1**, after the transformation of the data through the square root method.

5.2.2 Boxplots

Boxplots for each genotype tested were plotted together to visualize the distribution of the results (**Figure 5.3**). Each boxplot represents the values of the transformed data for all the replicates produced for each accession.

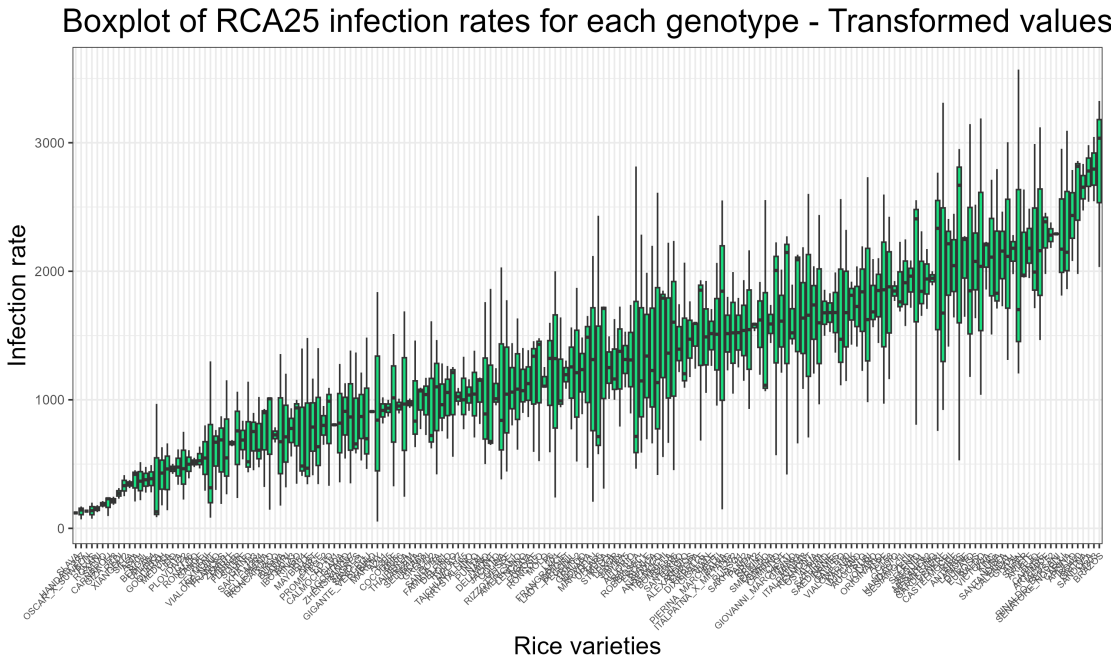


Figure 5.3: Boxplots showing the distribution of the results of the replicates for each rice accession, from the genotype showing the lowest level of colonization to the one with the highest values.

5.3 Pruning

Marker density plots allow to visualize the even coverage of the genome represented by markers. Here, the plots produced with *rMVP* package were used. In **Figure 5.4** , SNPs density plots before (62421 SNPs), and after pruning (14220 SNPs) are represented.

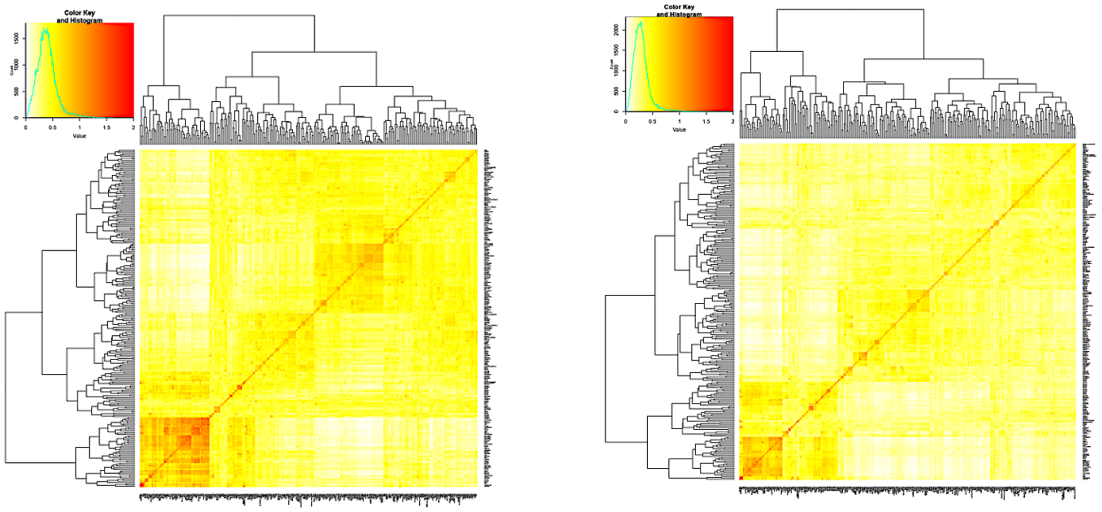
(a) **(b)**

Figure 5.4: SNPs density plots produced with *rMVP* package. In (**a**) the density of markers before the pruning step. In (**b**) the remaining SNPs after pruning.

5.4 Kinship and population structure

Other graphical representations useful to study the structure of the rice panel used in this study are the kinship heatmap, and the population structure. These elements are fundamental to perform robust GWAS analyses, but also the plots are valuable to visualize the relations between the accessions.

Figure 5.5 shows the kinship matrix, as a heatmap. The plot is produced with the *GAPIT* package. The heatmap shows how much the pedigrees of the genotypes considered are similar to each other, and the kinship is useful to reduce false positives associations given by the similarity of related genotypes. In **Figure 5.5a** the kinship heatmap is represented before the pruning step, while in **5.5b** the reduced intensity of the signal is due to the pruning, *i.e.* the removal of redundant markers in LD.



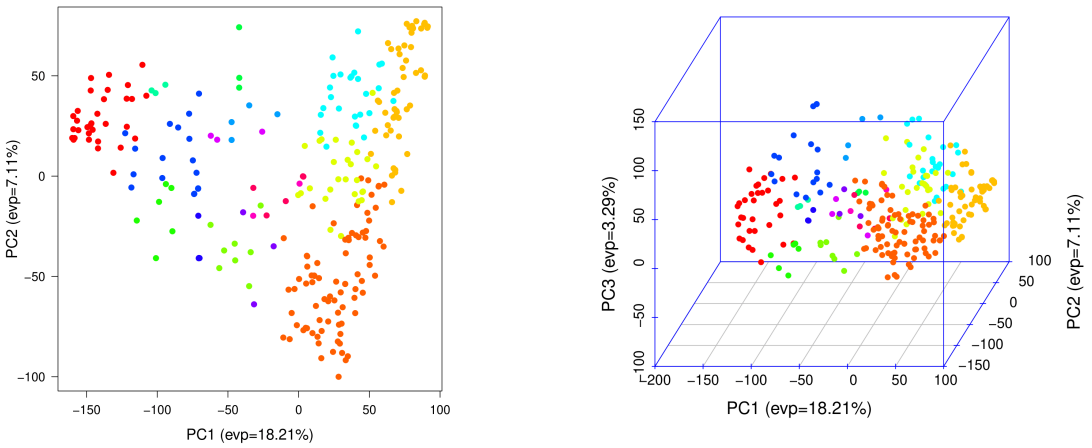
(a) **(b)**

Figure 5.5: Heatmap of the kinship matrix. (**a**) before pruning the genotypic data, (**b**) after pruning the genotypic data.

Another useful graph to define the population structure of the collection of genotypes is the PCA (Principal Component Analysis). The PCA is a statistical approach to represent big datasets in a simplified way, calculating Principal Components (PCs), that are pseudo-variables defined starting for the real ones. A single PC represents different original variables, and the PCs are defined based on the overall amount of variance explained by the variables clustered in that PC. In this way, PCs are defined from the one explaining the most variance (PC_1) to the one explaining the least (PC_n). PCA can be used to visualize the distribution of the individuals when studying the population structure, distributing the genotypes on a scatter plot in which each axis corresponds to a PC.

GAPIT provides the PCA plot, with the number of PCs selected during the compilation of the script **(Figure 5.6)**. In this case, the number of sub-populations, and thus of PCs, was set to 16 after the analysis of the population structure with

LEA package **(Figure 4.1)**. For the association test, *GAPIT* considers only the genotypes for which a corresponding phenotypic trait is provided. However, the PCA and NJ tree were made from all the Risinnova collection. For this reason, to provide a better overview of the structure of only the set of genotypes used in this experiment, a second PCA plot was produced **(Figure 5.7)**.



(a) **(b)**

Figure 5.6: PCA scatterplots produced with *GAPIT* package. The distribution of all the genotypes composing the Risinnova collection is showed, in a 2D (**a**), and a 3D (**b**) distribution.

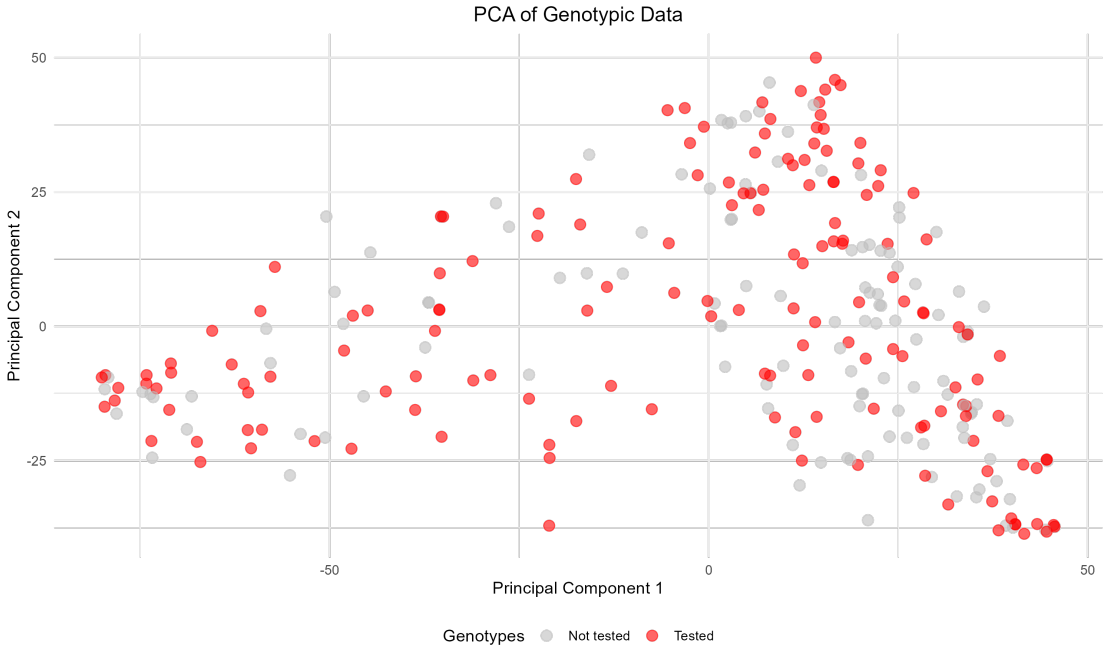
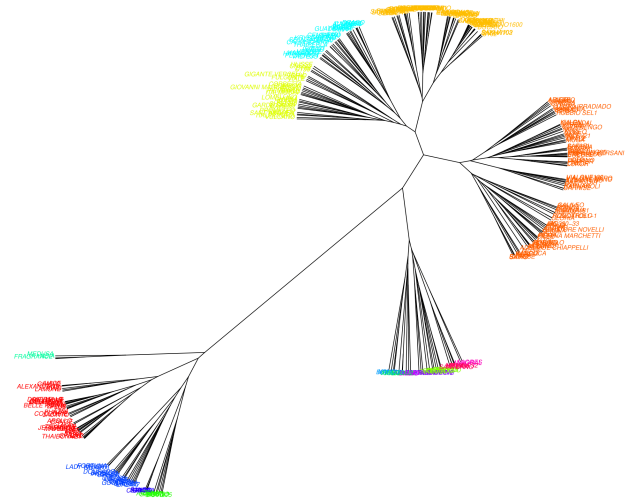
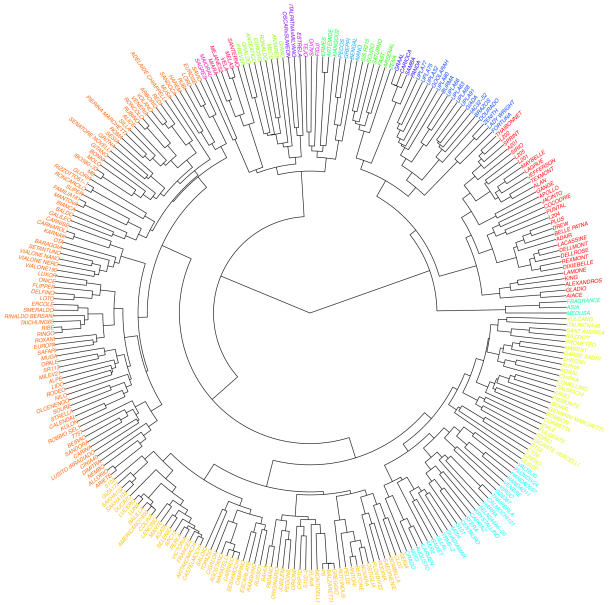


Figure 5.7: PCA scatterplot showing the distribution of the genotypes into sub-populations. Genotypes selected for the experiments are highlighted in red.

The Neighbor-Joining (NJ) tree gives a representation of the entire collection in a phylogenetic tree, allowing to visualize the descendance of each genotype. In **Figure 5.8** the NJ tree provided by *GAPIT* is represented.



(a)



(b)

Figure 5.8: Neighbor-Joining tree of the Risinnova collection produced with *GAPIT* package. The 16 different sub-populations are represented in different colours. In (**a**) the radial representation of the NJ tree, in (**b**) a circular NJ tree.

Figure 5.9: the Risinnova phylogenetic tree, with the genotypes used in the study highlighted in red.

5.5 Genome-Wide Association Studies

GWAS tests were performed with two different R packages: *rMVP*, and *GAPIT*, version 3. Both the packages provide multiple output files from the analysis. The two most important plots are represented by Manhattan plots, and QQ-plots. A Manhattan plot is a graphical representation used in GWAS tests to display the distribution of genetic markers across chromosomes. On the X-axis, the markers are organized by their genomic position, grouped by chromosome. On the Y-axis, the significance of each marker's association with a phenotypic trait is plotted, as the -$\log_{10}$(p-value). Peaks in the plot indicate genomic regions where markers show strong statistical associations with the trait, suggesting potential *loci* of interest.

Multiple models are available to perform a GWAS, from the simplest t-test to the most advanced *Blink* (Huang *et al.*, 2019), *SUPER* (Q. Wang *et al.*, 2014), and *FarmCPU* (Liu *et al.*, 2016). For this analysis, *FarmCPU* model was chosen because it provides statistically robust results.

GAPIT produces some descriptive plots about the analysed phenotypic data **Figure 5.10**.

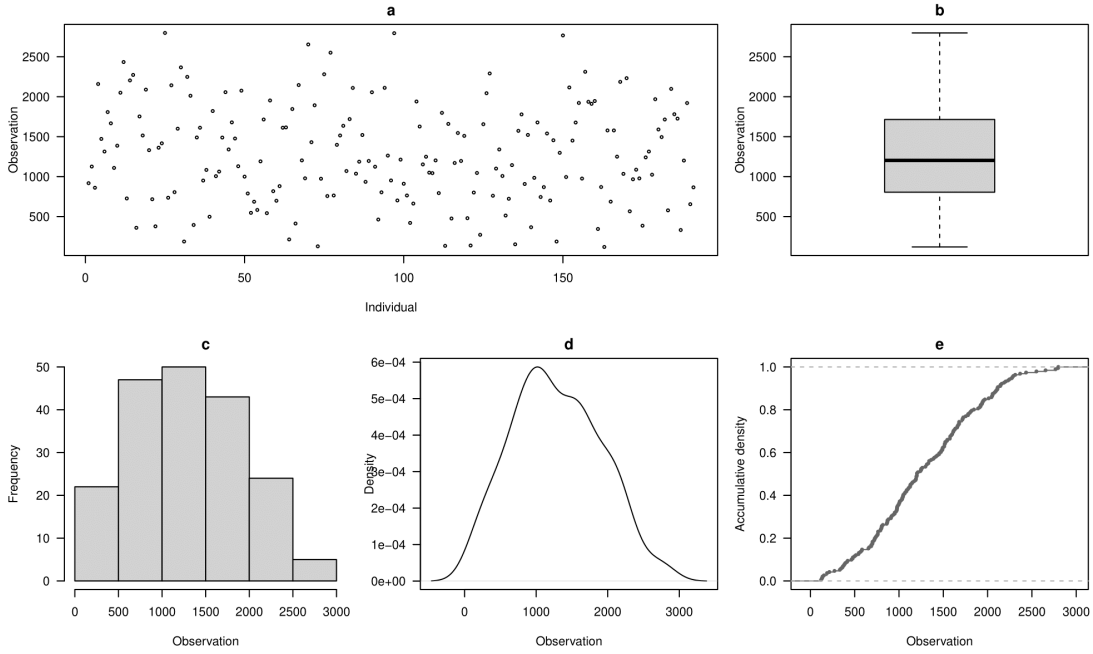
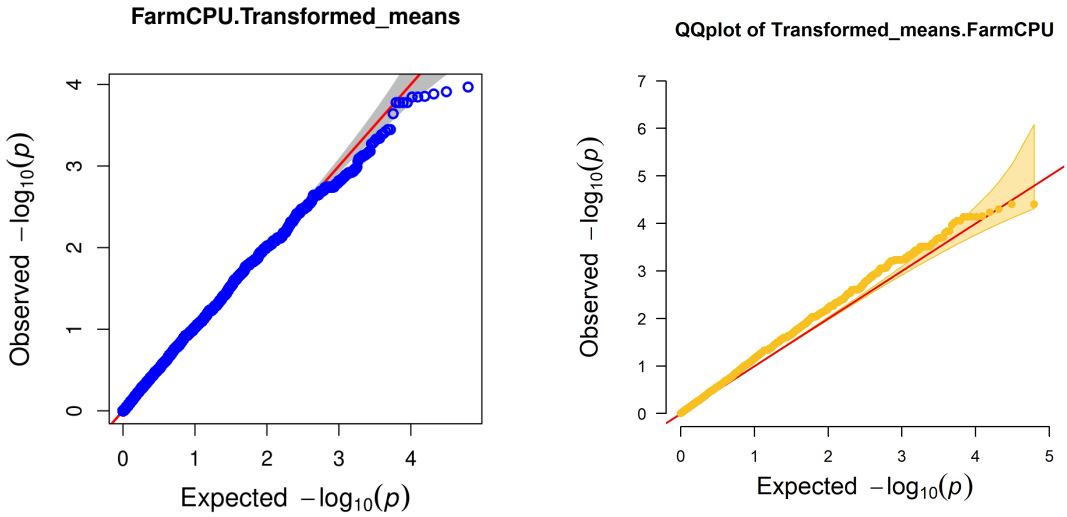


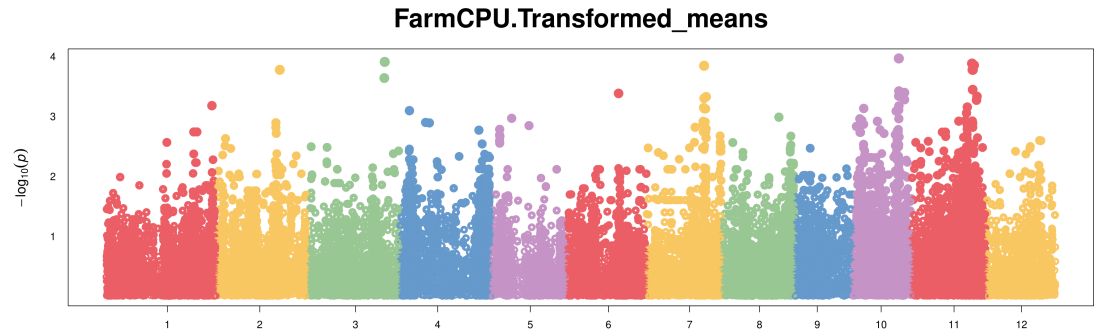
Figure 5.10: Descriptive plots produced with *GAPIT* package about the phenotypic data analysed. (**a**) Distribution of values. (**b**) boxplot, (**c**) frequency histogram, (**d**, **e**) density plots.

Together with the Manhattan plot, the QQ-plot is the second main graphical representation provided from GWAS analyses to look for possible association **(Figure 5.11)**. Each point represents a marker and is expected that almost all points lay approximately on the line, or inside the confidence area. Upper tail deviations could indicate a significant association between a marker and the trait considered.

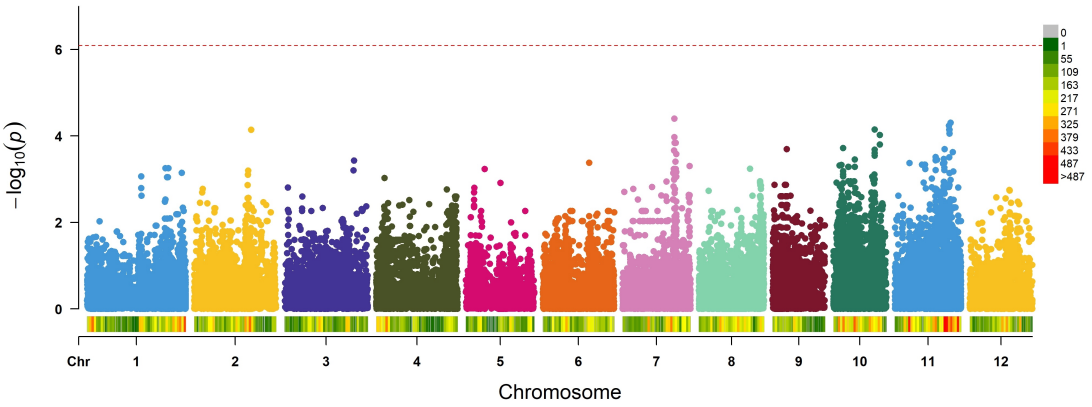


(a) **(b)**

Figure 5.11: QQ-plots showing the distribution of observed data to an expected distribution. In (**a**) the QQ-plot produced with *GAPIT*, in (**b**) the one obtained with *rMVP*. No significant associations resulted from the analyses.



(a)



(b)

Figure 5.12: Manhattan plot resulting from the GWAS analyses performed on the same dataset using the *GAPIT* (**a**) and *rMVP* (**b**) packages. The red dashed line in (**b**) represents the statistical threshold above which a marker can be considered significantly associated with the trait of interest. This line is not present in (**a**) because *GAPIT*, by default, optimizes the dimensions of the plot, and in this analysis, no markers were found to be associated with the colonization rate.

5.5.1 GWAS with a reduced dataset

The analysis with all the genotypes tested for the inoculation rates with *K. sacchari* RCA25 did not show any significant association. Considering the broad

variability of values in replicates recorded for many genotypes (see **Figure 5.3**), a further analysis was performed.

Defining a threshold based on standard deviation, a subset of 100 genotypes showing the least variability between accessions was selected **(Figure 5.13)**, and a new GWAS was performed with the reduced dataset (See the Paragraph 7 of the **R script** in **Supplementary material** for the relative code).

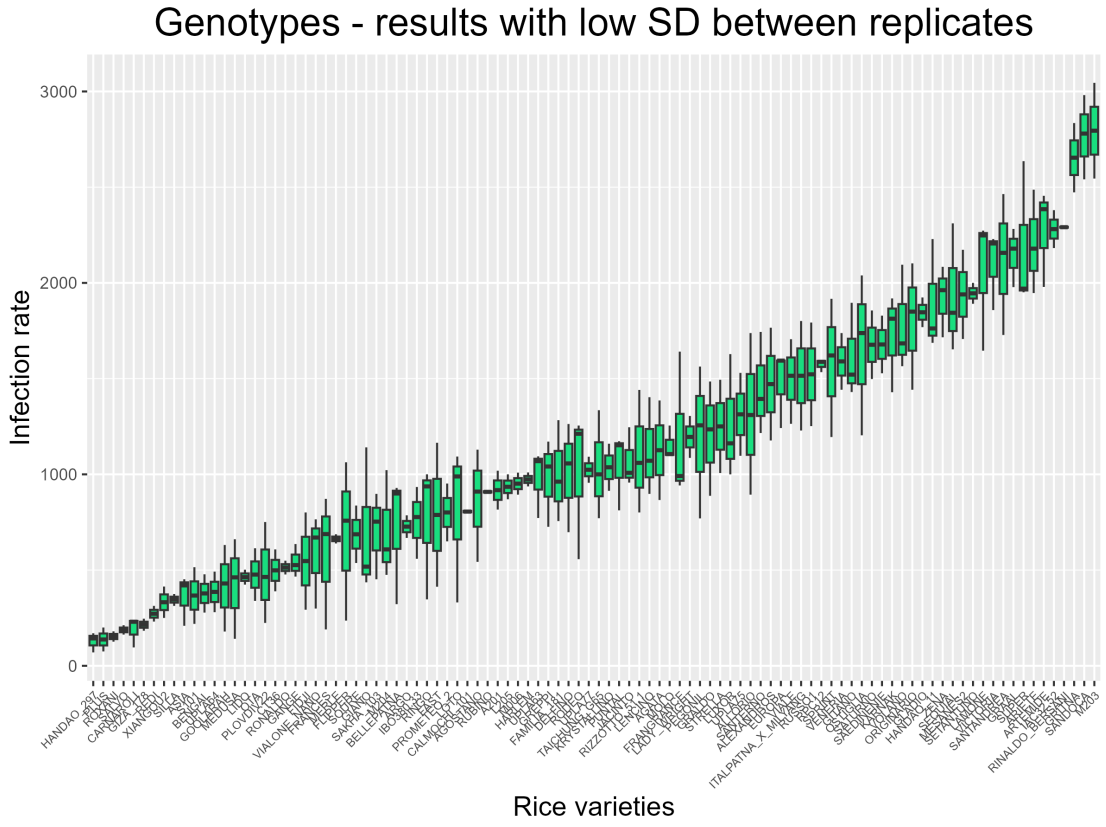


Figure 5.13: Boxplots of the replicates for the 100 genotypes composing the reduced dataset. The selection was based on the sd of the replicates, keeping the 100 accessions with the lowest values of sd.

The reduced dataset was used for a new GWAS test, and the resulting QQ and Manhattan plots are shown in **Figure 5.15** and **Figure 5.16**, respectively. Some plots produced with *GAPIT* and describing the data distribution of the reduced dataset are shown in **Figure 5.14**.

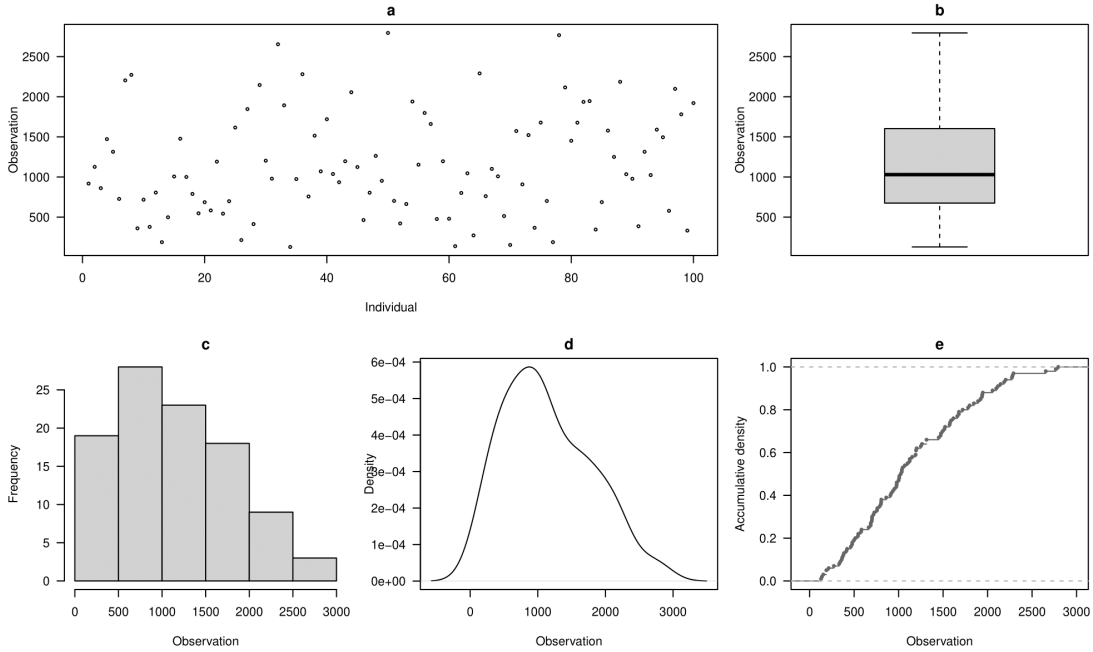
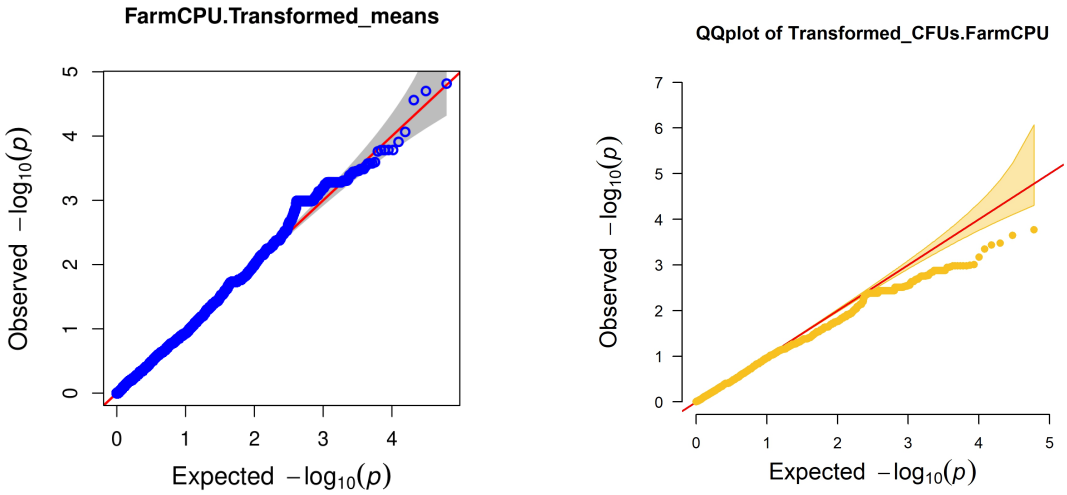
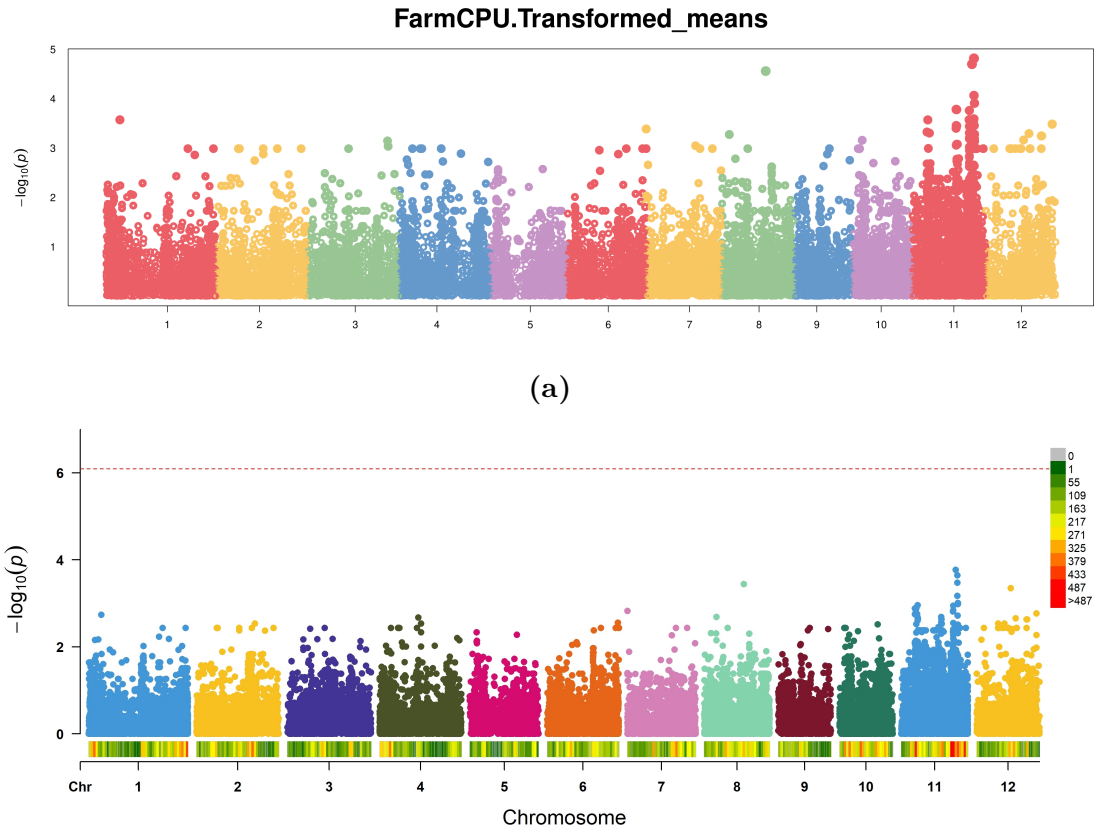


Figure 5.14: Descriptive plots of the reduced dataset provided after GWAS with GAPIT. (**a**) Distribution of values. (**b**) boxplot, (**c**) frequency histogram, (**d**, **e**) density plots.



(a) **(b)**

Figure 5.15: QQ-plots obtained from the GWAS analyses performed on the reduced dataset of the 100 genotypes having the lowest sd between replicates. In (**a**) the QQ-plot produced with *GAPIT*, in (**b**) the one obtained with *rMVP*.



(a)

(b)

Figure 5.16: Manhattan plots showing the results of the GWAS on the reduced dataset of 100 genotypes. In (**a**) the Manhattan plot obtained with *GAPIT*, in (**b**) the plot produced with *rMVP*.

After the analysis of the reduced dataset, no significant associations were detected. However, a region on chromosome 11 consistently showed a high peak in both tests across all replicates and in the reduced dataset, prompting further analyses.

In both GWAS tests performed, the highest peak included the SNP **S11_23670478**. Although this marker was not significantly associated, it exhib-

ited a consistent signal. A GWAS uses genotyped markers to identify associations. However, it is unlikely that the marker itself is directly responsible for the trait under investigation. Instead, the association is often due to LD. In GWAS, markers act as probes to identify genomic regions in LD that show a statistical association with the trait of interest. The extent of LD can vary significantly across species and is influenced by factors such as the mode of reproduction. For example, in plants, the type of pollination (self-pollination vs. cross-pollination) can significantly affect LD patterns.

To delineate the genomic region surrounding a detected marker that is likely to influence the trait, the LD-Decay value is calculated. This parameter represents the average distance over which the correlation between genetic markers significantly diminishes. In this specific case, the LD-Decay was observed to extend up to approximately 182 kilobases (kb) (**Figure 5.17**), defining the region of interest around the markers considered.

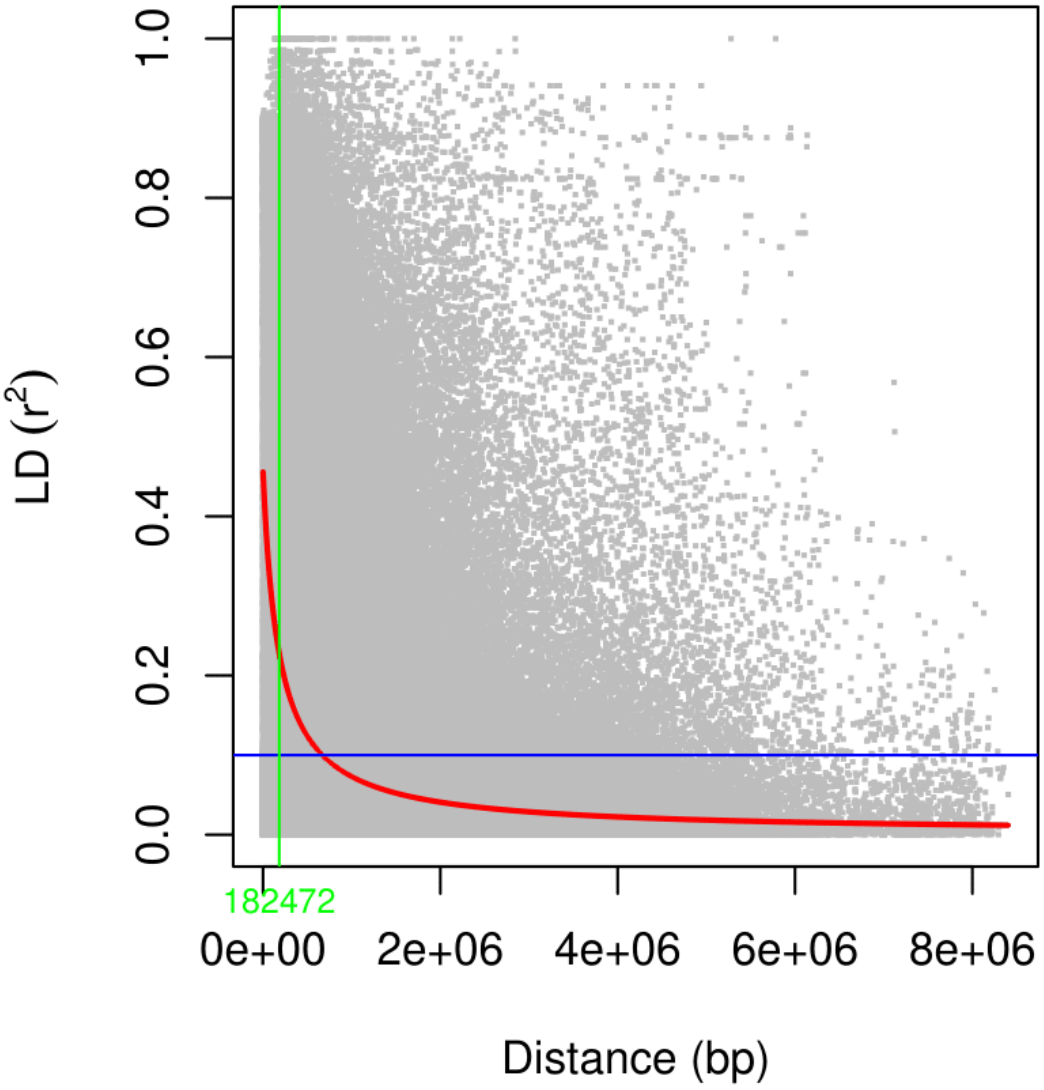


Figure 5.17: LD-decay plot, used to define the region to consider in LD with the markers of interest. The green line indicates the distance in base pairs (bp) that defines the window within which two regions of the DNA can be considered in LD.

Considering the calculated LD-Decay, a window of 364 Kb was considered (182 Kb before and 182 Kb after the exact position of the marker **S11_23670478**). The

region considered comprised from 23488478 bp to 23852478 bp on the chromosome 11. The genomic region was studied on the rice genome browser provided by the Rice Annotation Project Database (RAP-DB) (Sakai *et al.*, 2013), and based on the *O. sativa* spp. *japonica*, cultivar Nipponbare reference genome (**Figure 5.18**).

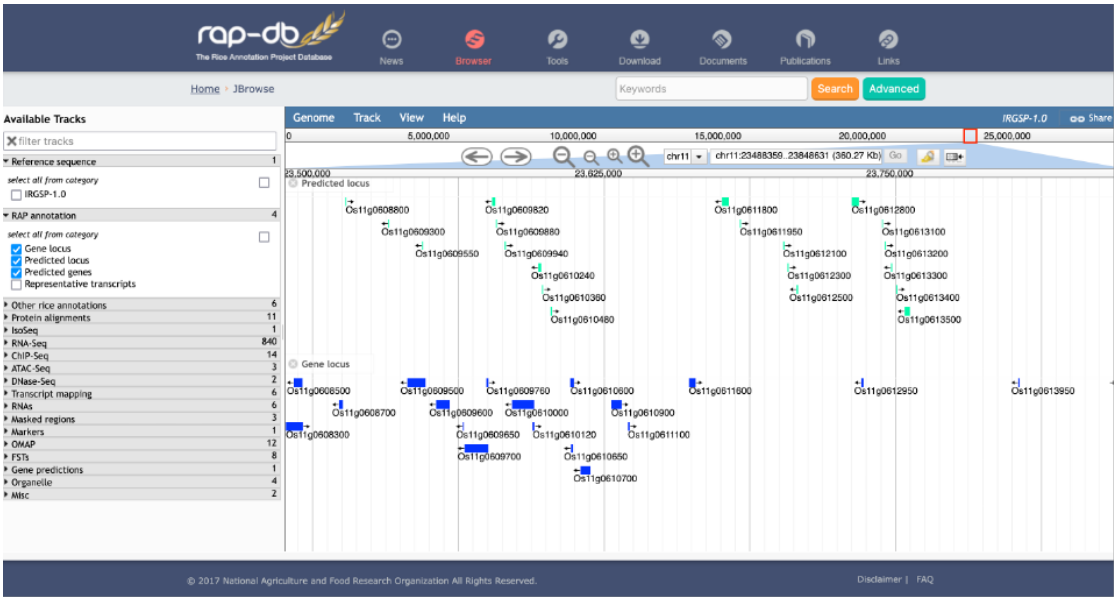


Figure 5.18: The RAP-DB genome browser. The region represented is a window of 364 Kb on chromosome 11.

The region of interest was furthermore studied using the tool BIOMART database on the **EnsemblPlants** website. The list of gene *loci* comprised in the selected region is shown in **Table 5.2**.

Table 5.2: list of gene *loci* comprised in the region 182 Kb upstream and downstream the marker S11_23670478 (ncRNA: non-coding RNA).

GENE STABLE ID	TRANSCRIPT ID	GENE DESCRIPTION	GENE NAME	GENE TYPE
Os11g0608300	Os11t0608300-01	Barley stem rust resistance protein	Receptor-like Cytoplasmic Kinase 337	Protein coding
Os11g0608500	Os11t0608500-00	Non-protein coding transcript		ncRNA
Os11g0608700	Os11t0608700-00	Similar to Jacalin-like lectin domain containing protein	Receptor-like Cytoplasmic Kinase 338, Jacalin-related lectin 24	Protein coding
Os11g0609500	Os11t0609500-01	Similar to Jacalin-like lectin domain containing protein	Receptor-like Cytoplasmic Kinase 339, Jacalin-related lectin 25	Protein coding
Os11g0609600	Os11t0609600-01	Similar to 14-3-3-like protein	G-box factor 14-3-3h protein	Protein coding
	Os11t0609600-02	14-3-3 protein, Flood adaptation, Anaerobic Germination (AG) and Anaerobic Seedling Development (ASD) tolerance (Nipponbare: AG sensitive, Weedy rice: AG tolerance		
	Os11t0609600-03	14-3-3 family protein, Temperature-dependent seed germination, Negative regulation of ABA signalling (Nipponbare: partial loss-of-function allele)		
Os11g0609650	Os11t0609650-00	Non-protein coding transcript		ncRNA
Os11g0609700	Os11t0609700-00	Disease resistance protein domain containing protein		Protein coding
Os11g0609760	Os11t0609760-00	Non-protein coding transcript		ncRNA
Os11g0610000	Os11t0610000-00	Similar to NB-ARC domain containing protein		Protein coding
Os11g0610120	Os11t0610120-00	Non-protein coding transcript		ncRNA
Os11g0610600	Os11t0610600-00	Zinc finger, RING-type domain containing protein	RING-type E3 ubiquitin ligase 333	Protein coding
Os11g0610650	Os11t0610650-00	Hypothetical protein		Protein coding
Os11g0610700	Os11t0610700-01	WD40 repeat-like domain containing protein	WD40-repeat transcription factor 190	Protein coding

Continued on next page

Table 5.2: list of gene *loci* comprised in the region 182 Kb upstream and downstream the marker S11_23670478 (ncRNA: non-coding RNA). (Continued)

Os11g0610900	Os11t0610900-01	Seryl-tRNA synthetase, Early chloroplast development at high temperature (Os11t0610900-01).	temperature-sensitive chlorophyll-deficient 11	Protein coding
	Os11t0610900-02	Hypothetical conserved gene		
Os11g0611100	Os11t0611100-00	Hypothetical conserved gene		Protein coding
Os11g0611600	Os11t0611600-01	Hypothetical protein		Protein coding
Os11g0612950	Os11t0612950-00	Similar to AtPPa1 (*Arabidopsis thaliana* pyrophosphorylase 1)		Protein coding
		Inorganic diphosphatase		
Os11g0613950	Os11t0613950-00	Non-protein coding transcript		ncRNA
Os11g0614400	Os11t0614400-01	Similar to Patatin-like protein 1	Patatin-related phospholipase A II *theta*	Protein coding
Os11g0614500	Os11t0614500-01	Similar to Patatin-like protein	Patatin-related phospholipase A II *iota*	Protein coding
	Os11t0614500-02	Similar to Patatin-like protein		

Discussion

In the described work, some aspects of the beneficial interactions between rice plants and growth-promoting microorganisms were studied. The aim of this research was to evaluate whether there are elements in the rice genome that are closely involved in the interaction with this type of microorganism. To this end, the colonization rate of the bacterium *Kosakonia sacchari* RCA25 was analyzed and quantified in a collection of approximately 200 varieties of rice (*O. sativa* spp. *japonica*), through GWAS analysis. The collection was studied to characterize the population structure of the genotypes, their relationships with one another, and the genotyping data. The data were cleaned and organized as needed for subsequent analyses. Various trials were conducted in the lab to optimize the inoculation protocol. The resulting data, collected by counting CFUs, were checked, cleaned, and organized for use in association tests.

Based on the available information and the results of the GWAS tests, the genes identified do not appear to represent valid candidates as genetic elements involved in the studied interaction. No significant associations were found in the GWAS study. However, during the execution of various association tests, a peak on chromosome 11 showed a consistent presence, albeit below the significance threshold. For this reason, the corresponding marker was analyzed in greater

detail using databases such as RAP-DB and EnsemblPlants. By calculating the LD-decay rate, a window of 364 Kb (from 23,488,478 bp to 23,852,478 bp) was defined for the marker (SNP S11_23670478) associated with the peak, to search for *loci* potentially relevant to the purposes of this study.

From the list of genes included in this region, some represent hypothetical genes or proteins for which the currently available information is insufficient, but they could be interesting subjects for further study: **Os11g0610650**, **Os11g0611600** (hypothetical proteins), and ***Os11g0611100*** (hypothetical conserved gene) (**Table 5.2**).

Other known genetic elements found in the region might be potentially involved in the plant-microorganism interactions investigated in this work, so they were further explored in online databases.

The gene ***OsRLCK337*** (Receptor-like Cytoplasmic Kinase 337, gene ID: ***Os11g0608300***) does not seem to be expressed in root tissues, so it is probably not involved in the presence of *K. sacchari* RCA25, an endophytic bacterium that colonizes the roots.

As for the gene ***Os11g0610000***, there is limited information available, but its product is similar to an NB-ARC domain-containing protein. NB-ARC (Nucleotide-Binding adaptor shared by APAF-1, R proteins, and CED-4) protein are regulators of signaling pathways and play fundamental roles in effector recognition and signal transduction in plant growth (Pan *et al.*, 2022). The NB-ARC domain is present in many proteins involved in resistance against pathogens, regulating the activity of resistance proteins (Takken *et al.*, 2006; van Ooijen *et al.*, 2008).

Os11g0614400 and ***Os11g0614500*** are genes coding for Patatin-related phospholipase A II *theta* and *iota*, respectively. These molecules are glycoproteins with lipid acyl-hydrolase activity, and catalyzing the hydrolysis of phospholipids, glycolipids, sulfolypids, and mono- and diacylglycerols (Rydel *et al.*, 2003). Some studies have shown that patatin-like proteins may be involved in interactions between plants and microorganisms, either reducing or increasing the host's resistance to pathogens, based on the species (Camera *et al.*, 2009).

Analyzing the results obtained, two aspects should be considered: the first is that it must be kept in mind that interactions between the plant and microorganisms are complex, presumably mediated by a large number of mechanisms and processes, and consequently by a large number of minor effect QTLs. Therefore, although no specific genomic regions strictly correlated with bacterial colonization rates were found, the wide variation observed among genotypes still suggests the possibility that there should be genomic elements involved in this type of interaction. The second aspect to consider is that, although a relatively large number of rice genotypes were analyzed in this experiment, they all belonged to the *japonica* subspecies. From the results obtained, further developments of the study could consider an in-depth analysis of the most promising genes found in LD with the SNP S11_23670478, which were briefly discussed here. It would also be useful to expand the study by including rice genotypes belonging to other subspecies, such as *indica*. Another possible development in the analysis could be the repetition of the inoculation experiment and the comparison between the two, to check the repeatability of the genotype averages. The number of replicates for each genotype could also be increased, and a further optimization of the inoculation conditions could lead to minimize the environmental effects on the success rate of the inoculum.

As a final consideration, the aim of this work was to investigate potential elements within the genome of the selected rice accessions that could influence the interaction between the plant and the diazotrophic bacterium *K. sacchari*, and, if present, to provide an initial evaluation of them. The study did not aim to analyze the positive effects of the bacterium on rice plant development, which is why parameters such as NUE (Nitrogen Use Efficiency) or growth under nitrogen deficiency were not assessed. These aspects could be of significant interest and may be considered in a follow-up study.

GWA studies have already proven to be proficient in researching the interactions between plants and microorganisms. For instance, a study on 340 rice accessions of *japonica* and *indica* identified 16 *loci* associated with blast resistance, with two *loci* in *japonica* and one in *indica* showing significant associations (Raboin *et al.*, 2016). Similarly, a panel of 162 African rice cultivars revealed 31 genomic regions linked to blast resistance (Mgonja *et al.*, 2016). GWAS has proven effective in identifying genomic regions associated with disease resistance in plants and pathogenicity in microbes (Agrahari *et al.*, 2020). Dalman *et al.*, (2013) used GWAS to classify genetic components underlying virulence in *Heterobasidion annosum*, a fungal pathogen that damages conifers. Additionally, a GWAS involving 179 accessions of *A. thaliana* identified associations between plant genotypes and the abundance of specific microbes in root microbiomes (X. He *et al.*, 2021). These examples highlight some of the numerous studies conducted on the topic. However, as evident, most of them focus on interactions with pathogenic microorganisms. Analyses of the genetic bases involved in interactions between plants and non-pathogenic microorganisms remain more limited. This work can contribute to understanding the elements involved in the complex profiles that regulate plant-microbe interactions. Future expansions of this study could further enhance our knowledge in this field.

References

Agrahari, R. K., Singh, P., Koyama, H., & Panda, S. K. (2020). Plant-microbe Interactions for Sustainable Agriculture in the Postgenomic Era. *Current Genomics*, *21*(3), 168–178. https://doi.org/10.2174/1389202921999200505082116

Andreozzi, A., Prieto, P., Mercado-Blanco, J., Monaco, S., Zampieri, E., Romano, S., Valè, G., Defez, R., & Bianco, C. (2019). Efficient colonization of the endophytes Herbaspirillum huttiense RCA24 and Enterobacter cloacae RCA25 influences the physiological parameters of Oryza sativa L. cv. Baldo rice. *Environmental Microbiology*, *21*(9), 3489–3504. https://doi.org/10.1111/1462-2920.14688

Aubry, S. (2023). Genebanking plant genetic resources in the postgenomic era. *Agriculture and Human Values*, *40*(3), 961–971. https://doi.org/10.1007/s10460-023-10417-7

Becker, M., Patz, S., Becker, Y., Berger, B., Drungowski, M., Bunk, B., Overmann, J., Spröer, C., Reetz, J., Tchuisseu Tchakounte, G. V., & Ruppel, S. (2018). Comparative Genomics Reveal a Flagellar System, a Type VI Secretion System and Plant Growth-Promoting Gene Clusters Unique to the Endophytic Bacterium Kosakonia radicincitans. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.01997

Bianco, C., Andreozzi, A., Romano, S., Fagorzi, C., Cangioli, L., Prieto, P., Cisse, F., Niangado, O., Sidibé, A., Pianezze, S., Perini, M., Mengoni, A., & Defez, R. (2021). Endophytes from African Rice (Oryza glaberrima L.) Efficiently Colonize Asian Rice (Oryza sativa L.) Stimulating the Activity of Its Antioxidant Enzymes and Increasing the Content of Nitrogen, Carbon, and Chlorophyll. *Microorganisms*, *9*(8), 1714. https://doi.org/10.3390/microorganisms9081714

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brady, C., Cleenwerck, I., Venter, S., Coutinho, T., & De Vos, P. (2013). Taxonomic evaluation of the genus Enterobacter based on multilocus sequence analysis (MLSA): Proposal to reclassify E. nimipressuralis and E. amnigenus into Lelliottia gen. nov. as Lelliottia nimipressuralis comb. nov. and Lelliottia amnigena comb. nov., respectively, E. gergoviae and E. pyrinus into Pluralibacter gen. nov. as Pluralibacter gergoviae comb. nov. and Pluralibacter pyrinus comb. nov., respectively, E. cowanii, E. radicincitans, E. oryzae and E. arachidis into Kosakonia gen. nov. as Kosakonia cowanii comb. nov., Kosakonia radicincitans comb. nov., Kosakonia oryzae comb. nov. and Kosakonia arachidis comb. nov., respectively, and E. turicensis, E. helveticus and E. pulveris into Cronobacter as Cronobacter zurichensis nom. nov., Cronobacter helveticus comb. nov. and Cronobacter pulveris comb. nov., respectively, and emended description of the genera Enterobacter and Cronobacter. *Systematic and Applied Microbiology*, *36*(5), 309–319. https://doi.org/10.1016/j.syapm.2013.03.005

Camera, S. L., Balagué, C., Göbel, C., Geoffroy, P., Legrand, M., Feussner, I., Roby, D., & Heitz, T. (2009). The Arabidopsis Patatin-Like Protein 2 (PLP2) Plays an Essential Role in Cell Death Execution and Differentially Affects Biosynthesis of Oxylipins and Resistance to Pathogens. *Molecular Plant-Microbe Interactions®*, *22*(4), 469–481. https://doi.org/10.1094/MPMI-22-4-0469

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8

Chen, M., Zhu, B., Lin, L., Yang, L., Li, Y., & An, Q. (2014). Complete genome sequence of Kosakonia sacchari type strain SP1T. *Standards in Genomic Sciences*, *9*(3), 1311–1318. https://doi.org/10.4056/sigs.5779977

Choi, J. Y., Platts, A. E., Fuller, D. Q., Hsing, Y.-I., Wing, R. A., & Purugganan, M. D. (2017). The rice paradox: Multiple origins but single domestication in Asian rice. *Molecular Biology and Evolution*, msx049. https://doi.org/10.1093/molbev/msx049

Choi, J. Y., & Purugganan, M. D. (2018). Multiple Origin but Single Domestication Led to Oryza sativa. *G3 Genes|Genomes|Genetics*, *8*(3), 797–803. https://doi.org/10.1534/g3.117.300334

Dalman, K., Himmelstrand, K., Olson, Å., Lind, M., Brandström-Durling, M., & Stenlid, J. (2013). A Genome-Wide Association Study Identifies Genomic Regions for Virulence in the Non-Model Organism Heterobasidion annosum s.s. *PLOS ONE*, *8*(1), e53525. https://doi.org/10.1371/journal.pone.0053525

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382

ggplot2: Elegant Graphics for Data Analysis (3e). (s.d.). Recuperato 20 gennaio 2025, da https://ggplot2-book.org/

Ghazy, M. I., EL-Naem, S. A., Hefeina, A. G., Sallam, A., & Eltaher, S. (2024). Genome-Wide Association Study of Rice Diversity Panel Reveals New QTLs for Tolerance to Water Deficit Under the Egyptian Conditions. *Rice*, *17*, 29. https://doi.org/10.1186/s12284-024-00703-1

Giri, K. (2019). The First Report of Indigenous Free-Living Diazotroph Kosakonia sacchari Isolated from Himalayan Alder-Based Shifting Cultivation System in Nagaland, India. *Journal of Soil Science and Plant Nutrition*, *19*(3), 574–579. https://doi.org/10.1007/s42729-019-00056-5

Gutaker, R. M., Groen, S. C., Bellis, E. S., Choi, J. Y., Pires, I. S., Bocinsky, R. K., Slayton, E. R., Wilkins, O., Castillo, C. C., Negrão, S., Oliveira, M. M., Fuller, D. Q., Guedes, J. A. d'Alpoim, Lasky, J. R., & Purugganan, M. D. (2020). Genomic history and ecology of the geographic spread of rice. *Nature Plants*, *6*(5), 492–502. https://doi.org/10.1038/s41477-020-0659-6

Hardoim, P. R., van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., Döring, M., & Sessitsch, A. (2015). The Hidden World within Plants: Ecological and Evolutionary Considerations for Defining Functioning of Microbial Endophytes. *Microbiology and Molecular Biology Reviews*, *79*(3), 293–320. https://doi.org/10.1128/MMBR.00050-14

He, W., Chen, C., Xiang, K., Wang, J., Zheng, P., Tembrock, L. R., Jin, D., & Wu, Z. (2021). The History and Diversity of Rice Domestication as Resolved From 1464 Complete Plastid Genomes. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.781793

He, X., Zhang, Q., Li, B., Jin, Y., Jiang, L., & Wu, R. (2021). Network mapping of root–microbe interactions in Arabidopsis thaliana. *Npj Biofilms and Microbiomes*, *7*(1), 72. https://doi.org/10.1038/s41522-021-00241-4

Huang, M., Liu, X., Zhou, Y., Summers, R. M., & Zhang, Z. (2019). BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience*, *8*(2), giy154. https://doi.org/10.1093/gigascience/giy154

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics (Oxford, England)*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Kämpfer, P., McInroy, J. A., Doijad, S., Chakraborty, T., & Glaeser, S. P. (2016). Kosakonia pseudosacchari sp. Nov., an endophyte of Zea mays. *Systematic and Applied Microbiology*, *39*(1), 1–7. https://doi.org/10.1016/j.syapm.2015.09.004

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLOS Genetics*, *12*(2), e1005767. https://doi.org/10.1371/journal.pgen.1005767

Manual of Seed Handling in Genebanks. (2006). Earthprint.

Mgonja, E. M., Balimponya, E. G., Kang, H., Bellizzi, M., Park, C. H., Li, Y., Mabagala, R., Sneller, C., Correll, J., Opiyo, S., Talbot, N. J., Mitchell, T., & Wang, G.-L. (2016). Genome-Wide Association Mapping of Rice Resistance Genes Against Magnaporthe oryzae Isolates from Four African Countries. *Phytopathology®*, *106*(11), 1359–1365. https://doi.org/10.1094/PHYTO-01-16-0028-R

Muthayya, S., Sugimoto, J. D., Montgomery, S., & Maberly, G. F. (2014). An overview of global rice production, supply, trade, and consumption. *Annals of the New York Academy of Sciences*, *1324*(1), 7–14. https://doi.org/10.1111/nyas.12540

Pan, Y.-H., Chen, L., Guo, H.-F., Feng, R., Lou, Q.-J., Rashid, M. A. R., Zhu, X.-Y., Qing, D.-J., Liang, H.-F., Gao, L.-J., Huang, C.-C., Zhao, Y., & Deng, G.-F. (2022). Systematic Analysis of NB-ARC Gene Family in Rice and Functional Characterization of GNP12. *Frontiers in Genetics*, *13*. https://doi.org/10.3389/fgene.2022.887217

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, *81*(3), 559–575.

Raboin, L.-M., Ballini, E., Tharreau, D., Ramanantsoanirina, A., Frouin, J., Courtois, B., & Ahmadi, N. (2016). Association mapping of resistance to rice blast in upland field conditions. *Rice*, *9*(1), 59. https://doi.org/10.1186/s12284-016-0131-4

Rodriguez, R., & Durán, P. (2020). Natural Holobiome Engineering by Using Native Extreme Microbiome to Counteract the Climate Change Effects. *Frontiers in Bioengineering and Biotechnology*, *8*, 568. https://doi.org/10.3389/fbioe.2020.00568

Rydel, T. J., Williams, J. M., Krieger, E., Moshiri, F., Stallings, W. C., Brown, S. M., Pershing, J. C., Purcell, J. P., & Alibhai, M. F. (2003). The Crystal Structure, Mutagenesis, and Activity Studies Reveal that Patatin Is a Lipid Acyl Hydrolase with a Ser-Asp Catalytic Dyad. *Biochemistry*, *42*(22), 6696–6708. https://doi.org/10.1021/bi027156r

Saini, M. R., Chandran, L. P., Barbadikar, K. M., Sevanthi, A. M. V., Chawla, G., Kaushik, M., Mulani, E., Phule, A. S., Govindannagari, R., Sonth, B., Sinha, S. K., Sundaram, R. M., & Mandal, P. K. (2022). Understanding plant–microbe interaction of rice and soybean with two contrasting diazotrophic bacteria through comparative transcriptome analysis. *Frontiers in Plant Science*, *13*, 939395. https://doi.org/10.3389/fpls.2022.939395

Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., Wakimoto, H., Yang, C., Iwamoto, M., Abe, T., Yamada, Y., Muto, A., Inokuchi, H., Ikemura, T., Matsumoto, T., Sasaki, T., & Itoh, T. (2013). Rice Annotation Project Database (RAP-DB): An integrative and interactive database for rice genomics. *Plant & Cell Physiology*, *54*(2), e6. https://doi.org/10.1093/pcp/pcs183

Shinjo, R., Uesaka, K., Ihara, K., Loshakova, K., Mizuno, Y., Yano, K., & Tanaka, A. (2016). Complete Genome Sequence of Kosakonia sacchari Strain BO-1, an Endophytic Diazotroph Isolated from a Sweet Potato. *Genome Announcements*, *4*(5), e00868-16. https://doi.org/10.1128/genomeA.00868-16

Siddika, A., Rashid, A. A., Khan, S. N., Khatun, A., Karim, M. M., Prasad, P. V. V., & Hasanuzzaman, M. (2024). Harnessing plant growth-promoting rhizobacteria, Bacillus subtilis and B. aryabhattai to combat salt stress in rice: A study on the regulation of antioxidant defense, ion homeostasis, and photosynthetic parameters. *Frontiers in Plant Science*, *15*. https://doi.org/10.3389/fpls.2024.1419764

Takken, F. L., Albrecht, M., & Tameling, W. I. (2006). Resistance proteins: Molecular switches of plant defence. *Current Opinion in Plant Biology*, *9*(4), 383–390. https://doi.org/10.1016/j.pbi.2006.05.009

van Ooijen, G., Mayr, G., Kasiem, M. M. A., Albrecht, M., Cornelissen, B. J. C., & Takken, F. L. W. (2008). Structure-function analysis of the NB-ARC domain of plant disease resistance proteins. *Journal of Experimental Botany*, *59*(6), 1383–1397. https://doi.org/10.1093/jxb/ern045

Wang, J., & Zhang, Z. (2021). GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics, Proteomics & Bioinformatics*, *19*(4), 629–640. https://doi.org/10.1016/j.gpb.2021.08.005

Wang, Q., Tian, F., Pan, Y., Buckler, E. S., & Zhang, Z. (2014). A SUPER Powerful Method for Genome Wide Association Study. *PLOS ONE*, *9*(9), e107684. https://doi.org/10.1371/journal.pone.0107684

Wickham, H., & Bryan, J. (2015). *readxl: Read Excel Files* (p. 1.4.3) [Dataset]. https://doi.org/10.32614/CRAN.package.readxl

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., & Liu, X. (2021). rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. *Genomics, Proteomics & Bioinformatics*, *19*(4), 619–628. https://doi.org/10.1016/j.gpb.2020.10.007

Zeigler, R. S., & Barclay, A. (2008). The Relevance of Rice. *Rice*, *1*(1), Articolo 1. https://doi.org/10.1007/s12284-008-9001-z

Supplementary material

8.1 R script of the analysis

Rice - RCA25 inoculation: Statistical analysis

Introduction

This script is compiled as a R Quarto Project. For this reason I have not defined a Working Directory (for more info about R Quarto Projects: https://quarto.org/docs/projects/quarto-projects.html).

1. Preparation of the work area

1.1 Setting the default R code chunks for the script

The following script is divided in different paragraphs, and the relative code blocks, for the different steps performed in the analysis of the data.

1.2 Installation of the required packages

The first thing to do to start our analysis is to install the required packages:

```
# Installing Librarian package:
  utils::install.packages("librarian") # "librarian" package allows to
activate already installed packages or, if they are not, to install and
activate them.

# Installing Bioconductor Manager:
  if (!require("BiocManager",
               quietly = TRUE))
    install.packages("BiocManager")
  BiocManager::install(version = "3.20") # Bioconductor is a package manager
that maintains some of the packages required for the analysis

# Installing GAPIT:
  install.packages("remotes")
  remotes::install_github("jiabowang/GAPIT")
  library(GAPIT) # this set of commands allows to install GAPIT package, used
to perform GWAS analyses

# Installing and activating required packages:
  librarian::shelf(c(
      "adegenet", # to convert data for the PCA
      "ape", # to perform phylogenetic analyses
      "ASRgenomics", # set of functions for geno data analyses
      "car", # for statistical and plotting functions
      "data.table", # to manage datasets
      "dplyr", # to extend data manipulation capabilities
```

```
      "ggfortify", # adds a set of plotting tools
      "ggplot2", # to produce plots and figures
      "ggpubr", # to increase some of the capabilities of ggplot2 package
      "ggrepel", # to avoid elements in a plot to overlap
      "ggthemes", # predefined aestetics for plots made with ggplot2
      "ggtree", # to produce the NJ-tree
      "gplots", # various commands to increase plotting capabilities
      "Hmisc", # for statistics and data manipulation
      "LEA", # to study the population structure
      "paletteer", # collection of color palettes
      "pastecs", # package for descriptive statistics
      "RColorBrewer", # to add color palettes
      "readxl", # to open .xlsx files
      "rMVP", # to perform the GWAS with MVP
      "rstatix", # tools to increase statistical capabilities
      "snpStats", # a set of functions to study SNPs markers
      "stats", # for some statistical analyses
      "tidyverse", # for a better data management
      "vcfR" # to manage .vcf files
  ))
```

2. Loading and preparation of data

Now it is time to load the data. **Four datasets** will be loaded:

1. `cfu_draft`: this is the raw dataset, containing all the **information about rice genotypes tested** through the inoculation of *Kosakonia sacchari* RCA25 bacterium: number of replicates, roots weight, CFUs counted, CFUs corrected for the dilution factor. This dataset will be associated with the `cfu_csv` object.
2. `geno`: the genotypic information about the rice panel. This in particular is a file which was already filtered (SNPs call rate = 90%, MAF (Minor Allele Frequency) = 5%, individuals call rate = 90%) with Plink2 software (https://www.cog-genomics.org/plink/2.0/), and imputed with Beagle software (http://faculty.washington.edu/browning/beagle/beagle.html). It comprehends 281 taxa and 62421 SNPs.
3. `geno_pruned`: the same genotypic dataset indicated above, but with the addition of the pruning step to remove redundant SNPs in LD (Linkage Disequilibrium). After pruning, the number of markers is 14220.
4. `risinnova_origins`: this is the dataset that contains **information about the origins of the rice varieties** composing the Risinnova collection. This dataset will be associated with the `risinnova_origins` object.

```
# Loading phenotypic (i.e. CFUs) data:
  cfu_draft <-
readxl::read_excel("Phenotype_data/Piastre_Risinnova_2023.xlsx",
                           sheet = 1) # this is the   dataset regarding CFUs
data
```

```
# Loading genotypic data:
  geno <- read.table("Genotype_data/hapmap/Risinnova_filtered_imputed_2024-
12-30.hmp.txt",
                     header = T,
                     row.names = 1)

# Loading genotypic PRUNED data:
  geno_pruned <-
read.table("Genotype_data/hapmap/Risinnova_filtered_pruned_imputed_2024-12-
30.hmp.txt",
                           header = T,
                           row.names = 1)

# Loading information about origins for our rice varieties:
  risinnova_origins <-
readxl::read_excel("Origins&PopStructure/Risinnova_countries_data.xlsx",
                                   sheet = 1) # this dataset will be used to
prepare a piechart showing the relative proportions of the different
countries for rice varieties composing the Risinnova collection

  head(cfu_draft)
```

After loading the data, I will prepare a **new subset containing only the name of the varieties and CFUs/roots ratio** (this will be the `vars_and_CFUs` object):

```
# Keeping only the columns with the necessary information:
  vars_and_CFUs_draft <- as.data.frame(cfu_draft[,c(1,2,9)])

# Removing NAs:
  vars_and_CFUs <- na.omit(vars_and_CFUs_draft) # cleaned data.frame with
rice varieties, number of replica and CFUs

# Rounding the CFUs values:
    vars_and_CFUs <- vars_and_CFUs %>%
    mutate(CFUsRoots_Ratio = round(CFUsRoots_Ratio, digits = 0))

# Changing the columns names of the resulting phenotypic dataset:
  colnames(vars_and_CFUs) <- c("Taxa",
                               "Replica",
                               "CFUs")

# The "Escarlate" genotype shows inoculation values of 0 for each of the
replicates. It is not possible to say for certain whether this is due to the
variety's resistance to interaction with the microorganism, but it can be
reasonably assumed that an error in the inoculation process is more likely.
Therefore, this genotype will be removed:
  vars_and_CFUs2 <- vars_and_CFUs[-c(130:132),]

# Resetting the row numbers:
```

```
  row.names(vars_and_CFUs2) <- NULL

# Converting all the genotypes names in all uppercase:
  vars_and_CFUs2$Taxa <- toupper(vars_and_CFUs2$Taxa)

# Some genotypes contain spaces inside their names. This is not advisable, so
these will be substituted with underscores ("_"):
 vars_and_CFUs2$Taxa <- gsub(" ", "_", vars_and_CFUs2$Taxa)
```

Then, I will **merge the 3 replicates available for each genotype**, in order **to obtain mean values**. The new object will be `vars_and_CFUs_means`. Then, I will briefly summarize some **basic info about the new dataset**, using `summary()` and `Hmisc::describe()` functions:

```
# Calculating the mean values of CFUs from the replicates:
  vars_and_CFUs_means <- aggregate(CFUs ~ Taxa,
                                   vars_and_CFUs2,
                                   FUN = mean)

# Rounding the CFUs values:
    vars_and_CFUs_means <- vars_and_CFUs_means %>%
    mutate(CFUs = round(CFUs, digits = 0))

# Changing the columns names in the new dataset:
  colnames(vars_and_CFUs_means) <- c('Taxa', 'Mean_CFUs')

  head(vars_and_CFUs_means)

# Briefly showing info about the new dataset:
  Hmisc::describe(vars_and_CFUs_means) # this function provide the number of
variables and observations in the dataset, n missing, mean,
5,10,25,50,75,90,95th percentiles, 5 lowest and 5 highest scores
```

3. Data analysis

3.1 Data distribution

Now it is time to produce the first plot. This will be an **histogram to visualize the distribution of CFUs** data, associated with the `vars_and_CFUs_means` object. Then, it will be possible to visualize the same thing but using a **density plot**:

```
# Histogram for the distribution of mean values:
  CFUs_mean_values_hist <- ggplot(vars_and_CFUs_means,
                                  aes(x = Mean_CFUs)) +
  geom_histogram(fill="#69b3a2",
                 color="#e9ecef") +
  theme(axis.text.x = element_text(size = 6,
                                   angle = 45,
```

```
                                    hjust = 1)) +
  theme_bw() +
  xlab("CFUs") +
  ylab("Rice varieties") +
  labs(title = "Histogram of CFUs mean values in tested rice varieties") +
  theme(plot.title = element_text(hjust = 0.5,
                                  size = 22),
        axis.title.x = element_text(size = 14),
        axis.title.y = element_text(size = 14))

# Saving the mean values histogram image:
  ggsave("Outputs/Figures/CFUs_mean_values_histogram.png",
         CFUs_mean_values_hist,
         width = 10,
         height = 6)

# Representing the density plot for mean values:
  CFUs_density_plot <- ggplot(vars_and_CFUs_means) +
  aes(x = Mean_CFUs) +
  geom_density() +
  labs(title = "Density plot of CFUs") +
  theme_bw()

# Saving the figure of the density plot:
  ggsave("Outputs/Figures/CFUs_density_plot.png",
           CFUs_density_plot,
           width = 10,
           height = 6)
```

3.2 Distribution of means for each variety

With the following code chunk, I will prepare a second plot, showing the distribution of CFUs mean values recorded for each genotype tested:

```
# Plotting mean values as a ordered barplot:
  CFUs_mean_values_barplot <- ggplot(vars_and_CFUs_means,
                                     aes(x = reorder(Taxa, Mean_CFUs),
                                         y = Mean_CFUs)) +
  geom_col(fill = "#0AB0E5",
           color="#e9ecef") +
  theme(axis.text.x = element_text(size = 6,
                                   angle = 45,
                                   hjust = 1)) +
  xlab("Rice varieties") +
  ylab("CFUs (mean values)") +
  labs(title = "Barplot of RCA25 mean infection rates in tested rice
varieties") +
  theme(plot.title = element_text(hjust = 0.5,
                                  size = 22),
        axis.title.x = element_text(size = 14),
```

```
        axis.title.y = element_text(size = 14)) +
  theme_bw()

# Saving the plot as a figure:
  ggsave("Outputs/Figures/CFUs_mean_values_barplot.png",
         CFUs_mean_values_barplot,
         width = 10,
         height = 6)
```

3.3 QQ-plot of CFUs Mean values

Quantile-Quantile plots (QQ-plots) can be useful to visualize the distribution of data, and if this is normal or not. If points are close to the reference line (sometimes referred as Henry's line), and within the confidence bands, the normality assumption can be considered as met. The bigger the deviation between the points and the reference line, and the more they lie outside the confidence bands, the less likely data show a normal distribution[1]:

```
# QQ-plot of CFUs inoculation rates:
  CFUs_QQplot <- ggqqplot(vars_and_CFUs_means$Mean_CFUs) +
    geom_qq_line(color = "blue") +
    theme_bw() +
    theme(axis.text.x = element_text(size = 6,
                                     angle = 45,
                                     hjust = 1)) +
    labs(title = "QQplot of CFUs mean values") +
    theme(plot.title = element_text(hjust = 0.5,
                                    size = 22),
          axis.title.x = element_text(size = 14),
          axis.title.y = element_text(size = 14))

# Saving the plot:
  ggsave("Outputs/Figures/CFUs_QQplot.png",
         CFUs_QQplot,
         width = 10,
         height = 6)
```

3.4 Data normalization

The histogram showed that the data are not normally distributed, and also in the QQ-plot many points do not lie inside the confidence bands. To perform analysis like ANOVA and GWAS, however, a normal distribution is an important aspect, especially in case of not very big amount of data.

When facing a non-normal distribution, the first step is usually to apply a transformation step on the data. This can be performed through different methods (e.g. logarithm or

---

[1] https://statsandr.com/blog/descriptive-statistics-in-r/#for-a-single-variable

square root transformation). In the next code block, the replicates values will be transformed through the square root method to normalize them:

```
# Transforming of replicates values:
  vars_and_CFUs2$sqrt_CFUs <- sqrt(vars_and_CFUs2$CFUs)

# Rounding the transformed values:
    vars_and_CFUs2 <- vars_and_CFUs2 %>%
    mutate(sqrt_CFUs = round(sqrt_CFUs, digits = 0))

# Removing the unnecessary CFUs column:
  vars_and_CFUs3 <- vars_and_CFUs2[,-3]

# Updating column names:
  colnames(vars_and_CFUs3) <- c("Taxa", "Replicate", "Transformed_CFUs")

# Calculating means of transformed values for each genotype:
  vars_and_CFUs_transformed_means <- aggregate(Transformed_CFUs ~ Taxa,
                                               vars_and_CFUs3, FUN = mean)

# Rounding the transformed mean values:
    vars_and_CFUs_transformed_means <- vars_and_CFUs_transformed_means %>%
    mutate(Transformed_CFUs = round(Transformed_CFUs, digits = 0))

# Changing the columns names in the new dataset:
  colnames(vars_and_CFUs_transformed_means) <- c('Taxa', 'Transformed_means')
  head(vars_and_CFUs_transformed_means)
```

3.5 Histogram and QQ-plot after transformation

After the square root transformation, the data should show a normal distribution. To verify this, I will now plot the histogram and QQ-plot of the means obtained from transformed CFUs data, followed by a Shapiro-Wilk test to confirm the normality of the data distribution:

```
# Plotting histogram of transformed mean values:
  CFUs_mean_norm_hist <- ggplot(vars_and_CFUs_transformed_means,
                                aes(x = Transformed_means)) +
  geom_histogram(fill="#69b3a2",
                 color="#e9ecef") +
  theme(axis.text.x = element_text(size = 6,
                                   angle = 45,
                                   hjust = 1)) +
  theme_bw() +
  xlab("CFUs") +
  ylab("Rice genotypes") +
  labs(title = "Histogram of CFUs - transformed mean values") +
  theme(plot.title = element_text(hjust = 0.5,
                                  size = 22),
        axis.title.x = element_text(size = 14),
```

```
        axis.title.y = element_text(size = 14))

# Saving the mean values histogram image:
  ggsave("Outputs/Figures/CFUs_transformedMeans_histogram.png",
         CFUs_mean_norm_hist,
         width = 10,
         height = 6)

# Plotting QQ-plot of CFUs inoculation rates after transformation:
  CFUs_norm_QQplot <-
ggqqplot(vars_and_CFUs_transformed_means$Transformed_means) +
    geom_qq_line(color = "blue") +
    theme_bw() +
    theme(axis.text.x = element_text(size = 6,
                                     angle = 45,
                                     hjust = 1)) +
    labs(title = "QQplot of CFUs - transformed mean values") +
    theme(plot.title = element_text(hjust = 0.5,
                                    size = 22),
          axis.title.x = element_text(size = 14),
          axis.title.y = element_text(size = 14))

# Saving the plot figure:
    ggsave("Outputs/Figures/CFUs_transformed_QQplot.png",
           CFUs_norm_QQplot,
           width = 10,
           height = 6)

# Performing Shapiro-Wilk test:
    shapiro.test(vars_and_CFUs_transformed_means$Transformed_means) #
Shapiro-Wilk test for the distribution of data. H0 assumes a normal
distribution
```

3.6 Boxplots of transformed replicates

The next graph will be a **distribution of boxplots**, each one representing the transformed results of the replicates for each tested genotype. It is important to highlight that **not for all the genotypes tested 3 replicates are available**: in some cases, bad plates results were obtained in lab. These were indicated as NAs (Not Available), and removed during the data upload. Besides, other tests were removed in previous QC steps.

Tukey's fences will be applied during the definition of the plot to remove possible outliers.

The boxplots **will be organized** in a way that shows the distribution of the rice varieties **from the one with the lowest infection rate to the most infected one**:

```
# Tukey's fences for removing outliers:
  fences <- vars_and_CFUs3 %>%
    group_by(Taxa) %>%
    summarise(lower_fence = quantile(Transformed_CFUs, 0.25) - 1.5 *
```

```
IQR(Transformed_CFUs),
              upper_fence = quantile(Transformed_CFUs, 0.75) + 1.5 *
IQR(Transformed_CFUs))

  filtered_data <- vars_and_CFUs3 %>%
    left_join(fences, by = "Taxa") %>%
    filter(Transformed_CFUs >= lower_fence & Transformed_CFUs <= upper_fence)

# Plotting boxplots without outliers (no differences):
  # Organizing the boxplots by infection rates amount:
  CFUs_boxplots_ordered <- ggplot(filtered_data,
                                  aes(x = reorder(Taxa, Transformed_CFUs),
                                      y = Transformed_CFUs)) +
  geom_boxplot(fill = "#19DE7F",
               outlier.colour = "red") +
  theme_bw() +
  theme(axis.text.x = element_text(size = 6,
                                   angle = 45,
                                   hjust = 1)) +
  ggtitle("Boxplot of RCA25 infection rates for each genotype - Transformed
values") +
  labs(x = "Rice varieties",
       y = "Infection rate") +
  theme(plot.title = element_text(hjust = 0.5,
                                  size = 21),
        axis.title.x = element_text(size = 16),
        axis.title.y = element_text(size = 16))

# Saving the plot image:
  ggsave("Outputs/Figures/Transformed_CFUs_boxplots_ordered.png",
         CFUs_boxplots_ordered,
         width = 10,
         height = 6)
```

3.7 Origins pie chart

It is time to evaluate the origins of the rice varieties. I will use a **piechart to visualize the distribution of the rice varieties tested**, dividing the slices by countries:

```
# Selecting the data about the origin:
  rice_countries <- risinnova_origins %>% dplyr::select(Variety, Origin) # a
df with only the columns for the varieties and origins is produced

# Grouping the varieties by countries:
  rice_countries_groups <- rice_countries %>%
    group_by(Origin) %>%
    summarise(n = n()) # here the different rice varieties are grouped by the
origin

# Calculating percentages:
```

```r
  percent_df <- rice_countries_groups %>%
    mutate(percent = round(n / sum(n)*100, 1))  # Create the pie chart

  pie(rice_countries_groups$n,
      labels = percent_df$Origin,
      border = "white",
      col = paletteer_d("ggthemes::Classic_Cyclic"))

  df2 <- rice_countries_groups %>%
    mutate(csum = rev(cumsum(rev(n))),
           pos = n/2 + lead(csum, 1),
           pos = if_else(is.na(pos), n/2, pos)) # Getting the positions for
labels in the piechart

  df2$percentages <- round(df2$n / sum(df2$n) * 100, 1)

  colors <- c("#1f77b4",
              "#ff7f0e",
              "#2ca02c",
              "#d62728",
              "#9467bd",
              "#008080",
              "#e377c2",
              "#7f7f7f",
              "#bcbd22",
              "#17becf",
              "#aec7e8",
              "#ffbb78",
              "#98df8a",
              "#ff9896",
              "#c5b0d5",
              "#c49c94",
              "#f7b6d2",
              "#b39eb5",
              "#dbdb8d",
              "#fb5858",
              "#266d57",
              "#9db2fb")

# Preparing the piechart showing the origins composition of the rice
varieties:
  origins_piechart <- ggplot(rice_countries_groups, aes(x="", y=n,
fill=Origin)) +
    geom_col(width = 1, color = 1) +
    coord_polar(theta = "y") +
    geom_label_repel(data = df2,
                     aes(y = pos,
                         label = paste0(n, " ", "(", percentages, "%", ")")),
                     size = 4.5,
```

```
                      nudge_x = 1,
                      show.legend = FALSE) +
    guides(fill = guide_legend(title = "Country of origin")) +
    theme_void() +
    scale_fill_manual(values = colors) +
    theme(legend.title = element_text(size = 12,
                                      hjust = 0.5))  # Increase legend title
font size

# Saving the origins piechart:
  ggsave("Outputs/Figures/origins_piechart.png",
         origins_piechart,
         width = 10,
         height = 6)
```

4. Statistical analysis

In the following code block, summary statistics of values in the `vars_and_CFUs_means2` dataset will be calculated:

```
# Calculating some basic statistical parameters:
  var(vars_and_CFUs_means$Mean_CFUs)
  sd(vars_and_CFUs_means$Mean_CFUs)
  summary(vars_and_CFUs_means) # this function gives the main descriptive
statistical values: range (min and max values), 1st and 3rd quantiles,
median, and mean.

  stats_table <- stat.desc(vars_and_CFUs_means)
  print(stats_table)
```

Parameter	Value
Min	14493
Max	8131811
Mean	2216069
Median	1721052
1° quantile	812210
3° quantile	3229018
SD (standard Deviation)	1793215

4.1 Analysis of variance

I will perform a One-way ANOVA test to look if there are significant differences between the results of different genotypes. In case of rejection of the null hypothesis H_0 (p-values < 0.05), significant differences among the group means should be present:

```
# Performing ANOVA test:
  anova <- aov(Transformed_CFUs ~ Taxa,
               data = vars_and_CFUs3)
  summary(anova)

# Performing TukeyHSD test:
  TukeyHSD(anova)
```

Df: Degrees of freedom

Sum sq: Sums of squares

Mean Sq: Mean squares

---

5. Population structure

In the following paragraphs, I will start the analysis of the population structure on the rice panel. Multiple approaches can be used to study the population structure, the most common are the Principal Component Analysis (PCA), the Neighbor-Joining (NJ) tree, and the ancestry matrix.

5.1 Principal Component Analysis

The PCA gives an overview of the population structure through the calculation of Principal Components (PCs), that describe the dataset dividing values into groups based on the variance explained. GAPIT package, which will be used to perform the GWAS, produce a PCA plot, but for the entire Risinnova collection. So, I will plot all the genotypes of the collection, highlighting in red the accessions used in the experiment:

```
# Loading the geno file in VCF format:
  geno_vcf <- read.vcfR("Genotype_data/VCF/Risinnova_filtered_imputed_2024-
12-30.vcf")

# Converting VCF file in genlight object for the PCA:
  geno_genlight <- vcfR2genlight(geno_vcf)

# Executing the PCA:
  pca_result <- glPca(geno_genlight, nf = 3)  # nf = number of PCs

# Extracting PCs scores:
  pca_scores <- pca_result$scores

# Producing a dataframe with values of the first PCs:
  pca_df <- data.frame(PC1 = pca_scores[, 1], PC2 = pca_scores[, 2])

# Visualizing the PCA scatterplot:
ggplot(pca_df, aes(x = PC1, y = PC2)) +
```

```
  geom_point(size = 3, alpha = 0.6) +
  theme_minimal() +
  labs(title = "PCA of Genotypic Data (VCF)",
       x = "Principal Component 1",
       y = "Principal Component 2")

# Selecting only tested genotypes:
  geno_tested <- vars_and_CFUs_transformed_means$Taxa

# Adding a column "Color" to pca_df:
  pca_df$Color <- ifelse(rownames(pca_df) %in% geno_tested, "Tested", "Not
tested")

# Producing the plot with "tested" and "not tested" genotypes:
  pca_tested <- ggplot(pca_df, aes(x = PC1, y = PC2, color = Color)) +
    geom_point(size = 3, alpha = 0.6) +
    scale_color_manual(values = c("Tested" = "red", "Not tested" = "gray")) +
    theme_minimal() +
    labs(title = "PCA of Genotypic Data",
         x = "Principal Component 1",
         y = "Principal Component 2",
         color = "Genotypes") +
    theme(legend.position = "bottom") +
    theme(plot.title=element_text(hjust=0.5))

# Saving the PCA scatterplot:
  ggsave("Outputs/Figures/PCA_tested.png",
         pca_tested,
         width = 10,
         height = 6)
```

5.2 Neighbor-Joining Tree

Here I define the NJ Tree of the genotypes composing the Risinnova collection, highlighting the accessions used in the test, as done for the PCA in the previous paragraph:

```
# Extracting genotypes matrix (0, 1, 2 for homozigous and heterozygous):
  geno_matrix <- as.matrix(geno_genlight)

# Calculating Hamming distance (allelic differences):
  dist_matrix <- dist(geno_matrix,
                      method = "manhattan") / (2 * ncol(geno_matrix))

# Producing the nj-tree:
  nj_tree <- nj(dist_matrix)

# Visualizing the tree:
  plot(nj_tree,
       type = "unrooted",
```

```
       main = "Neighbor-Joining Tree")

# Defining the genotypes to highlight inside the tree:
  nj_tree$tip.label.highlight <- ifelse(nj_tree$tip.label %in% geno_tested,
"Tested", "Not tested")

# Creating a data frame with data about tested genotypes:
  highlight_data <- data.frame(
    label = nj_tree$tip.label,  # Nomi dei genotipi
    highlight = nj_tree$tip.label.highlight  # Colonna di evidenziazione
  )

# Adding the data to the plot:
  nj_tree <- ggtree(nj_tree, layout = "rectangular") %<+% highlight_data +
    geom_tiplab(
      aes(color = highlight, size = highlight),
      fontface = "bold"
    ) +
    scale_color_manual(values = c("Tested" = "red", "Not tested" =
"#5b5b5b")) +
    scale_size_manual(values = c("Tested" = 2, "Not tested" = 1.5)) +
    theme_tree() +
    theme(plot.title=element_text(hjust=0.5)) +
    labs(
      title = "Phylogenetic Tree of the Risinnova collection",
      color = "Condition"
    ) +
    guides(
      color = guide_legend(override.aes = list(shape = 1)),
      size = "none"
    )

# Saving the NJ tree:
  ggsave("Outputs/Figures/NJ_tree_rectangular.png",
         nj_tree,
         width = 12,
         height = 20)
```

5.3 Population structure - Ancestry matrix

The analysis of the population structure will be performed with the *LEA* package (https://www.bioconductor.org/packages/release/bioc/html/LEA.html). In this first block I will convert the genotypic data in the format required by LEA:

```
# Converting the .vcf PRUNED genotypic file in .geno file and loading the
latter:
  vcf2geno("Genotype_data/VCF/Risinnova_filtered_pruned_imputed_2024-12-
30.vcf", "Genotype_data/Risinnova_filtered_pruned_imputed.geno")
  geno_geno <-
read.geno("Genotype_data/Risinnova_filtered_pruned_imputed.geno")
```

In the next step, I will produce a .lfmm file. LFMM (Latent Factor Mixed Models) are factor regression models in which the response variable is a genotypic matrix, while the explanatory variables are environmental measures of ecological interest or trait values [@10.18129/B9.bioc.LEA]:

```
write.lfmm(geno_geno, "Origins&PopStructure/genotypes.lfmm")
```

The *LEA* package allows to infer the number of components that explain the most of the variance. A good rule is to count the points before the "knee":

```
# Making and saving the PCA:
  pc = pca("Origins&PopStructure/genotypes.lfmm", scale = TRUE)
  tw = tracy.widom(pc)
  plot(tw$percentage, pch = 19, col = "darkblue", cex = .8)
```

In the following code chunk, the main part of the population structure analysis with *LEA* will be performed. The first part will consists in defining a project, based on the number of clusters (Ks), *i.e.* the hypothesized number of sub-populations, chosen in the previous block. *LEA* will compute the selected number of Ks, and the relative runs, or best hypothesis, for each. These results will then be plotted to choose some of the most promising Ks, with the relative best runs provided by *LEA*.

```
# Calculating the population structure:
 pops_project <- snmf("Genotype_data/Risinnova_filtered_pruned_imputed.geno",
                      K = 1:30,
                      entropy = TRUE,
                      repetitions = 10,
                      project = "new")

# Loading the project:
  geno_project <- load.snmfProject("/Users/lorenzo/Nextcloud/Rice-
Endophytes/Origins&PopStructure/Risinnova_filtered_pruned_imputed.snmfProject
")

# Plotting the results of snmf():
  ancestral_pops <- plot(geno_project,
                         col = "blue",
                         pch = 19,
                         cex = 1.2)

# Preparing a colors palette for the population plot:
  my.colors <- c(
    "tomato",
    "lightblue",
    "olivedrab",
    "blue",
    "pink",
    "orange",
    "grey",
    "green",
```

```
    "#5572ad",
    "red",
    "darkblue",
    "#e815a8",
    "#a2e815",
    "#13bbb4",
    "#5b50ae",
    "#b7571c",
    "#4CAF50",
    "#2196F3",
    "#FF5722"
    )

# Choosing the best run:
  cross_entropy <- cross.entropy(geno_project,
                                 K = 19)
  best_run <- which.min(cross_entropy)
  print(best_run)
```

To test different hypotheses, I will write below a **table of interesting Ks with the relative best run**:

K	BEST RUN
6	4
9	4
16	4
19	4

Now that we have defined the most promising Ks and runs, the next step consists in plotting these as barcharts, to visualize the proposed subdivisions of the clusters.

K = 6, run = 4:

```
# Plotting the population structure, K=6, run=4, sorting by Q:
barchart(geno_project,
         K = 6,
         run = 4,
         border = NA,
         space = 0,
         col = my.colors,
         sort.by.Q = T,
         xlab = "Individuals",
         ylab = "Ancestry proportions",
         main = "Ancestry matrix",
         cex.names = 1.8,
         cex.axis = 1.8,
         cex.lab = 1.8,
         cex.main = 1.8)
```

K = 9, run = 4:

```
# Plotting the population structure, K=9, run=4, sorting by Q:
  barchart(geno_project,
         K = 9,
         run = 4,
         border = NA,
         space = 0,
         col = my.colors,
         sort.by.Q = T,
         xlab = "Individuals",
         ylab = "Ancestry proportions",
         main = "Ancestry matrix",
         cex.names = 1.8,
         cex.axis = 1.8,
         cex.lab = 1.8,
         cex.main = 1.8)
```

K = 16, run = 4:

```
# Plotting the population structure, K=16, run=4, sorting by Q:
  barchart(geno_project,
           K = 16,
           run = 4,
           border = NA,
           space = 0,
           col = my.colors,
           sort.by.Q = T,
           xlab = "Individuals",
           ylab = "Ancestry proportions",
           main = "Ancestry matrix",
           cex.names = 1.8,
           cex.axis = 1.8,
           cex.lab = 1.8,
           cex.main = 1.8)
```

K = 19, run = 4:

```
# Plotting the population structure, K=19, run=4, sorting by Q:
  barchart(geno_project,
           K = 19,
           run = 4,
           border = NA,
           space = 0,
           col = my.colors,
           sort.by.Q = T,
           xlab = "Individuals",
           ylab = "Ancestry proportions",
           main = "Ancestry matrix",
           cex.names = 1.8,
           cex.axis = 1.8,
```

```
            cex.lab = 1.8,
            cex.main = 1.8)
```

6. GWAS

We now have all the elements required to perform association studies through GWAS. Different tests will be conducted using the same datasets but with different packages. The model chosen for all the tests is FarmCPU (Fixed and Random Model Circulating Probability Unification), one of the most robust methods available. It is particularly suited for relatively small datasets, as it is not the most computationally efficient option.

6.1 GWAS with rMVP package

In the next code block, I will start the **first Genome-Wide Association** test, using the R package `rMVP` (https://github.com/xiaolei-lab/rMVP):

```
# rMVP requires the data to be implemented as a file, so I will save the
final dataset about pheno data in a .txt file:
  write.table(vars_and_CFUs_transformed_means,
"Phenotype_data/vars_and_CFUs_transformed_means.txt", sep = "\t")

## rMVP requires the hapmap version with "#" to "rs" and "assembly" columns:
  MVP.Data(fileHMP="Genotype_data/hapmap/Risinnova_filtered_imputed_2024-12-
30.hmp_hashtag.txt",
           filePhe="Phenotype_data/vars_and_CFUs_transformed_means.txt",
           #sep.hmp="\t",
           sep.phe="\t",
           #SNP.effect="Add",
           fileKin=FALSE,
           filePC=FALSE,
           #maxLine=10000,
           out="Results/2025-01-19/rMVP/AllReps/allReps_mvp")

  genotype_allReplicates <- attach.big.matrix("Results/2025-01-
19/rMVP/AllReps/allReps_mvp.geno.desc")
  phenotype_allReplicates <- read.table("Results/2025-01-
19/rMVP/AllReps/allReps_mvp.phe",head=TRUE)
  map_allReplicates <- read.table("Results/2025-01-
19/rMVP/AllReps/allReps_mvp.geno.map" , head = TRUE)

  imMVP <- MVP(
    phe=phenotype_allReplicates,
    geno=genotype_allReplicates,
    map=map_allReplicates,
    #K=Kinship,
    #CV.GLM=Covariates,
    #CV.MLM=Covariates,
    #CV.FarmCPU=Covariates,
```

```
    nPC.GLM=16,
    nPC.MLM=16,
    nPC.FarmCPU=16,
    #maxLine=10000,
    #ncpus=10,
    vc.method="BRENT",
    method.bin="static",
    threshold=0.05,
    method=c("MLM", "FarmCPU"),
    file.output=c("pmap", "pmap.signal", "plot", "log"),
    out = "Results/2025-01-19/rMVP/AllReps/")
```

6.2 GWAS with GAPIT package

Now I will perform the same GWAS test, but using GAPIT v.3 package (https://github.com/jiabowang/GAPIT):

```
# Phenotype dataset:
  pheno_GAPIT <-
read.table("Phenotype_data/vars_and_CFUs_transformed_means.txt", header = T)

# Genotype dataset:
  geno_GAPIT <-
read.table("Genotype_data/hapmap/Risinnova_filtered_imputed_2024-12-
30.hmp.txt", header = F)

 # GWAS:
  myGAPIT_FarmCPU <- GAPIT(
    Y=pheno_GAPIT,
    G=geno_GAPIT,
    model=c("MLM", "FarmCPU"),
    NJtree.group=16,
    Multiple_analysis=TRUE,
    Geno.View.output = T,
    PCA.total = 16,
    file.output=T
  )
```

7. Selection of the varieties with lower SD

In some cases, replicates for a rice variety showed a high amount of variability between values. To make the analysis more stringent, I will **select only a subset of varieties based on the standard deviation** (sd):

```
# Creation of a new df with the sd values for the different replicas for each
rice variety:
  data_sd <- vars_and_CFUs3 %>%
    group_by(Taxa) %>%
    summarise_at(vars(Transformed_CFUs),
```

```
                list(name=sd))

  colnames(data_sd) <- c("Taxa", "sd")

# Sorting the sd dataset based on sd:
  data_sd_sorted <- data_sd %>% arrange(sd)

# Selecting the first n varieties with the least values of sd:
  top_100 <- head(data_sd_sorted, 100)

# Using the list of 100 to filter the original df:
  vars_and_CFUs_first100 <- vars_and_CFUs3 %>%
    inner_join(top_100,
               by = "Taxa")

# Select only varieties and their CFU values
  filtered_cleaned_data <- vars_and_CFUs_first100[, c("Taxa",
"Transformed_CFUs")]

# Calculating mean values for the selected subset of varities:
  CFUs_sdFiltered_means <- aggregate(Transformed_CFUs ~ Taxa,
                                     filtered_cleaned_data, FUN = mean)

  colnames(CFUs_sdFiltered_means) <- c('Taxa', 'Transformed_means')

# Rounding the mean CFUs values of filtered_cleaned_data object:
  CFUs_sdFiltered_means2 <- CFUs_sdFiltered_means %>%
    mutate(Transformed_means = round(Transformed_means,
                                     digits = 0))

# Frequency table in order to view how many varieties are kept after the
filtering:
  frequency_table <- table(filtered_cleaned_data$Taxa)

# Performing a Shapiro-Wilk test to check normality:
  shapiro_test(CFUs_sdFiltered_means$Transformed_means)

# Histogram of the reduced dataset:
  reduced_histogram <- ggplot(CFUs_sdFiltered_means2,
                              aes(x = Transformed_means)) +
  geom_histogram(fill="#69b3a2",
                 color="#e9ecef") +
  theme(axis.text.x = element_text(size = 6,
                                   angle = 45,
                                   hjust = 1)) +
  xlab("CFUs") +
  ylab("Rice varieties") +
  labs(title = "Histogram of CFUs mean values in tested rice varieties") +
  theme(plot.title = element_text(hjust = 0.5, size = 22),
```

```
        axis.title.x = element_text(size = 14),
        axis.title.y = element_text(size = 14))

# Saving the histogram of reduced dataset:
  ggsave("Outputs/Figures/reduced_histogram.png",
         reduced_histogram,
         width = 8,
         height = 6)
```

7.1 Boxplots of the reduced dataset

The representation of selected values through **boxplots** will be useful to visualize the number of remaining varieties and the range of replica values for each of them:

```
# boxplots of only varieties with reduced sd:
  reduced_boxplots <- ggplot(filtered_cleaned_data, aes(x = reorder(Taxa,

Transformed_CFUs),
                                                        y =
Transformed_CFUs)) +   geom_boxplot(fill = "#19DE7F",

outlier.shape = NA,

notch = FALSE,

outliers = FALSE) +
  theme(axis.text.x = element_text(size = 6,
                                   angle = 45,
                                   hjust = 1)) +
  ggtitle("Genotypes - results with low SD between replicates") +
  labs(x = "Rice varieties",
       y = "Infection rate") +
  theme(plot.title = element_text(hjust = 0.5,
                                  size = 20),
        axis.title.x = element_text(size = 14),
        axis.title.y = element_text(size = 14))

# Saving the boxplots of reduced dataset:
  ggsave("Outputs/Figures/Reduced_boxplots_100vars.png", reduced_boxplots,
width = 8, height = 6)
```

8. GWAS with the reduced dataset

Now I will run another GWAS to check if there are **differences between the analysis performed with all the available replicates, and the one with only the best two**:

8.1 rMVP - reduced dataset

```
# rMVP requires the data to be implemented as a file, so I will save the
final dataset about pheno data in a .txt file:
```

```
  write.table(CFUs_sdFiltered_means2,
"Phenotype_data/CFUs_sdFiltered_means2.txt", sep = "\t")

# rMVP requires the hapmap version with "#" to "rs" and "assembly" columns:
  MVP.Data(fileHMP="Genotype_data/hapmap/Risinnova_filtered_imputed_2024-12-
30.hmp_hashtag.txt",
           filePhe="Phenotype_data/CFUs_sdFiltered_means2.txt",
           #sep.hmp="\t",
           sep.phe="\t",
           #SNP.effect="Add",
           fileKin=FALSE,
           filePC=FALSE,
           #maxLine=10000,
           out="Results/2025-01-19/rMVP/ReducedReps/reducedReps_mvp")

  genotype_reducedReplicates <- attach.big.matrix("Results/2025-01-
19/rMVP/ReducedReps/reducedReps_mvp.geno.desc")
  phenotype_reducedReplicates <- read.table("Results/2025-01-
19/rMVP/ReducedReps/reducedReps_mvp.phe",head=TRUE)
  map_reducedReplicates <- read.table("Results/2025-01-
19/rMVP/ReducedReps/reducedReps_mvp.geno.map" , head = TRUE)

  imMVP <- MVP(
    phe=phenotype_reducedReplicates,
    geno=genotype_reducedReplicates,
    map=map_reducedReplicates,
    #K=Kinship,
    #CV.GLM=Covariates,
    #CV.MLM=Covariates,
    #CV.FarmCPU=Covariates,
    nPC.GLM=16,
    nPC.MLM=16,
    nPC.FarmCPU=16,
    #maxLine=10000,
    #ncpus=10,
    vc.method="BRENT",
    method.bin="static",
    threshold=0.05,
    method=c("MLM", "FarmCPU"),
    file.output=c("pmap", "pmap.signal", "plot", "log"),
    out = "Results/2025-01-19/rMVP/ReducedReps/")
```

8.2 GAPIT - reduced dataset

```
 # GWAS:
  myGAPIT_FarmCPU <- GAPIT(
    Y=CFUs_sdFiltered_means2,
    G=geno_GAPIT,
    model=c("MLM", "FarmCPU"),
    NJtree.group=16,
```

```
    Multiple_analysis=TRUE,
    Geno.View.output = T,
    PCA.total = 16,
    file.output=T
  )
```

9. LD-decay plot

LD-decay allows to define the window of a genetic sequence represented by a single marker. This is a fundamental information for post-GWAS analyses, when regions of significant markers (if present) are studied in genome browsers.

For more detail, please see following reference:

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., ... & Buckler, E. S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome - Proceedings of the national academy of sciences, 98(20), 11479-11484 (https://doi.org/10.1073/pnas.201394398).

The LD matrix required to produce the LD-decay plot was produced using TASSEL 5.0 software (https://tassel.bitbucket.io/).

(The script used in the following code block was modified from the one proposed in the GitHub page: https://github.com/mohsinali1990/My_scripts/blob/main/LD%20decay%20Plot%20from%20TASSEL%20LDoutput.R).

```
# Importing TASSEL LD output file:
  ld <- read.delim("Risinnova_filtered_pruned_LD.Decay.txt",stringsAsFactors
= FALSE,header=TRUE, sep = "\t")
  str(ld)

# Removing sites that have NaN for distance or r2:
  ld_sub <- ld[ld$R.2 != "NaN",]
  ld_sub$dist <- as.numeric(ld_sub$Dist_bp)
  ld_sub2 <- ld_sub[ld_sub$dist != "NaN",]
  ld_sub2$rsq <- ld_sub2$R.2

  file <- ld_sub2[,c(1,2,7,8,15:19)]

# C values range from about 0.5 to 2, start with 0.1:
Cstart <- c(C=0.1)

# Fitting a non linear model using the arbitrary C value (N is the number of
the genotypes that have the SNP site):
  modelC <- nls(rsq ~ ( (10+C*dist)/( (2+C*dist) * (11+C*dist) ) ) *
                  ( 1+( (3+C*dist) * (12+12*C*dist+(C*dist)^2) ) / (
2*N*(2+C*dist) * (11+C*dist) ) ),
                data=file, start=Cstart, control=nls.control(maxiter=100))
```

```
# Extracting rho, the recombination parameter in 4Nr:
  rho <- summary(modelC)$parameters[1]

# Feeding in the new value of rho to obtain LD values adjusted for their
distances along the chromosome/genome:
  newrsq <- ( (10+rho*file$dist) / ( (2+rho*file$dist) * (11+rho*file$dist) )
) *
    ( 1 + ( (3+rho * file$dist) * (12+12*rho*file$dist + (rho*file$dist)^2) )
/
        (2*file$N*(2+rho*file$dist) * (11+rho*file$dist) ) )

  newfile <- data.frame(file$dist, newrsq)

  maxld <- max(newfile$newrsq,na.rm=TRUE) #using max LD value from adjusted
data
  halfdecay = maxld*0.5
  halfdecaydist <- newfile$file.dist[which.min(abs(newfile$newrsq-
halfdecay))]
  newfile <- newfile[order(newfile$file.dist),]

# Plotting the values:
  pdf("LD_decay.pdf", height=5, width = 5)
  mar.default <- c(5,4,4,2) + 0.1
  par(mar = mar.default + c(0, 4, 0, 0))
  plot(file$dist, file$rsq, pch=".", cex=2, xlab="Distance (bp)",
ylab=expression(LD ~ (r^2)), col="grey")
  lines(newfile$file.dist, newfile$newrsq, col="red", lwd=2)
  abline(h=0.1, col="blue") # if you need to add horizental line
  abline(v=halfdecaydist, col="green")
  mtext(round(halfdecaydist,2), side=1, line=0.05, at=halfdecaydist,
cex=0.75, col="green")
  dev.off()
```

8.2 Publications

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 20 March 2024
DOI 10.3389/fpls.2024.1349401

OPEN ACCESS

EDITED BY
Dilfuza Egamberdieva,
Leibniz Center for Agricultural Landscape Research (ZALF), Germany

REVIEWED BY
Malik Mobeen Ahmad,
Integral University, India
Evans N. Nyaboga,
University of Nairobi, Kenya

*CORRESPONDENCE
Lorenzo Sena
lorenzo.sena@unimore.it
Nicola Pecchioni
nicola.pecchioni@unimore.it

RECEIVED 04 December 2023
ACCEPTED 05 March 2024
PUBLISHED 20 March 2024

CITATION
Sena L, Mica E, Valè G, Vaccino P and Pecchioni N (2024) Exploring the potential of endophyte-plant interactions for improving crop sustainable yields in a changing climate. *Front. Plant Sci.* 15:1349401. doi: 10.3389/fpls.2024.1349401

Exploring the potential of endophyte-plant interactions for improving crop sustainable yields in a changing climate

Lorenzo Sena[1,2]*, Erica Mica[3], Giampiero Valè[3], Patrizia Vaccino[2] and Nicola Pecchioni[1,2,4]*

[1]Dipartimento di Scienze della Vita, Sede Agraria, UNIMORE - Università di Modena e Reggio Emilia, Reggio Emilia, Italy, [2]Centro di Ricerca Cerealicoltura e Colture Industriali, CREA – Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Vercelli, Italy, [3]Dipartimento per lo Sviluppo Sostenibile e la Transizione Ecologica, UPO – Università del Piemonte Orientale, Complesso San Giuseppe, Vercelli, Italy, [4]Centro di Ricerca Cerealicoltura e Colture Industriali, CREA – Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Foggia, Italy

Climate change poses a major threat to global food security, significantly reducing crop yields as cause of abiotic stresses, and for boosting the spread of new and old pathogens and pests. Sustainable crop management as a route to mitigation poses the challenge of recruiting an array of solutions and tools for the new aims. Among these, the deployment of positive interactions between the micro-biotic components of agroecosystems and plants can play a highly significant role, as part of the agro-ecological revolution. Endophytic microorganisms have emerged as a promising solution to tackle this challenge. Among these, Arbuscular Mycorrhizal Fungi (AMF) and endophytic bacteria and fungi have demonstrated their potential to alleviate abiotic stresses such as drought and heat stress, as well as the impacts of biotic stresses. They can enhance crop yields in a sustainable way also by other mechanisms, such as improving the nutrient uptake, or by direct effects on plant physiology. In this review we summarize and update on the main types of endophytes, we highlight several studies that demonstrate their efficacy in improving sustainable yields and explore possible avenues for implementing crop-microbiota interactions. The mechanisms underlying these interactions are highly complex and require a comprehensive understanding. For this reason, omic technologies such as genomics, transcriptomics, proteomics, and metabolomics have been employed to unravel, by a higher level of information, the complex network of interactions between plants and microorganisms. Therefore, we also discuss the various omic approaches and techniques that have been used so far to study plant-endophyte interactions.

KEYWORDS

endophytes, climate change, stress tolerance, omic techniques, arbuscular mycorrhizal fungi, microbiota

1 Introduction

The main objective of this review is to to provide a summarized update of the existing knowledge about the interactions between plants and endophytic microorganisms, focusing on their potential for improving crop production and resilience. With this aim, we reviewed their potential role for facing key challenges, such as biotic and abiotic stresses, in the context of climate change. Among the plant-microorganisms interactions, we have restricted our analysis to the endophytic ones, and among these, to the ones between endophytes and agricultural plants.

Before analysing the impact of the endophytic interactions on the stress resiliences, enclosing in a wider view the plant mineral nutrition and its relationships with quality, we have first defined the scenario of climate change in which the microbiota can have a renewed role and importance. Then, we identified the boundaries and characteristics of the endophytes within the microbiota, the general mechanisms and ontogenesis of their interactions with agricultural plants, as well as their classification in groups significant for the agricultural systems. We also explored the relevant 'omic' techniques as essential tools for analysing complex microbial communities and their interactions with plants. Omics are a natural choice for studying the complex symbiosis between plants and microorganisms in view of their exploitation (Plett and Martin, 2018; Sharma et al., 2020; Sandrini et al., 2022). Most papers analysed in this review are recent reviews and original papers published in the last decade, in particular the original research papers. Databases used during the research process included PubMed (https://pubmed.ncbi.nlm.nih.gov/), Litmaps (https://www.litmaps.com/), Open Knowledge Maps (https://openknowledgemaps.org/). Keywords used during the first phase of information collection comprehended: endophytes, abiotic stress, biotic stress, omics, PGPM. Several reviews have examined the interaction between plants and Plant Growth Promoting Microorganisms (PGPMs) (Vandenkoornhuyse et al., 2015; Santoyo et al., 2016; Khatoon et al., 2020), including potential benefits in biotic and abiotic stress scenarios (Miliute et al., 2015; Yan et al., 2019; Kamran et al., 2022), and have been here considered and updated. Also, the impact of climate change has been thoroughly studied and reviewed (Driga and Drigas, 2019; Lynch et al., 2021), as it has specific aspects for plant and microorganism interactions (Compant et al., 2010; Trenberth et al., 2014; Classen et al., 2015; Shahzad et al., 2021), and both these fields have been here contextualized with respect to topics treated in this review. Abiotic stressors such as drought, heat, and salinity have been included since of highly significant concern (Lipiec et al., 2013; Evelin et al., 2019; Ma et al., 2020; Angon et al., 2022). The role of microorganisms in protecting from biotic stresses (Pandey et al., 2019; Chaudhary et al., 2022), including viruses (Bao and Roossinck, 2013), nematodes (Bernard et al., 2017; Gamalero and Glick, 2020; Pulavarty et al., 2021), insects (Bradshaw et al., 2016), fungi, and bacteria (Muthu Narayanan et al., 2022) have been previously discussed and are here gathered and updated.

In this review, we combined various aspects, usually considered individually, that characterize the symbiosis between endophytes and plants: mechanisms of selection and interaction, effects on biotic and abiotic stress factors, omic techniques for the study of such complex symbioses, consequences from the nutritional point of view of the interaction between endophytes and agricultural crops. For each type of stress, we considered examples where endophytes have demonstrated beneficial effects on agricultural plants, and we critically analysed potential limits of microbe-based approach. Finally, we also briefly discussed the possible future perspectives of the use of endophytes in crop production under climate change, highlighting possible limitations and improvements.

2 The climate change scenario and plant-microorganisms interactions

Plants, in their evolutionary path outside the oceans, have established important relationships with various microorganisms, such as bacteria, fungi, protists, and viruses (Plett and Martin, 2018). These microorganisms can live associated with different plant tissues and organs and form the plant microbiota. This can be divided in different microbial communities, based on the plant parts they colonize: the phyllosphere, which includes all the aboveground plant tissues (*i.e.*, stems and leaves), the endosphere, represented by the internal tissues, the spermosphere, *i.e.* the seed, and the rhizosphere, which comprehends the roots surface and the soil surrounding it, reached by the root exudates (Johnston-Monje and Raizada, 2011; Levy et al., 2018; Sharma et al., 2020). The different conditions of each of these habitats lead to diverse microbial communities, even within the same plant.

Through millions of years, pathogenic, competing, mutualistic, or symbiotic associations have been established between plants and microorganisms. Fossil records provide substantial evidence that over the past 450 million years, virtually all plants have formed symbiotic relationships with microbes since their first colonization of land. Various studies have documented microbial symbionts in fossilized plant specimens dating back to this era (Redecker et al., 2000; Krings et al., 2007; Bonfante and Genre, 2010; Kawaguchi and Minamisawa, 2010; Genre et al., 2020). Partida-Martínez and Heil (2011) suggested that the plant microbiota plays additional essential roles in phenotypic and epigenetic plasticity, as well as in the continuous evolution of plants.

The interactions between plants, soil, and microbes have played and continue to play a vital role by influencing various processes that contribute to plant health and productivity (Ahmed et al., 2019; Khatoon et al., 2020). Endophytes, the microorganisms of the endosphere, can provide benefits to the whole plant, either by promoting plant growth (Johnson et al., 2003; Rodriguez et al., 2008), eliciting the production of metabolites and useful chemicals such as antibiotics and agrochemicals (Kaul et al., 2012; Kusari et al., 2014; Wang et al., 2015), or helping plants to cope with stresses (Yan et al., 2019; Chen et al., 2021).

Since the first alerts of incoming human-driven climate change (Driga and Drigas, 2019), it became clear that the phenomenon would also impact the relationships between plants and

microorganisms (Classen et al., 2015; Singh et al., 2019). Climate change manifests mainly as a global increase in temperature, dry periods, rapid changes of meteorological conditions (*e.g.*, flash droughts), rainfall intensity and uneven distribution (Easterling et al., 2000; Hammond et al., 2012), with differences in impact linked to various geographical regions (Surjan et al., 2016). All these deviations from a previously more stable climate, particularly the rise in temperature, are shifting plant phenology and the global distribution of plants (Sykes, 2009; Geissler et al., 2023), and significantly increase the threats to survival of natural environments (Abbass et al., 2022).

Agricultural systems are both the subject of climate change impacts, more negative than positive (Shahzad et al., 2021; Yadav et al., 2021), and contributors to it, like other human activities that require energy inputs (Hatfield et al., 2020). In fact, although on a smaller scale per unit area compared to other human activities, due to the vast extent of agricultural systems, they can contribute to GreenHouse Gases (GHGs) emissions.

Some agricultural practices, such as intense tillage, irrigation and extensive fertilizers usage, in addition to the ever-increasing use of machinery operated by fossil fuels, lead to increased emissions of GHGs in the atmosphere (Lin et al., 2011; Galic et al., 2019). As for CH_4 and N_2O, agricultural activities are responsible for around half and three-quarters of all anthropogenic emissions, respectively (Lynch et al., 2021). Agriculture, especially in the form of livestock production and rice cultivation, is one of the main sources of CH_4, which has a Global Warming Potential (GWP) much stronger than CO_2 (Saunois et al., 2020). As for N_2O, agriculture remains the main source of this GHG, primarily through the use of nitrogen fertilizers, both synthetic and natural. The lifetime of N_2O in the atmosphere is about 120 years, and its GWP is about 210 times higher than that of CO_2 (Singh, 2000).

Today, converting conventional agriculture into a sustainable, yet high-yielding system, is crucial to meet both our future food needs and the integrity of the biosphere. In fact, an increasing food demand cannot be met by simply improving current agricultural practices based on fossil carbon inputs, which are detrimental to the environment (Santoyo et al., 2017; Slepetiene et al., 2020). Indeed, the continuous use of fertilizers and pesticides derived from chemical synthesis has been shown to deteriorate agroecosystems, reduce soil biodiversity, and impair natural predators of insects (McLaughlin and Mineau, 1995; Alavaisha et al., 2019).

For this reason, it was recently proposed that the deployment of plant-microbe interactions be used as one of the strategies for converting agricultural systems from the traditional mechanistic approach to the new paradigm of agroecological intensification.

3 The plant microbiota

The metabolism and morphology of plants and their microbiota are intrinsically connected, with a dynamic interplay between both, maintaining the function of the holobiont. The holobiont concept suggests a new perspective of organisms as meta-organisms, composed of a host organism and its associated microorganisms, co-evolved as species assemblages (Berg et al., 2016). In the plant kingdom, microbiota fulfil important functions for the holobiont, promoting its growth and increasing tolerance against biotic and abiotic stressors, as well as for the ecosystem, decomposing crop residues and contributing to nutrient cycling (Chaparro et al., 2012; Mendes and Raaijmakers, 2015; Vandenkoornhuyse et al., 2015).

The composition and functional diversity of the plant microbiota is influenced by biotic factors including age or developmental stage, species or cultivar, and plant health, as well as by abiotic factors such as soil properties, nutrient status, and climatic conditions (Berg and Smalla, 2009).

Three main classes of microorganisms are reviewed as part of the plant microbiota: endophytic fungi and bacteria, residing inside the plant tissues; rhizospheric microorganisms, residing in the soil surrounding plant roots; and mycorrhizal fungi.

3.1 Endophytic bacteria and fungi

An endophyte is a microorganism that lives, at least for a portion of its life cycle, inside plant tissues without producing any symptoms of disease. Among endophytic microorganism are bacteria, fungi, actinomycetes, and viruses (Bao and Roossinck, 2013; Stępniewska and Kuźniar, 2013), and they are found in almost every plant (Partida-Martínez and Heil, 2011). Endophytes can be classified as systemic (or true endophytes) or non-systemic (or transient endophytes), depending on their life cycle and the type of relationship they establish with the plants. Indeed, there is a huge variability of symbiotic lifestyles, from mutualism to parasitism, depending on genotypic and/or environmental factors. True endophytes co-evolved with their hosts, creating mutualistic relationships, and are often vertically transmitted, while transient endophytes could shift from a pathogenic to a mutualistic behaviour, depending on external conditions (Wani et al., 2015). In any case, several factors may influence the host response to endophytic interactions: mainly host genotype, nutrient availability, environment, field management practices, and microorganism strain (Hesse et al., 2003; Malinowski and Belesky, 2006; Qawasmeh et al., 2012).

Endophytes can be found in the root tissues, where they are more abundant, but also in the aerial parts of the plant (leaf, flower) and in the seeds (Hallmann, 2001).

Endophytic bacteria could be considered as a subgroup of the rhizospheric bacteria, that acquired the ability of colonizing their host plants (Reinhold-Hurek and Hurek, 1998). In fact, rhizosphere is a highly competitive environment (Raaijmakers et al., 2002), while the internal tissue of the host may represent a protected ecological niche with minor perturbations from the external conditions of the soil or, in general, of the environment (Kasmir et al., 2011), and this could have created an evolutionary drive from the first to the second. Endophytic and rhizospheric bacteria implement very similar strategies to promote plant growth, but usually endophytic microorganisms have a higher beneficial potential.

Among bacterial endophytes, *Proteobacteria* are the most widely represented, including α-, β-, and γ-*Proteobacteria*; the latter taxon being the most diverse and widespread (Miliute et al., 2015; Santoyo et al., 2016). Other classes frequently isolated from

plant tissues are *Actinobacteria*, *Bacteroidetes*, and *Firmicutes* (Reinhold-Hurek and Hurek, 2011). Rarer, but still present, are *Acidobacteria*, *Planctomycetes* and *Verrucomicrobia* (Santoyo et al., 2016). The most common bacterial genera are *Bacillus* (*Firmicutes*), and *Pseudomonas* (*Proteobacteria*). *Rhizobia* spp. are also included among endophytic bacteria, as they colonize internal root tissues of *Fabaceae*, developing the typical nodules for nitrogen fixation.

Among endophytic fungi, on the other hand, we can find the families *Clavicipitaceae* (associated with grasses), *Cladosporiaceae*, *Glomerellaceae*, *Sebacinaceae*, *Pleosporaceae*, and *Hypocreaceae*, among which the most representative genus is *Trichoderma* (Hardoim et al., 2015).

Key molecular and metabolic pathways at the base of host-microbe recognition and strain selection by different plant genotypes are starting to be elucidated. It is clear how different genotypes grown in different soil/environmental combinations are enriched with different endophytic strains (Granér et al., 2003; Rashid et al., 2012; Edwards et al., 2015). The role of plant genotype on strain selection and microbial population composition has been widely studied and even if it is not the main force driving microbial diversity, it is able to modulate it (Weinert et al., 2011; Bulgarelli et al., 2015; Walters et al., 2018). Interestingly, it has been shown that plant domestication, and lately the development of high-yielding genotypes, caused a reduction in the plant capacity of associating with useful microorganism (Porter and Sachs, 2020; Valente et al., 2020), since human-centered breeding neglected the traits related to microbiota association.

Considering endopythes role in promoting plant growth, especially in nutrient-deprived conditions, and in increasing plant defence against pathogen attack, either directly or indirectly, they are now considered a tool for crop management. These could sustain agricultural practices with fewer chemical inputs. However, research is still needed to further the knowledge both on the plant-side, trying to identify the genetic factors responsible for a more efficient microbial colonization, and on the microbic-side, to isolate the most promising micro-organisms and the most effective synthetic communities. Once these aspects are clarified, it will be possible to engineer plants and microbes to make their interaction more effective. It will also be possible to explore the use of root exudates or organic compounds that might serve as pre-biotics. Finally, we should be able to overcome and the bottleneck, as of the applicability of this research to open fields conditions, that remains challenging.

Another relevant application of this category of microorganisms is phytoremediation. It has been shown that some bacterial strains are tolerant to high concentrations of heavy metals, as Cd, Cu and Zn, and other pollutants. These strains favour their accumulation into the plants, promoting their growth (Ullah et al., 2015; Ma et al., 2016), even in a stressful environment.

3.2 Mycorrhizal fungi

Mycorrhizae are likely to have played a crucial role in the evolution of terrestrial plants (Bonfante and Genre, 2008; Partida-Martínez and Heil, 2011). Today, mycorrhizal fungi can be divided into ectomycorrhizae, when the hyphae colonize the root intercellular spaces, and endomycorrhizae, when they penetrate inside the plant cells. Endomycorrhizae are further divided into orchid (OM), ericoid (ERM) and arbuscular mycorrhizae (AMs) (Bonfante and Genre, 2010; Chen et al., 2021).

Ectomycorrhizae are mostly associated with woody perennial trees such as *Pinaceae*, *Fagaceae*, *Dipterocarpaceae* and *Caesalpinoidaceae*, contributing to the wellness of most forest ecosystems (van der Heijden et al., 2015). EM fungi are phylogenetically diverse and belong to *Basidiomycetes*, *Ascomycetes* and *Zygomycetes*, representing the orders *Pezizales*, *Agaricales*, *Helotiales*, *Boletales*, and *Cantharellales*. EM hyphae grow partially inside the root intercellular space and partially outside, creating a mantle covering the tip of colonized lateral roots, called the Hartig net (Shi et al., 2023).

EM fungi live in symbiosis with their hosts, but are also facultative saprotrophs, decomposing complex organic matter present in the soil and making nitrogen and phosphate available for the plants (Martin and Nehls, 2009). In turn, the EM fungi receive photosynthates from the plant. Genomic data and functional studies show the presence of specialized families of phosphate, ammonium, and nitrate transporters (Jargeat et al., 2003; Casieri et al., 2013; Becquer et al., 2014; Stuart and Plett, 2019), supporting evidence of their fundamental role in different genomes of EM fungi. The entire metabolic chain transporting N/P from the soil to the plants through EM hyphae has been fully elucidated in recent years, and many advances have been made thanks to high-quality genomic sequences available (Nehls and Plassard, 2018).

Among endomycorrhizae, Arbuscular Mycorrhiza (AM) are the most represented, as these symbioses are formed by the 70-90% of terrestrial plants species, while the fungi all belong to the monophyletic *phylum Glomeromycota* (Schüβler et al., 2001). AM fungi are obligate biotrophs, considered organisms with no or rare sexual reproduction, and present aseptate hyphae developing inside the plant cells, where they form the characteristic tree-shaped hyphal structure. In AM symbiosis, the fungi support the plants by supplying mainly P-based nutrients and water, while the plant supplies the fungi with carbon nutrition (Parniske, 2008). It is estimated that almost 20% of the photosynthetic products of terrestrial plants are allocated to AM (Bago, 2000). The N contribution is less pronounced in AM compared to EM, even though some publications have shown that the portion of N transported to the plant cells from AM is not negligible and depends on soil pH, moisture, and nutrient concentration (Govindarajulu et al., 2005; Jin et al., 2005; Tanaka and Yano, 2005).

Typically, a soil ecosystem characterized by mycorrhizal symbiosis features a wide variety of plant-fungi relationships, thus offering a broad functional diversity (van der Heijden et al., 1998; Burleigh et al., 2002). This diversity arises from the presence of different plant species and fungi, each with the potential to select the most cooperative partner (Werner and Kiers, 2015). Plant roots tend to be enriched with fungal species or isolates that ensure optimal growth benefits (Bever et al., 2009; Kiers et al., 2011; Verbruggen et al., 2012), while AM fungi typically select plants that can allocate the highest amount of C nutrients (Lekberg et al.,

2010; Kiers et al., 2011). However, the genetic and ecological mechanisms underlying this partner selection remain unclear, and understanding them could reveal crucial insights for future agricultural applications.

The success of the mutualistic relationship depends on many factors, such as the combination of plant-fungi genotypes, their molecular and metabolic regulation, and soil characteristics (pH, structure, moisture), nutrient availability, and colonization rate. Different plant genotypes, under controlled conditions, may respond differently to AM in terms of plant growth and yield, as shown in crops such as maize (*Zea mays* L.) (Ramírez-Flores et al., 2019) and sorghum (*Sorghum bicolor* (L.) Moench) (Watts-Williams et al., 2019), or in terms of stress resistance, as demonstrated in rice (*Oryza sativa* L.) (Chareesri et al., 2020) and bread wheat (*Triticum aestivum* L.) (Lehnert et al., 2018). The results of these studies represent milestones for future breeding programs, supporting more sustainable agriculture.

Besides AM, endomycorrhizae are also represented by ericoid (ERM) and orchid (OM) mycorrhizae. As for ERM, the fungi colonize plants of the *Ericaceae* family, such as *Calluna*, *Vaccinium* and *Erica*, typically found on nutrient-poor and acidic soils (Soudzilovskaia et al., 2019).Thus, they represent an essential way to mobilize organic material in infertile soils. In OM, the fungi belong to *Basidiomycetes*, mainly to *Rhizoctonia* species, with which most orchids are associated (Favre-Godal et al., 2020). Orchids strongly depend on the nutrients coming from the fungi, especially for the initial stages of seed germination and growth.

It should be noted that there are also plants that establish different types of mycorrhiza, either spatially, temporally, or simultaneously, within the same root system. For example, this is the case with plants from *Populus*, *Fraxinus* and *Eucalyptus* genera (Ambriz et al., 2010; Teste et al., 2020), as well as other plants known as dual-mycorrhizal plant species. It must be noted that there are also plants establishing different types of mycorrhiza, in a spatially or temporally distinguished manner, or simultaneously, within the same root system. For example, this is the case of plants from *Populus*, *Fraxinus* and *Eucalyptus* genera (Ambriz et al., 2010; Teste et al., 2020) and other plants that are called dual-mycorrhizal plant species.

It is worth mentioning that often a single fungus may connect the root systems of several plants, creating what are known as common mycorrhizal networks (Figueiredo et al., 2021). This facilitates the exchange of signalling compounds and nutrients, increases pathogen resistance, and promotes plant growth.

4 Plant-endophyte interactions

Successful endophyte colonization involves compatible plant-microbe interactions (Khare et al., 2018). Several steps can be identified to accomplish the whole process, which includes attraction, recognition, and colonization (Figure 1).

4.1 Attraction

Some endophytes are seed-borne and are present in germinated plants, thus representing a bridge across host plant generations (Coombs and Franco, 2003). Also, plants with vegetative propagation can transmit their endophytic microbiota to the next generation (Kamran et al., 2022). In others, a chemotactic response of

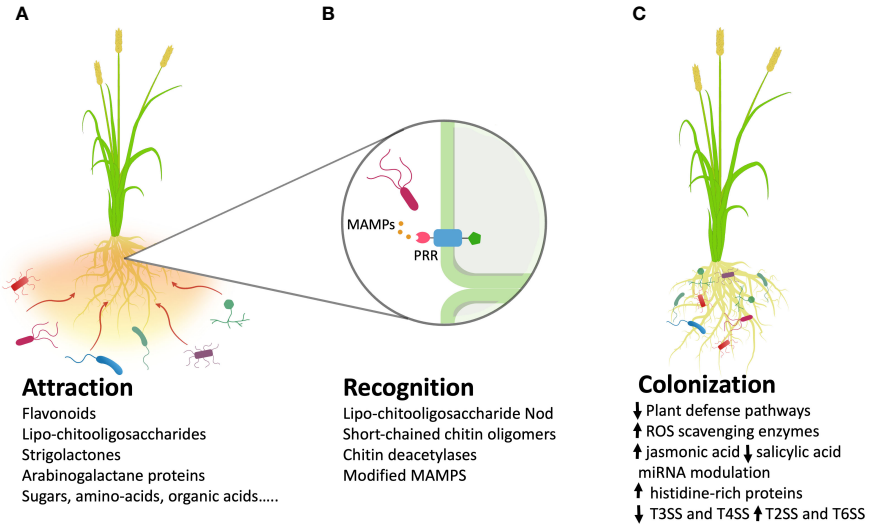


FIGURE 1
Metabolites and processes involved in three crucial steps of plant roots colonization by bacterial endophytes. **(A)** Release of molecular exudates from the roots favors the chemotactic response by the endophytes presents in the surrounding soil; **(B)** The recognition step is a complex phase in which plant receptors recognize microbial molecules that trigger molecular pathways. A typical recognition mechanism acts through MAMPs (Microbe-Associated Molecular Patterns) and plant PRRs (Pattern Recognition Receptors). **(C)** Once inside the plant, endophytes can influence many processes, for example modulating the levels of phytohormones or increasing ROS (Reactive Oxygen Species)-scavenging enzymes. (MAMPS, Microbe-Associated Molecular Patterns; ROS, Reactive Oxygen Species; T3SS, type III secretion system; T4SS, type IV secretion system; T2SS, type II secretion system; T6SS, type VI secretion system; arrow up, up regulated metabolites/processes; arrow down, down regulated metabolites/processes).

endophytes to host plant root exudates has been observed. These exudates are rich in biomolecules (including nutrients and water), and thus attract or are recognized by friendly endophytic microbes (Compant et al., 2010; Brader et al., 2014). Flavonoids are one such group of metabolites secreted by several plants and categorized as chemo-attractants, playing an important role in endophytic interaction with the root hair. Flavonoids are used in bioformulations to affect successful infection of legume roots by rhizobia (Arora and Mishra, 2016). They are also reported to play a role with non-rhizobial endophytes, and it has been proven that in the presence of these metabolites, the colonization of roots in rice and wheat by the endophytic *Serratia* sp. EDA2 and *Azorhizobium caulinodans* ORS571 is far more effective (Webster et al., 1998; Balachandar et al., 2006). Lipo-ChitoOligosaccharides (LCO), also called Nod factors, are well-characterized signal molecules activating the Common Symbiotic Pathway (CSP) in rhizobia-legume and arbuscular mycorrhizal associations (Gough and Cullimore, 2011). Recently, StrigoLactone (SL) secreted by roots of *Arabidopsis thaliana* was found to act as a signal molecule for colonization of endophytic *Mucor* sp (Rozpądek et al., 2018). SL treatment may also activate the synthesis and release of short-chain chitin oligomers, whose perception by the plant can stimulate the symbiotic signalling pathway during early stages of host colonization (López-Ráez et al., 2017). Additionally, ArabinoGalactan Proteins (AGPs), which are highly glycosylated members of the Hydroxyproline-Rich GlycoProteins (HRGPs) superfamily of plant cell wall proteins, play a crucial role in establishing the interaction of plant with microbes (including endophyte) at several stages (Nguema-Ona et al., 2013). Several other root exudates, including sugars, amino acids, organic acids, phenolic compounds, and other secondary metabolites, are now known to be secreted by plant roots, which selectively invite the mutualistic microbes, particularly the endophytes (Chagas et al., 2017). A bacterial endophyte can also utilize the hyphae of a fungal pathogen to gain access from the soil to plant roots, thereby protecting the host from infection (Palmieri et al., 2020). showed that the endophytic rhizobacterium *Rahnella aquatilis* utilizes hyphae of the fungal pathogen *Fusarium oxysporum* to access and colonize plant roots. Metabolomic and multi-omics approaches, as those described below, would most likely increase the knowledge about metabolites released by plant seeds and roots involved in attracting favourable endophytic microorganisms. This information is necessary to address the realization of bioformulations or genetic engineering approaches to increase the production of chemo-attractants and thus the colonization of plant tissues by bacterial endophytes under normal and stressful plant growth conditions (Arif et al., 2020).

4.2 Recognition

The strategies that plants use to distinguish beneficial microbes, such as endophytes, from pathogens, are still a matter of research and not completely understood. Plants possess various PRRs (Pattern Recognition Receptors) that recognize M/PAMP (Microbe/Pathogen-Associated Molecular Patterns) ligands and initiate immune reactions (Hacquard et al., 2017). The most characterized MAMPs include flagellin, elongation factor Tu, peptidoglycan, lipopolysaccharides, bacterial cold shock proteins, bacterial superoxide dismutase, Beta-Glycan, β-glucans from oomycetes, and chitin (Newman et al., 2013). These MAMPs are recognized on the surface of plant cells by PRRs, which include receptor-like kinases and receptor-like proteins (Tang et al., 2017). Both pathogens and symbionts can be recognized by PRRs, because the M/PAMPs are not specific to pathogens. To avoid recognition by the host plant and the subsequent immune response, pathogens and symbionts have evolved complex extracellular invasion strategies. Due to the similarity of pathogen and symbiont genomes (Reinhardt et al., 2021), common extracellular strategies exist between them. They can be divided into three categories: avoiding accumulation of MAMP precursors, reducing hydrolytic MAMP release, and preventing MAMP perception (Buscaill and van der Hoorn, 2021). These strategies can involve different microbial effectors. Symbionts have developed various strategies to allow their potential hosts to better distinguish them from pathogens during the recognition phase. For example, LCO Nod Factors are perceived by legumes, activating the symbiotic pathway (Radutoiu et al., 2003; Bozsoki et al., 2020). In rice, short-length chitooligosaccharide (CO4) triggers symbiotic signal transduction with the symbiotic complex receptor MYR1–CERK1.This suppresses the formation of the CEBiP-CERK1 heteromer that would mount the immune response, while long-chain chitooligosaccharide (CO8) induces immune signalling through CEBiP-CERK1 (Chiu and Paszkowski, 2021; Zhang et al., 2021). It has also been observed that fungal endophytes produce chitin deacetylases, which deacetylate chitosan oligomers that are thus not perceived by plant receptors (Cord-Landwehr et al., 2016). There is also evidence where endophytic bacteria are known to produce their own MAMPs, which are either not recognized by PRRs of plants or trigger in plants a comparatively weak and transient defence reaction compared to pathogenic interactions (Vandenkoornhuyse et al., 2015). Along this line, it was shown that in grapevine (*Vitis vinifera* L.), thanks to an alteration in sequence, the perception of flagellin from an endophytic *Burkholderia phytofirmans* by LRR-RLK (Leucine-Rich Repeat-Receptor-Like Kinase) FLAGELLIN-SENSITIVE 2 (FLS 2) PRR was different from the perception of those of bacterial pathogens, such as *Pseudomonas aeruginosa* or *Xanthomonas campestris* (Trdá et al., 2015). Still, there are knowledge gaps about the genetic mechanisms that differentiate recognition strategies deployed by beneficial with respect to pathogenic microbes, that need to be filled through comparative genomic studies between the two microbial categories, complemented with functional analyses. Availability of complete information on gene functions involved in the endophytic recognition would allow targeted modifications of favourable strains through gene editing and/or over-expressing approaches that would improve the plants capability in recognizing microbial symbionts and protecting them from the immune response.

4.3 Colonization

Potential entry points for endophytes are cracks formed at the emergence of lateral roots, zones of root elongation, root hair cells, and wounds. Other sources include stomata, particularly of young stems and leaves, lenticels, and germinating radicles. For successful

colonization, some bacteria must find their way to these apertures. *Klebsiella pneumoniae* 342 (Kp342) can colonize the lateral root junctions in wheat and alfalfa (*Medicago sativa* L.) (Dong et al., 2003). Similarly, *Herbaspirillum seropedicae* and *Gluconacetobacter diazotrophicus* dominate colonization at lateral root junctions (James et al., 1997; Luna et al., 2012). Some endophytes enter through infection colonization, where cellulolytic and pectinolytic enzymes produced by endophytes come into play (Miliute et al., 2015), such as pectate lyase, which has been implicated in the colonization of *Klebsiella* strains (Kovtunovych et al., 1999). Symbionts can colonize hosts while overcoming the response to Damage-Associated Molecular Patterns (DAMPs) and MAMPs, while a response against pathogens is still possible in the presence of non-pathogenic microbes (Zhou et al., 2020). Several studies have proven that there is a downregulation of plant defence pathways during the colonization of plants by mutualistic partners, such as rhizobia or AMF (Fouad et al., 2014; Benhiba et al., 2015). In the case of an oxidative burst or generation of Reactive Oxygen Species (ROS) as plant defence system, endophytes protect themselves by producing enzymes such as superoxide dismutases, catalases, peroxidases, alkyl hydroperoxide reductases, and glutathione-S-transferases (Zeidler et al., 2004). The root endophytic fungus *Serendipita indica* secretes a histidine-rich protein to improve its access to micronutrients and to influence oxidative stress and reactive oxygen homeostasis to facilitate the colonization of the host plant (Nostadt et al., 2020). Also, symbionts could induce Jasmonic Acid (JA) and suppress Salicylic Acid (SA) formation to Induced Systemic Resistance (ISR), whereas pathogens typically enhance the SA biosynthesis to mediate Systemic Acquired Resistance (SAR) in plants (Martínez-Medina et al., 2017). Moreover, during mutualistic interactions, late induction of SA/JA/ET signalling pathways prevents the microbe from 'overstepping' and 'overpowering' the plant (Plett and Martin, 2018). It is reported that most miRNAs induced in the host during the establishment of endophytes also target hormone-response pathways (Formey et al., 2014). During AMF infection, the miRNA E4D3Z3Y01BW0TQ is upregulated and disrupts Gibberellic Acid (GA) signalling pathway, known for repressive action against mutualistic associations (Formey et al., 2014; Martín-Rodríguez et al., 2015; Wu et al., 2016). The plant may also induce the expression of different groups of genes during colonization by diverse sets of microbes. For example, during the establishment of symbiosis, the majority of pathways targeted by miRNAs for plant defence system are turned off, thus preventing the obstacle to the proliferation of endophytes (Plett and Martin, 2018). For AMF, two receptor-like kinases called Arbuscular Receptor-like Kinase 1 (ARK1) and ARK2 are required for the sustenance of the symbiotic interaction in several plant species (Montero et al., 2021). Moreover, AMF are separated from the plant cytoplasm by a specialized host-derived membrane, which represents the main interface facilitating the bidirectional exchange of nutrients and information and protects the microbial symbionts from the immune response (Huang et al., 2021). The biosynthesis of this peri-arbuscular membrane is controlled by a gene called *GLUCOSAMINE INOSITOL PHOSPHORYLCERAMIDE TRANSFERASE1 (GINT1)* (Moore et al., 2021). Protein Secretion Systems (SSs) in bacteria also modulate the plant immune system. Among all known SSs, Type III Secretion System (T3SS) and Type IV Secretion System (T4SS) are essential for delivering Effector Proteins (EFs) by the pathogenic bacteria into the plant, but these are either absent or present in low abundance in mutualistic endophytic bacteria (Green and Mecsas, 2016; Liu et al., 2017). Notable exceptions can be seen in some rhizobial strains where T3SS is important for nodulation of some legumes (Ausmees et al., 2004; Okazaki et al., 2016, 2013). The T3SS is also a determinant for rice endophyte colonization by non-photosynthetic *Bradyrhizobium* spp (Piromyou et al., 2015). Furthermore, the Type 2 Secretion System (T2SS) was demonstrated to be required for suppressing MAMP-triggered immunity in efficient root colonizer bacteria and, notably, enhanced the colonization capacity of other tested commensal bacteria in *Arabidopsis* (Teixeira et al., 2021). On the other hand, in mutualistic proteobacterial endophytes, Type VI Secretion Systems (T6SSs) are present, and are also commonly found in commensal and pathogenic plant-associated bacteria. However, they are associated with important functions, which are apart from virulence, usually such as competition against other bacteria (Reinhold-Hurek and Hurek, 2011; Bernal et al., 2018). From this picture, it emerges that colonization involves a plethora of traits from both, plants and microorganisms and available data most likely shed light only on a small fraction of the involved processes. Considering the plant side, in addition to the information provided above, recent investigations highlighted that plant genes can shape the microbiota composition (Zhang et al., 2019; Deng et al., 2021; Escudero-Martinez et al., 2022; Oyserman et al., 2022; Escudero-Martinez and Bulgarelli, 2023) and that wild germplasm is supposed to support higher microbiome diversity than domesticated counterparts (e.g (Pérez-Jaramillo et al., 2018; Nerva et al., 2022). Taken together, these results indicate that there is room for genetics interventions addressed to increase both plant and beneficial microbial aptitude in establishing favourable interactions and that further functional and multi-omics investigations can increase the available targets for improving endophytic colonization by plant growth promoting microorganisms. Once plants and microbial effective targets are identified, these could be modified/introgressed/engineered into their respective genomes (Arif et al., 2020; Nerva et al., 2022; Escudero-Martinez and Bulgarelli, 2023).

5 Omics for the study of plant-endophyte interactions

The intricate network of interactions among the various actors of the microbiota requires the use of advanced techniques with higher likelihood of obtaining global information from the organisms. This is to decipher a complex system and attempt to clarify the role of each organism at the genetic, transcriptional, metabolic, and physiological/phenotypic level.

The microbiota consists of several microorganisms inhabiting soil layers and distinct plant tissues (Bulgarelli et al., 2013; Compant et al., 2019), among which different relationships can be established, depending on environmental factors. A multi-layer communication web organizes the connections among the microorganisms, between the different plants growing in the same soil, and between plants and microorganisms (Hassani et al., 2018; Xiong et al., 2021). Much has been learned about these mechanisms in recent years, thanks to the advent of Next Generation Sequencing (NGS) and, more

broadly speaking, to the "omic" technologies, *i.e.* genomics, transcriptomics, proteomics, and metabolomics. The plant-microbe scientific community has greatly benefited from them.

5.1 Genomics and metagenomics

The development of NGS technologies has allowed to perform whole genome sequences of numerous fungi and bacteria. Overcoming the limit of traditional culture-dependent identification approaches, it has enabled the identification of as much microbial diversity as possible. Meta-genomic approaches nowadays almost routinely make use of DNA extraction from the whole soil/tissue microbial population, allowing the analysis of its gene/taxa content using next generation sequencing (Allan, 2014). The sequencing can involve the whole genome, which is then tentatively assembled and annotated, or only the 16S rRNAs. These data can be used to study the microbial diversity and to evaluate the absolute abundance of different bacterial strains, taking into account the different copy number of 16S rRNA genes in distinct bacterial genomes (Case et al., 2007; Wang et al., 2020).

It is worth mentioning that the availability of several AM fungi genomes has allowed for the study of the evolution of these organisms, which are considered as living fossils and ancient asexuals (Parniske, 2008). Their genome size is highly variable, from the 39.6 Mb of *Paraglomus occultum* (Malar C et al., 2022) to 784 Mb of *Gigaspora margarita* (Venice et al., 2020), with large genomes hosting a higher number of genes and a high proportion of transposable and active elements (Venice et al., 2020; Dallaire et al., 2021); differences that could explain their intra-specific variability.

In parallel, the study of the epigenome variability is emerging as a tool to understand a hidden layer of variability (Chaturvedi et al., 2021). Another interesting example of recent scientific advances given by the most recent sequencing technologies concerns the use of long-read sequencing and chromatin conformation capture techniques that made it possible to understand the genomic organization of multi-nucleate coenocytic hyphae of AM, demonstrating their heterokaryotic nature and supporting rare sexual reproduction events (Yildirir et al., 2022; Sperschneider et al., 2023).

5.2 Transcriptomics

This approach, coupled to advanced bioinformatic pipelines, for example using algorithms of artificial intelligence, could be considered as the most useful omic science for understanding the network of interactions. It has largely benefited from NGS technologies, whose recent advances have significantly increased the sensitivity to catch the rarest transcripts. Moreover, long-read sequencing technologies in the Iso-Seq approach, among others, allow to cover the entire transcript length thereby distinguishing rare isoform resulting from alternative splicing events (Li et al., 2017; Zhang et al., 2019).

Transcriptomics has been successfully applied to uncover the plant molecular strategies used to recruit the most favourable microbial organisms in response to diverse abiotic and biotic stresses, and to understand the microbial molecular networks used to successfully establish the symbiotic relationships (Sheibani-Tezerji et al., 2015). Furthermore, transcriptomic studies applied to bacterial cells may help decipher which strains and cells, among the endophytic or rhizospheric population, are transcriptionally active (Sharma et al., 2004; Knauth et al., 2005; Sheibani-Tezerji et al., 2015), surmounting DNA-based technologies that cannot distinguish non-viable cells. The completion of whole genome sequencing of new microbial species and strains will be crucial allowing the identification of the microbial response to different soil characteristics and plant genotypes.

RNA-seq has also been applied to the population of small RNAs, to identify and characterize plant miRNAs involved in host-microbiota communications. These small non-coding RNAs are important key regulators of different plant biological pathways, from organ development to stress response. It has been shown that microorganisms might stimulate the expression of plant miRNAs, modulating drought tolerance response, nutrient uptake, or facilitating symbiosis establishment (Mohsenifard et al., 2017; Pentimone et al., 2018; Kord et al., 2019; Tao et al., 2023). Besides plant miRNAs, other small RNA-like molecules are coded by fungi and bacteria, that could be involved in an intriguing system of cross-kingdom RNAi-mediated regulation, for example during mycorrhizal colonization (Silvestri et al., 2019), or nodule formation (Ren et al., 2019). The intriguing hypothesis of small RNAs as mobile cell-to-cell signalling molecules (Huang et al., 2019) has been explored in detail thanks to the possibility to purify Extracellular Vesicles (EV). In fact, EVs have been shown to transport small RNAs between plant host and microorganisms in both pathogenic and mutualistic interactions (Cai et al., 2019), thus their further analysis will deepen our understanding of below-ground inter- and intra-kingdom communications. Single cell transcriptomics, coupled with enhanced microscopy techniques will greatly improve our understanding of endophyte bacteria and AM fungi lifestyle inside the plants (Yin et al., 2023) and of root cells regulatory network.

5.3 Proteomics and metabolomics

In parallel with next-generation and third-generation (or single molecule) (Schadt et al., 2010) sequencing technologies, which have been successfully applied to the study of all the DNA/RNA populations present in a tissue, the analyses aimed at characterizing the entire set of proteins, with their post-translational modifications, and metabolites have evolved. This evolution is to comprehensively study all the molecules in a microbe/plant biological system, thereby increasing their sensitivity and throughput. Proteomics has been applied to plant tissues to understand how the presence of an endophyte, for example, may modulate the synthesis of different plant proteins (Lery et al., 2011), revealing their role in cellular recognition. The analysis of the protein-protein interactions, also called an interactome, is essential to unveil molecular mechanisms at the base of symbiotic relationships.

Metabolomics and proteomics have been used to analyse root exudates, containing both primary and secondary metabolites, to understand how biotic and abiotic factors might modulate their composition, and as a result, attract and associate with different microorganisms. However, analysing either the metabolites or the proteins present in a colonized plant tissue, or both, is challenging. This is because it is difficult to distinguish between molecules produced by either the plants or the fungi/bacteria. Recently, to resolve this issue, several techniques have been developed to narrow the analyses to the single-cell level, such as Mass Spectrometry Imaging (Boughton and Thinagaran, 2018), Laser ablation electrospray ionization (Bhattacharjee et al., 2020), live single-cell mass spectrometry (Masuda et al., 2018), and the spatial metabolomics pipeline (Geier et al., 2020).

In addition to soluble metabolites, plants can diffuse Volatile Organic Compounds (VOCs).Metabolomics is essential to uncover the role of these signalling molecules and their modulation in response to environmental stimuli and genotype interactions. However, their role in soil matrices could be less abundant and relevant than in aerial open-air environments.

Metabolomic analyses have shown that plants can influence their microbiota by secreting various metabolites. In turn, the microbiome can influence the metabolome of the host plant (Haichar et al., 2008).

It is now clear that to acquire global information on the interconnections existing among plants and the microbiota, single omics technologies should be integrated into a multi-omics approach (Chen et al., 2021). To this end, the development of bioinformatic tools and networking models that can integrate and visualize information is essential. This will provide a comprehensive view of the regulatory network connecting all the molecular levels from the genome to the metabolic pathways. By employing a multi-omics approach, it will be possible to deepen our knowledge on the complex interactions between plants and their growth-promoting microbial counterparts. This will be fundamental to understand how to engineer microbial communities and plants for a more sustainable agriculture. In this scenario it is fundamental to develop high-throughput phenotyping platforms, to measure and analyze qualitative and quantitative traits on a large scale, developing suitable phenomics approaches, that could be non -invasive and able to work in the field as well, in order to fill the gap with other omics techniques (Großkinsky et al., 2015).

6 The role of endophytes in protection from abiotic stresses

Plants, as sessile organisms, face continuous exposure to environmental stresses. These include both biotic factors, such as pathogens, pests and herbivores, and abiotic factors, such as heat, cold, drought, salinity, waterlogging, heavy metal toxicity, nutrient deficiency, and oxidative stresses (Cramer et al., 2011). Climate change has been increasing the negative effects of these abiotic stresses, leading to both faster events of severe stresses (e.g., flash droughts (Pendergrass et al., 2020)), and to more prolonged periods of stress, with several negative impacts on plant growth and productivity, up to more than -50% (Mahajan and Tuteja, 2005; Lohani et al., 2020; Sandrini et al., 2022).

Abiotic stress like heat or cold extremes can cause changes in membrane fluidity and protein structure, while the presence of salt or heavy metals in the soil can alter the physiological processes of enzymes and molecular interactions (Zhang et al., 2022). Salinity can also negatively impact the photosynthetic components, reducing the assimilation of CO_2 and the absorption of light. This, in turn, can lead to an increase of ROS and oxidative stress (Ma et al., 2020). It is important to consider that heat, drought, and salt stress are commonly present together, exacerbating the detrimental effects on plants. To sense and respond to abiotic stresses, plants have evolved multiple complex mechanisms, which have been extensively reviewed in the last decades (Zhang et al., 2022).

Considering this negative scenario, plant-associated microorganisms appear to be promising allies for modern agriculture to face climate change (Figure 2; Table 1). For example, many rhizobacteria produce osmoprotectants in the presence of stress conditions, while other bacteria, like *Pseudomonas* spp., produce ExoPolySaccharides (EPS) to increase water retention in case of drought stress (Rathinasabapathi, 2000; Grover et al., 2011). The mechanisms of both interactions and potential advantages exploitable in agriculture are reviewed for each stress type.

B).

6.1 Temperature

Temperature changes are not necessarily harmful to plants; they can play a crucial physiological role in regulating internal clocks and controlling processes like the opening and closing of flower corollas. Some species require exposure to low temperatures to initiate important developmental processes, such as vernalization for flowering or germination (Ruelland and Zachowski, 2010). However, temperatures that are too low can induce a range of physiological responses that can be detrimental to their survival. Chilling, intended as a few degrees above 0°C air temperature, can cause reductions in enzymatic activity, rigidification of membranes, destabilization of protein complexes, and stabilization of RNA secondary structure, while also promoting the accumulation of ROS. Additionally, chilling can impair photosynthesis and increase membrane permeability, resulting in the leakage of cellular contents. Freezing (below 0°C) stress, on the other hand, can cause more severe damage, as ice formation within cells leads to mechanical disruption and cell/tissue/plant death (Ruelland and Zachowski, 2010).

There are not many reports available in literature about the protective effects of endophytes against the low temperature stresses, perhaps because such conditions also limit the growth and multiplication of microorganisms. As an example of protective effects towards cold stress, the endophytic fungus *Epichloe gansuensis* increases the biosynthesis of alkaloids and unsaturated fatty acids during the seed germination of Drunken horsegrass (*Achnatherum inebrians*), thereby increasing tolerance to cold stress (Chen et al., 2016). It was also reported that the endophytic rhizobacterium *Parabulkholderia. phytofirmans* PsJN induced the

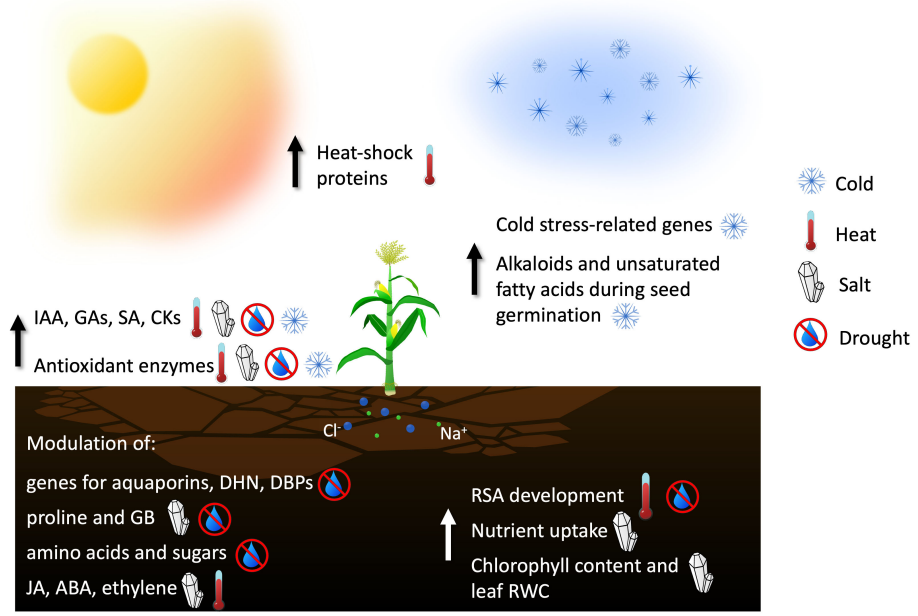


FIGURE 2
Main contributions of endophyte microorganisms in enhancing plant tolerance to abiotic stresses through increasing the synthesis of antioxidant molecules and heat-shock proteins, modulating the amount of phytohormones, or stimulating the development of the root structure (IAA, Indole-3-Acetic Acid, indicating auxins in general; GAs, Gibberellins; SA, Salicilic Acid; CKs, Citokinins; DHN, Dehydrin family of proteins; DBPs, Dehydration responsive element Binding Proteins; GB, Glycine Betaine; JA, Jasmonic Acid; ABA, Abscissic Acid; RSA, Root Structure Architecture; RWC, Relative Water Content).

upregulation of some cold stress-related genes in grapevine (Theocharis et al., 2012; Chen et al., 2021).

Heat stress is defined as a temperature rise of 10-15°C above ambient, with the optimal range for plant growth about 15-24°C. Among all the different abiotic stress factors, it has the most detrimental effects on plants, reducing flower fertility and modifying crop growth with a detrimental final effect on yield (Shaffique et al., 2022). Its effects include an increase in membrane fluidity, the formation of ROS, and alterations to photosynthesis and respiration processes (Banerjee and Roychoudhury, 2018; Shaffique et al., 2022). Heat stress triggers a cascade of physiological responses that result in the release of Heat Shock Proteins (HSPs), a class of molecular chaperones that facilitate protein folding and prevent aggregation under conditions of cellular stress. These proteins can assist unfolded proteins in refolding into their proper conformation or direct them towards degradation through ubiquitination processes (Ruelland and Zachowski, 2010; Banerjee and Roychoudhury, 2018). The often simultaneous presence of heat and drought stress exhibits holistic features, as the combined effects are greater than those caused by each stress individually (Lipiec et al., 2013).

The inoculation of the endophytic bacterium *Brevibacterium linens* RS16 in rice plants reduced the emission of the plant stress hormone ethylene due to its 1-aminocyclopropane-1-carboxylate (ACC) deaminase activity. Plants inoculated with *B. Linens* RS16 also showed increased levels of small HSPs (Choi et al., 2022).

A recent report demonstrates that the application of the plant growth-promoting root endophyte *Paraburkholderia phytofirmans* PsJN enhances the development of the Root System Architecture (RSA) in *Arabidopsis thaliana*, under both normal and high-temperature conditions. This allows the plant to access a larger soil area, thereby better managing abiotic stresses such as heat and drought (Macabuhay et al., 2022).

The simultaneous presence of heat and salinity stress can greatly impact crops. The inoculation of the endophytic fungus *Trichoderma virens* SB10, along with Glycine Betaine (GB) treatments, conferred significant tolerance in soybean (*Glycine max* L.) plants against these two stresses. In presence of GB, *T. virens* SB10 enhanced the production of hormones like gibberellins, IAA, and SA. The co-treatment with the fungus and GB also led to a reduction in proline accumulation and Na^+ uptake and an increase in macronutrient (N, Ca, K) uptake. Effects on *GmHKT1* and *GmSOS1* gene expression, two major genes involved in salt tolerance (Singh and Roychoudhury, 2021), were also recorded, leading to the maintenance of a high K^+/Na^+ ratio. Treated plants exhibited higher growth rates and an increase in antioxidant activities due to the upregulation of Ascorbate PeroXidases (APX), SuperOxide Dismutases (SOD), PerOXidases (POD) and reduced Glutathione (GSH) enzymes (Bilal et al., 2023).

6.2 Drought and salinity

These two stresses represent the main abiotic stress factors that limit crop production globally (Trenberth et al., 2014; Rodriguez and Durán, 2020).

TABLE 1 Overview of mechanisms of protection of plants from abiotic stresses by beneficial endophytes.

A)

Endophyte Species	Host plant	Increased Tolerance to Stress	Mechanisms	References
		Single Stress		
Epichloe gansuensis	Drunken horsegrass (*Achnatherum inebrians*)	Cold	Increase alkaloids biosynthesis and unsaturated fatty acids during seed germination	Chen et al., 2016
Piriformospora indica	Maize (*Zea mays* L.)	Drought	enhance the expression of genes involved in the drought stress response of maize hosts by increasing auxin, ABA, SA, and cytokinin levels	Zhang et al., 2018
Trichoderma harzianum	Rice (*Oryza sativa*)	Drought	Modulate activity of genes for aquaporin and dehydrin, dehydration responsive element binding protein, and SuperOxide Dismutase	Pandey et al., 2016
Piriformospora indica	Barley (*Hordeum vulgare*)	Drought	Increase the production of high temperature stress-responsive proteins	Ghaffari et al., 2019
AMF culture (mainly *Rhizophagus intraradices*; *Funneliformis mosseae*; *F. geosporum*)	Wheat (*Triticum aestivum*)	Drought	Increase of reachable soil water thanks to hyphae	Mathur et al., 2019
Porostereum spadiceum AGH786	Soybean (*Glycine max*)	Salt	Decrease of JA and ABA, increase of GA3	Hamayun et al., 2017
Yarrowia lipolytica FH1	Maize (*Zea mays* L.)	Salt	Secretion of exogenous IAA and regulation of endogenous IAA and ABA. Effects on production of peroxidase, catalase and proline	Gul Jan et al., 2019

B)

Species	Host plant	Increased Tolerance to Stress	Mechanisms	References
		Single Stress		
Burkholderia phytofirmans PsJN (PGPR)	Grapevine (*Vitis vinifera* L.)	Cold	Induction the up-regulation of cold stress-related genes	Theocharis et al., 2012
Brevibacterium linens RS16	Rice (*Oryza sativa* L.)	Heat	Reduction of ethylene levels, enhancement in GST expression, increased levels of small HSPs	Choi et al., 2022
Bacillus aryabhattai SRB02	Soybean (*Glycine max*)	Heat	Increase in phytohormones production, modification of stomata behaviour and root structure	Park et al., 2017
Streptomyces sp. PGPA39	Tomato (*Solanum lycopersicum*)	Salt	Decrease of ethylene through ACC deaminase production	Palaniyandi et al., 2014
		Multiple Stresses		
Paraburkholderia phytofirmans PsJN (PGPR)	Thale cress (*Arabidopsis thaliana*)	Heat and Drought	Enhancement in the development of RSA	Macabuhay et al., 2022
Bacillus spp.	Fennel (*Foeniculum vulgare* Mill.)	Drought and Salt	Increase of availability of P	Mishra et al., 2016

A): Fungi; B): Bacteria.

Drought is defined as a period when the available water is insufficient for an organism or environment to function at its best (Kamran et al., 2022). Drought stress represents one of the most critical threats to plant productivity, and thus to global food production, affecting all the stages of plant growth. It leads to a reduction in turgor pressure, affecting cell division, enlargement, and differentiation (Farooq et al., 2009, 2009).

Drought thus affects plants in many ways: typical symptoms include stunted plants, scorching, rolling, and yellowing of leaves, and permanent wilting (Seleiman et al., 2021). Moderate drought stress can induce modifications in the RSA and in the allocation of resources by the plant. In the case of severe drought stress, the roots shrink, and alterations occur at PhotoSystem II (PS II) (Ma et al., 2020).

Soil salinity is defined as the increased amount of sodium (Na^+) and especially chloride (Cl^-) ions in soils, resulting primarily from natural events such as weathering of parent rocks, seawater, or atmospheric deposition. Other than that, anthropogenic processes, such as poor drainage facilities, irrigation with brackish groundwater, unsuitable water management, and 'cultural' errors in irrigated agriculture, can increase soil salinity (Evelin et al., 2019). Salinity causes ionic imbalance and alters metabolic pathways in plant cells, like the synthesis of proteins and the function of some enzymes and ribosomes. Besides, Na^+ competes with other essential nutrients like phosphate, nitrate and potassium (Angon et al., 2022).

From several reports, it became evident that the plant microbiome can play a role in protecting against high salinity and drought (Yang et al., 2009; Rolli et al., 2015; Berg et al., 2016).

The endophytic fungus *Piriformospora indica* enhances the expression of genes involved in the drought stress response of maize hosts by increasing auxin, ABA (ABscissic Acid), SA, and cytokinin levels (Zhang et al., 2018). Also, *Trichoderma harzianum* was shown to improve drought tolerance in rice, by modulating the activity of genes for aquaporin and dehydrin, Dehydration responsive element-Binding Protein (DBP), and SOD (Pandey et al., 2016).

Symbiotic relationships between plants and endophytic fungi such as *Piriformospora indica* can enhance the adaptation of plants to drought stress by regulating amino acid and soluble sugar metabolism. For instance, *P. indica* was found to improve the adaptation of barley (*Hordeum vulgare* L.) to drought stress (Ghaffari et al., 2019). Soybean inoculated with the endophytic fungus *Porostereum spadiceum* AGH786 under saline conditions showed reduced effects of salinity. The endophyte caused decreasing levels of JA and ABA and increasing levels of GA3, leading to improved plant growth (Hamayun et al., 2017). Similarly, researchers observed a positive effect of the interaction between the endophytic fungus strain *Yarrowia lipolytica* and maize under saline conditions, which improved plant growth attributes such as leaf relative water content, levels of oxidative enzymes and chlorophyll content through the enhancement of metabolism and hormones (ABA and IAA) secretions (Gul Jan et al., 2019).

Endophytic microbes can alleviate the salt-generated oxidative stress in plants by activating genes for ion transporters, ROS scavenging, and by activating the production and signalling of phytohormones such as auxin, JA, and Ethylene (ET) (Brotman et al., 2013; Ghaffari et al., 2016; Qin et al., 2018; Eida et al., 2019).

Finally, seed bio-priming is a novel beneficial technique aiming to employ bio-stimulating agents like growth-promoting microorganisms to improve the physiological functioning of seeds and their stress resilience (Chakraborti et al., 2022). Two salt-tolerant endophytic fungi, *Paecilomyces lilacinus* KUCC-244 and *Trichoderma hamatum* Th-16 were used for bio-priming wheat and mung bean (*Vigna radiata* L.) seeds. Results showed that both endophytes, in particular *T. hamatum* Th-16, increased the growth and chlorophyll content of wheat and mung bean plants under extreme salinity conditions. The primed plants also exhibited increased activities of antioxidant enzymes and enhanced photosynthetic attributes (Irshad et al., 2023).

7 The role of endophytes in biotic stresses tolerance

Biotic stress occurs when the plant is damaged by phytopathogens, such as bacteria, fungi, viruses, insects, or herbivores that feed and thrive at the plant's expense. It is the primary cause of harvest losses, especially those caused by bacterial and fungal phytopathogens (Chaudhary et al., 2022). It is estimated that biotic stresses cause approximately 17-30% of global crop production loss (Muthu Narayanan et al., 2022).

Until now, the standard procedure to combat plant pathogens infection has been the use of chemicals. This mechanistic approach often does not consider any ecosystemic interaction. However, pesticides can be hazardous, and the occurrence of pesticide resistance is another significant factor to consider (Hawkins et al., 2019).

Against these types of stresses, a great deal of research activity and farm-scale applications of beneficial organisms are reported, much more than against other crop limitations, such as abiotic stresses. Biological control represents a promising strategy to manage plant pathogens sustainably. The first uses of insect parasites date back to the end of the 19th century (Hajek et al., 2007). Biological control involves applying either beneficial organisms, or substances produced by them, such as enzymes, phytohormones, and secondary metabolites, to alleviate the negative effects caused by pathogens and stimulate favourable reactions in the plant (Muthu Narayanan et al., 2022).

Endophytes can use direct mechanisms to exert their biocontrol effects against phytopathogens, including the production of siderophores, to limit the availability of metal ions to pathogens, or the synthesis of antifungal compounds, or competition for a biological niche. They can also counteract pathogens through indirect mechanisms, by inducing SAR and ISR in the host plant (Pandey et al., 2019; Kamle et al., 2020; Figure 3; Table 2).

7.1 Bacterial pathogens

Phytopathogenic bacteria are predominantly represented by the genera *Agrobacterium*, *Bacillus*, *Burkholderia*, *Clavibacter*, *Erwinia*, *Pantoea*, *Pseudomonas*, *Ralstonia*, *Streptomyces*, *Xanthomonas*, and *Xylella* (Muthu Narayanan et al., 2022). Bacterial diseases can be systemic or localized, with the most common symptoms in plants being galls and overgrowth, wilting, rot, scabs, necrosis, chlorosis, and blights (Nazarov et al., 2020).

Fungal metabolites produced by *Trichoderma harzianum* have shown strong antibacterial activity against *Ralstonia solanacearum*, a phytopathogenic bacterium that causes bacterial wilt disease in tomato (*Solanum lycopersicum* L.) plants (Yan and Khan, 2021).

Endophytic actinobacteria were isolated from Chilean native potatoes (*Solanum tuberosum* subsp. *tuberosum* L.) and they were demonstrated to act against *Pectobacterium carotovorum* subsp. *Carotovorum* and *P. atrosepticum*, bacterial pathogens that cause tissue maceration symptoms in potato tubers. One of the isolates,

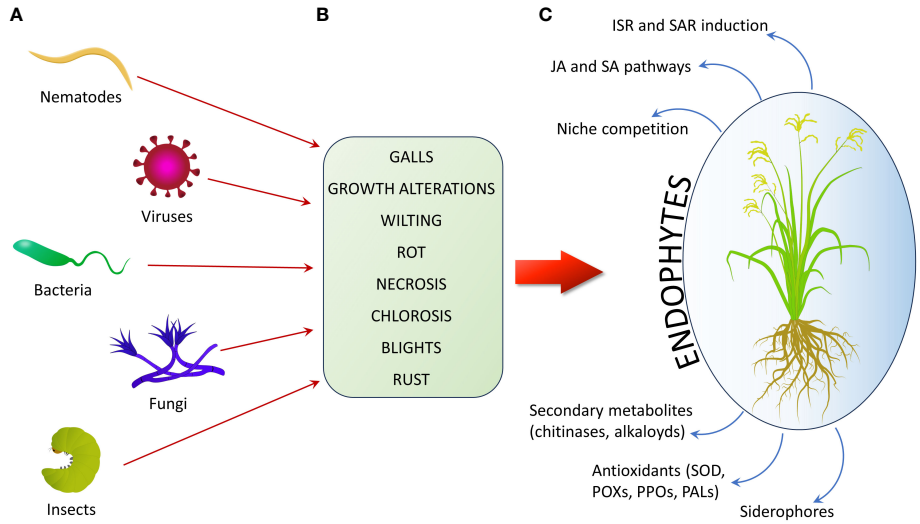


FIGURE 3
Schematic example of plant damages caused by phytopathogens and relative mechanisms by which endophytes can help crops contrasting pests. **(A)** Main groups of plant pathogens. **(B)** Principal types of damages caused by phytopathogens. **(C)** Main plant defense responses enhanced by endophytes to contrast plant parasites (ISR, Induced Systemic Resistance; SAR, Systemic Acquired Resistance; JA, Jasmonic Acid; SA, Salicylic Acid; SOD, Super-Oxide Dismutase; POX, PerOXidase; PPO, PolyPhenol Oxidase; PAL, Phenylalanine Ammonia Lyase).

Streptomyces sp. TP199, was found to inhibit the growth of *Pectobacterium* sp., reducing tissue maceration symptoms. Moreover, strain TP199 showed metal-dependent Acyl Homoserine Lactones (AHL) quorum quenching activity, which can inhibit communication between bacterial cells (Padilla-Gálvez et al., 2021).

7.2 Fungal pathogens

The most common plant pathogenic fungi are *Alternaria* spp., *Aspergillus* spp., *Colletotrichum* spp., *Fusarium* spp., *Phytophthora* spp., *Pythium*, and *Pyricularia* spp., while anthracnose, dieback, gall, powdery mildew, blight, rust, rot, wilt, and smut are examples of diseases caused by these fungal phytopathogens (Muthu Narayanan et al., 2022).

Some *Bacillus* strains significantly increased antioxidant enzymes like SOD, PerOXidase, PolyPhenol Oxidase and Phenylalanine Ammonia Lyase in leaves and roots of the rice plant, contrasting the fungus *Pyricularia oryzae*. They also secreted different biocontrol molecules such as proteases, glucanases, and siderophores in the rice rhizosphere (Rais et al., 2017). In wheat, *Fusarium graminearum* is the cause of Fusarium head blight as well as Fusarium foot and root rot. Colombo and colleagues (Colombo et al., 2019) studied the biocontrol activity of *Streptomyces* spp. on *F. graminearum* in spring wheat and one strain, DEF09, effectively inhibited FHB under controlled and field conditions by blocking the spread of the pathogen at the infection site.

Endophytes with biocontrol potential against *Rhizoctonia solani*, a fungal pathogen causing sheath blight disease in maize, were isolated from *Stevia rebaudiana* plants. Three bacterial isolates, identified as *Ochrobactrum ciceri* SR1EB1, *Achromobacter spanius* SR1EB11 and *Bacillus licheniformis* SR2EB5, showed growth inhibition effects against *R. solani* (Vyas and Singh, 2023).

Furthermore, the gram-positive bacterium *Micromonospora*, isolated from nitrogen-fixing nodules of leguminous plants, showed biocontrol effects on fungal pathogens and stimulation of plant immunity in tomato. Root inoculation with *Micromonospora* strains showed reduced infection from the fungal pathogen *Botrytis cinerea*, and investigations on defence mechanisms highlighted an increased induction in JA-related defence pathways (Martínez-Hidalgo et al., 2015).

7.3 Viral pathogens

Plant viruses are globally diffused plant pathogens, obligatory parasites due to their need for replication within plant cells. Plant viruses are primarily RNA viruses, and the ones considered most important are Tobacco Mosaic Virus (TMV), belonging to the family *Virgaviridae*, Tomato Spotted Wilt Virus (TSWV) (*Tospoviridae*), Tomato Yellow Leaf Curl Virus (TYLCV) (*Geminiviridae*), Cucumber Mosaic Virus (CMV) (*Bromoviridae*), Potato Virus Y (PVY) (*Potyviridae*), Cauliflower Mosaic Virus (CaMV) (*Caulimoviridae*), African Cassava Mosaic Virus (ACMV) (*Geminiviridae*), Plum Pox Virus (PPV) (*Potyviridae*), Brome Mosaic Virus (BMV) (*Bromoviridae*), Potato Virus X (PVX) (*Alphaflexiviridae*) (Scholthof et al., 2011). Symptoms of viral diseases in plants include growth suppression, deformation, discoloration, necrosis, and impaired reproduction (Nazarov et al., 2020). Investigating the contribution of microbiota towards viral infection (Zhou et al., 2015), identified two new butyrolactones (aspernolides C and D) along with two previously known butyrolactones (aspernolides A and B) from a culture of the endophytic fungus *Aspergillus versicolor*. When tested against

TABLE 2 Overview of mechanisms of protection of plants from biotic stresses (plant pathogens, pests, and parasites) by beneficial endophytes.

A)

Pathogen/Pest	Endophyte species	Host Plant	Effects	Mechanisms	References
Bacteria					
Ralstonia solanacearum	*Trichoderma harzianum*	Tomato (*S. lycopersicum* L.)	Destruction of bacterial cells	Release of fungal metabolites	Yan and Khan, 2021
Viruses					
Tobacco Mosaic Virus (TMV)	*Aspergillus versicolor*	Tobacco (*Nicotiana glutinosa* L.)	Moderation of anti-TMV activity	Synthesis of aspernolides C and D butyrolactones	Zhou et al., 2015
Maize Chlorotic Mottle Virus (MCMV) and SugarCane Mosaic Virus (SCMV)	*Trichoderma harzianum; Metarhizium anisopliae*	Maize (*Zea mays* L.)	Reduction of the pathogenic effects of SCMV (*T. harzianum*); Decrease in titer of SCMV (*M. anisopliae*); No evident effects on MCMV	Not yet defined, but possibly due to activation of defence-related genes and the interference exerted by endophytes on virus movement	Kiarie et al., 2020
Pests/Parasites					
Meloidogyne javanica (Nematode)	*Trichoderma harzianum* BI	Tomato (*S. lycopersicum* L. var. Roma VF)	Reduction of nematode eggs hatching, increase of antioxidant enzymes	Penetration inside the nematode egg mass matrix	Sahebani and Hadavi, 2008
Meloidogyne incognita (Nematode)	*Trichoderma harzianum* T-78	Tomato (*S. lycopersicum* L.)	Reduction of nematode eggs clusters, delay in the development of eggs	Enhancement of JA- and SA-regulated defences	Martínez-Medina et al., 2017
Insects and herbivores	*Clavicipitaceae*	(The paper does not focus on a specific species of plant)	Feeding deterrence, delayed development, increased mortality	Production of bioactive alkaloids like ergot, indole-diterpenes, lolines, peramine	Panaccione et al., 2014

B)

Pathogen	Endophyte species	Host Plant	Effects	Mechanisms	References
Fungi					
Botritys cinerea	*Micromonospora* spp.	Tomato (*S. lycopersicum* L.)	Reduction in the infection rates	Increased induction of JA-related pathways	Martínez-Hidalgo et al., 2015
Pyricularia oryzae	*Bacillus* spp.	Rice (*Oryza sativa* L.)	Increase of antioxidant response, reduction of blast disease symptoms	Increased synthesis of antioxidant enzymes, secretion of proteases, glucanases, siderophores	Rais et al., 2017
Fusarium graminearum	*Streptomyces* strain DEF09	Spring wheat (*Triticum aestivum* L.)	Inhibition of fungal spreading from the infection site	Chitin degradation, synthesis of antifungal molecules	Colombo et al., 2019
Rhizoctonia solani	*Ochrobactrum ciceri* SR_1EB_1; *Achromobacter spanius* SR_1EB_{11}; *Bacillus licheniformis* SR_2EB_5	Stevia (*Stevia rebaudiana* Bertoni)	Growth inhibition of hyphae	Stripping of fungal hyphae and accumulation of debris	Vyas and Singh, 2023
Bacteria					
Pectobacterium sp.	*Streptomyces* sp. TP199	Potato (*S. tuberosum* subsp. *Tuberosum* L.)	Reduction of tubers tissue maceration	Synthesis of antimicrobial compounds, interference on communication signals	Padilla-Gálvez et al., 2021
Pests/Parasites					
Meloidogyne javanica (Nematode)	*Streptomyces* strain SA	Banana (*Musa acuminata* AAA Cavendish)	Inhibition of nematodes	Synthesis of antibiotics effective against nematodes (avermectin, nanchangmycin, milbemycin), niche competition	Su et al., 2017

A): Fungi; B): Bacteria.

viruses, both aspernolides C and D exhibited moderate anti-TMV activity. Similarly, Kiarie et al. (2020) tested the ability of fungal endophytes to contrast Maize Lethal Necrosis (MLN), a serious disease affecting maize crops in eastern Africa. This disease is caused by the co-infection of maize plants with Maize Chlorotic Mottle Virus (MCMV) (*Tombusviridae*) and Sugarcane Mosaic Virus (SCMV) (*Potyviridae*). Maize plants inoculated with *T. harzianum* and *Metarhizium anisopliae* showed reduced severity and titer of SCMV, respectively, indicating their potential to induce resistance against SCMV. However, no significant effect was observed on the MCMV.

7.4 Insects and herbivores

Invasive insects cause at least $70 billion in crop losses every year (Bradshaw et al., 2016). Panaccione et al. (2014) summarized a series of studies on the role of bioactive alkaloids produced by endophytic fungi in protecting plants against herbivores. Four major types of bioactive alkaloids (ergot alkaloids, indole-diterpenes, lolines and peramine) are produced by fungi from the *Clavicipitaceae* family. Symbioses between plants and these endophytes have significant effects on both insects and mammalian herbivores, largely due to the production of these bioactive alkaloids. Ergot alkaloids contribute to herbivore resistance, also affecting nematodes. They also act through feeding deterrence, delayed development, and increased mortality of insects.

Loline alkaloids exhibit insecticidal and feeding-deterrent activity. Lolines are often present in grasses with fungal endophytes of the genera *Epichloë* and *Neotyphodium* (Wilkinson et al., 2000). Genetic analyses to determine whether the production of lolines in plants is active against aphids highlighted that the endophyte *Epichloë festucae* showed heritable variation in the expression of loline alkaloids. Analyses on Lol+ (alkaloid expression) and Lol- (no expression) linked alkaloid expression to activity against aphids, and the levels of alkaloids in the plants were correlated with the level of anti-aphid activity (Wilkinson et al., 2000).

Peramine is the most widely distributed of the four classes of *epichloae*-derived secondary metabolites. It is another alkaloid that acts as a strong feeding deterrent for different insects. Peramine is water-soluble and is dispersed throughout the plant (Rowan, 1993; Panaccione et al., 2014).

Additional information about the role of Endophytic EntomoPathogenic Fungi (EEPFs) was recently made available by (Samal et al., 2023).

On the other hand, different endophytes could be used not only to prevent herbivore damage in plants but, in some cases, to favour this phenomenon for domestic herbivores, reducing undesirable molecules present in plants for livestock nutrition (Bluett et al., 2005a, 2005b).

7.5 Nematodes

Nematodes are small, non-segmented invertebrates. They are the most abundant animals on Earth and are fundamental to the soil-food web (Gamalero and Glick, 2020). The phylum *Nematoda* comprises more than 30,000 species (Bernard et al., 2017), classified into five groups: bacterivores, fungivores, herbivores, omnivores, and predators (van den Hoogen et al., 2020).

Nematodes include the so-called Plant-Parasitic Nematodes (PPN), among which some of the most important are root-knot nematodes (*Meloidogyne* spp.), cyst nematodes (*Heterodera* spp. and *Globodera* spp.), and root-lesion nematodes (*Pratylenchus* spp.) (Kumar and Dara, 2021). Over 4,100 species of PPNs have been identified, causing an estimated $80–$118 billion dollars per year of damage to crops (Bernard et al., 2017). PPNs can damage the host plant through a needle-like oral structure called stylet, used to release specific enzymes inside plant tissues (Pulavarty et al., 2021, p. 202), and the group of root-knot nematodes develop root knots by forming a complex of multinucleate hypertrophied giant cells, which cause visible knots or galls at the root level (Martínez-Medina et al., 2017). More detailed information about nematodes, their characteristics, and modes of action can be found in Jones et al. (2011).

The main way to fight PPNs is to use chemical nematocides, but these are expensive and harmful to the environment, and EU regulations are constantly reducing the nematocides available for agriculture (Poveda et al., 2020). Therefore, the biocontrol of nematode infection is becoming a promising possibility.

Endophytes isolated from banana (*Musa acuminata* AAA Cavendish) plant roots infected with *Meloidogyne* spp. were tested against *Meloidogyne javanica*, and one strain, named SA and identified as *Streptomyces* spp., showed an inhibition rate of more than 50% *in vitro* and a biocontrol efficiency of more than 70% in sterile soil against the nematode (Su et al., 2017).

Fungi could also represent a valuable source of biocontrol agents against plant-infecting nematodes. The fungus *Trichoderma harzianum* strain BI was used against *M. javanica* (Sahebani and Hadavi, 2008), reducing the infection rates of the nematode through penetration inside the nematode egg mass matrix, leading to reduced hatch levels. *T. harzianum* BI also increased the activity of resistance-related enzymes in plants, such as POX, PPO, and PAL. Further investigation showed that chitinase activity in *T. harzianum* BI culture filtrates increased in the presence of colloidal chitin and nematode eggs, implying its potential for degradation of chitin present in nematode eggs.

The root endophyte strain *T. harzianum* T-78 was used to study the protective effects on tomato plants against the root-knot species *M. incognita* (Martínez-Medina et al., 2017) using a split-root system, in which the two halves of the plant roots were allowed to grow in two different pots, one for the treatment and the second as a control. T-78 inoculation decreased the amount of root galls, and a significant reduction in the number of nematode egg clusters was observed in systemic tests. Moreover, the expression profile of the SA-responsive marker genes Pathogenesis-Related protein 1a (PR1a) and Pathogenesis-Related protein P6 (PR-P6) was higher in T-78-pretreated plants infected with the nematode and SA concentrations were higher compared with the non-pretreated ones. Finally, the expression analysis of the JA-responsive genes Proteinase Inhibitors II (PI II) and MultiCystatin (MC) after *M. incognita* infection showed that JA signalling was down-regulated in plants not inoculated with T-78, while in tomato plants pre-

inoculated with T-78, the inhibition in genes PI II and MC showed a significant reduction.

More specific reviews are available to deepen the knowledge about the use of endophytes as biocontrol agents against nematodes (e.g., Gamalero and Glick, 2020; Kumar and Dara, 2021).

8 Role of endophytes in nutrition and quality of the final products

Plant–microbe interactions play a crucial role in improving soil nutrition and enriching micronutrients through metal solubilization, mobilization, and translocation to different parts of the plant. Micronutrient deficiency, also known as "hidden hunger", threatens the health of billions of people worldwide, particularly in developing countries. Additionally, the intensification of crop production is causing a gradual depletion of micronutrients in agricultural soils, compromising the nutritional value of food. Iron and zinc deficiencies are widespread in the human diet, leading to several malnutrition symptoms. To overcome these deficiencies, biofortification, the process of increasing the bioavailable concentrations of essential elements in the edible portion of crop plants, is commonly achieved through plant breeding and agronomic practices. Microbial communities can be exploited as a valid alternative due to their ability to increase metal solubilization in the soil and enhance their mobilization to the plant parts. This is achieved through the production of siderophores or other chelating factors, upregulation of Zn and Fe transporters, acidification of the rhizosphere through organic acid secretion and proton extrusion, reduction of anti-nutritional factors (e.g. phytic acids), and secretion of phenolics or phytohormones like signalling molecules (Singh and Prasanna, 2020). Both bacteria and fungi have demonstrated a positive effect in improving mineral contents in the edible parts of plants, although a major representation of mycorrhizal fungi underlines the importance of this category of endophytes for supporting plant nutrition.

Bacterial endophytes have proven to be effective in the biofortification of wheat grains with Zn (Ramesh et al., 2014; Abaid Ullah et al., 2015; Singh et al., 2017). Wang et al. (2014) found a positive influence of rice inoculation with an endophyte recovered from Zn hyper-accumulator *Sedum alfredii* on the bioavailability of Zn in the soil and its accumulation in rice grains. Vaid et al. (2014) tested the effect of zinc-solubilizing bacteria on growth promotion and zinc nutrition of rice: bacterial inoculations significantly enhanced the total Zn uptake as well as grain methionine concentration, besides increasing the mean of many agronomic traits, like dry matter yield, productive tillers/plant, number of panicles/plants, number of grains/panicle, grain and straw yield. The screening of 129 strains of endophytic bacteria from maize stem and leaves showed that 24.5% of these isolates were siderophore producers, 14% could solubilize insoluble Zn compounds and 33% of them had phytase activity (Verma et al., 2022). Rana et al. (2012) reported a significant increase in Fe, Mn, and Cu content in wheat grains upon inoculation with the bacterial strain *Providencia* spp. isolated from the wheat rhizosphere. The screening of a set of 213 endophytes from several wheat genotypes allowed Singh et al. (2018) to identify promising endophytes for Zn and Fe accumulation in wheat grains. At the same time, in grains after endophyte inoculations, phytic acid, an anti-nutritional factor, was significantly decreased.

Subramanian et al. (2013) reported that the inoculation of maize plants with AM fungi improved the availability of micronutrients in soils, particularly Zn, as a consequence of rhizospheric acidification and siderophore production, and produced grains with 10-15% higher Fe and Zn contents, while the anti-nutritional factor phytic acid decreased. In wheat, the application of a consortium of AMF resulted in an increase of micronutrients (Cu, Fe, Zn, Mn) and macronutrients (N, P, K) content in the grain (Mäder et al., 2011). AMF were also able to improve selenium (Se) level in the grain, alone or in association with selenobacteria (Duran et al., 2013). Tang et al. (2022) focused on the effects of endophytic fungus *Phomopsis liquidambaris* on the absorption and distribution of 14 essential mineral elements in the vegetative organs and in grains of rice: the results indicated that *P. liquidambaris* significantly increased the accumulation of N, P, Fe, Mn, Zn, Mo, and Se in rice grains, accompanied by a significant increase in yield and protein content. AM fungal inoculation was also effective in improving the nutritional value of chickpea (*Cicer arietinum* L.) grain by protein, Fe and Zn grain biofortification (Pellegrino and Bedini, 2014).

The interactions of endophytic fungi with plant tissues can also boost secondary metabolite production, resulting in the development of several bioactive compounds. In lettuce (*Lactuca sativa* L.), AM fungi, in addition to increasing fresh weight, improved the ascorbate level (Baslam et al., 2011). In spinach (*Spinacia oleracea* L.), they augmented the concentration of total phenolic compounds, flavonoids and phenolic acids (Khalid et al., 2017). In tomato, AMF inoculation, in addition to increasing fruit N, P, and Cu concentration, allowed for higher antioxidant concentration and carotenoid contents (Hart et al., 2015), while in strawberry (*Fragaria x ananassa* var. Selva) increased concentrations of anthocyanins (Lingua et al., 2013) and of sucrose, glucose, and two vitamins, ascorbic and folic acid (Bona et al., 2015).

Heavy metals contamination of agricultural soils is an important issue all around the world, posing serious risks to food safety. Indeed, although they are not essential elements for a plant's life, crops uptake heavy metals in soils by root systems, they transport them to aerial parts through the xylem and the phloem, and accumulate them in edible parts, threatening the food chain, and ultimately human health. Under natural conditions, heavy metals in soils originate geologically; however, their amounts are continually increased in soils by anthropogenic sources, i.e., atmospheric deposition of Particulate Matter (PM) from industrial activity and transportation, agricultural activity, such as wastewater irrigation, the application of pesticides and fertilizers (Shi et al., 2018). Among heavy metals, cadmium (Cd) and arsenic (As) are the major contaminant in agricultural soils. Cadmium, with a biological half-life of 10-30 years, has been classified as a potent human carcinogen. Endophytes and AM fungi are involved in alleviating metal toxicity to the host plant. Bacteria evolved various mechanisms to avoid heavy metal stress including: (a)

transport of metals across the cytoplasmic membrane; (b) biosorption and bioaccumulation to the cell walls; (c) metal entrapment in the extracellular capsules; (d) heavy metals precipitation; and (e) metal detoxification via oxidation–reduction (Zubair et al., 2016). Several mechanisms have been hypothesized for AMF-mediated detoxification of heavy metals, including (i) bound to cell wall and deposit in the vacuoles of AMF, (ii) sequestration by the help of siderophores in the soil or into root apoplasm, (iii) bound to metallothioneins or phytochelatins inside the fungal or plant cells, and (iv) transporters at the tonoplast of both plants and fungi catalyse the transport of metals from the cytoplasm and allow their compartmentalization into vacuoles (Jan and Parray, 2016).

Among cereals, rice plants tend to accumulate more Cd than others, and this is of particular concern in the larger rice-growing areas, where populations are relying on rice for most of their caloric intake (Hu et al., 2016). Zhou et al. (2021) identified an endophytic bacterium from Cd-contaminated soil capable of promoting rice growth and reducing Cd concentration in rice grain under Cd-contaminated conditions.

9 Conclusions and future perspectives

With a focus on endophytes, we have only scratched the surface of the enormous amount of data and information that has been produced by the scientific community worldwide, regarding the complex interactions between plants and microorganisms. Despite significant progress, many challenges remain, both from the technical and the legislative sides.

From a technical point of view, a deep understanding of the biology, way of action, and main interactions of the endophytes in the complex system of microorganism-plant-environment is necessary to optimize their usage in a one-health vision. For example, it is necessary to overcome the technical limits due to the very low fraction of culturable microorganisms inhabiting the soil, the difficulty in maintaining a stable inoculated microbial symbiont in the soil, and in increasing the plant's aptitude to associate with beneficial microorganisms. Related to this last point, it is relevant to deepen the knowledge related to processes involved in the establishment of successful associations, namely in the attraction, recognition and colonization steps, in order to allow knowledge-driven interventions. Recent investigations uncovered, from the plant side, sorghum, tomato and barley loci affecting microbiome composition (Deng et al., 2021; Escudero-Martinez et al., 2022; Oyserman et al., 2022), thus opening the way for exploiting host-genetics to manipulate and select the crop microbiota (Escudero-Martinez and Bulgarelli, 2023), while from the microbial side new approaches of microbiome engineering that boost the positive impact of the associations are emerging (Arif et al., 2020; Nerva et al., 2022). Increasing the available targets for improving endophytic colonization by plant growth promoting microorganisms rise the possibility that in the near future it will be possible to improve plant association with beneficial microbiota through plant breeding and microbiome engineering approaches. Along these concepts is the realization of knowledge-based synthetic microbial communities (SynComs) or Artificial Microbial Consortia (AMC), that generate a defined microbial system with known taxonomic and functional profiles, thus containing multiple functions for plant growth promotion (Arif et al., 2020; Nerva et al., 2022). This can potentially solve some of the drawbacks of traditional microbial biofertilizers, such a host incompatibility, ineffective competitiveness with indigenous microbes, and inadaptability to the local environment.

Field studies provide the most natural conditions for exploring the roles of endophytes. However, the strong impact of environmental factors makes them highly variable, and often take to unpredictable results. Moreover, agricultural systems and systems intensively used by humans are often characterized by a shift (often a reduction) in microbial diversity; and this may also be extended to plants raised in pot experiments, where we expect a reduced diversity or altered structure of the microbiota (Berg et al., 2016), that pose other limits to their extension to agricultural field systems.

When transitioning from controlled experiments to field applications of potential microbial formulations, the selected strain would have to interact with naturally occurring soil microbes and endophytes. The administered endophytes should be able to colonize a broad spectrum of crops and they should acquire a niche in the plant habitat, avoiding at the same time possible negative alterations of the ecosystem. Additionally, it is also crucial to consider the proper formulation to maximize the beneficial impact on crops, while reducing costs and number of applications (Verma et al., 2021).

Another element of concern is the ability of some endophyte to become a pathogen or produce toxins. For example, fungal endophytes from genera *Fusarium*, *Alternaria*, or *Aspergillus*, possess qualities as PGPM, but are also mycotoxins producers (Stranska et al., 2022).

From a legislative point of view, in the translation from research to application, in each area of the world, the categorizations and rules imposed by current and future legislation on the microbial compounds must be taken into great account for their deployment in the real world. For example, in Europe beneficial microorganisms are divided in two main categories, owing to the target. The microbes that act against biotic targets (e.g. pathogens, pests) are defined as Biological Control Agents (BCA), are enclosed in the plant protection products, and are ruled by the EC Regulation no. 1107/2009.The microorganisms whose target is the mitigation of an abiotic stress (e.g. freezing temperatures, salinity) are defined as Microbial Plant Biostimulants (MPB), and are ruled by EC Regulation no. 1009/2009, within the category PFC 6(A).

The rapidly changing climate presents a complex and daunting challenge that requires urgent solutions, with anthropogenic causes of pollution and environmental degradation continuing to worsen rapidly despite warnings from experts. Even though association of plant with beneficial microorganisms is demonstrated to protect plants from a changing environment, it should also be considered that diverse environmental conditions, including climate changes, can produce currently unpredictable outcomes on the interactions between host plant and endophytic microbiota.

However, there is hope for a better future. As is normal in research, big changes start from small discoveries. The approaches we have explored in this review offer potential solutions to

counteract the negative consequences of environmental stressors. Fortunately, ever innovating omic techniques and the ever-expanding set of genomic technologies offer powerful tools to help researchers gain deeper insights into the complex relationships between plants and their microbial partners. By using these tools, together with beneficial endophytes, to develop more eco-friendly and efficient agronomic practices, we can work towards a more sustainable future for our planet.

Author contributions

LS: Conceptualization, Investigation, Visualization, Writing – original draft, Writing – review & editing. EM: Writing – original draft, Writing – review & editing. GV: Conceptualization, Supervision, Visualization, Writing – original draft, Writing – review & editing. PV: Writing – original draft, Writing – review & editing. NP: Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

GV and EM would like to acknowledge the AGER Project "Enabling the potential of the unexplored: exploiting tailored microbial consortia to enhance environmental, societal and economic sustainability and resilience of Italian agro-ecosystems - Micro4Life" (Rif. 2022-2903). All authors wish also to thank Dr. Valerio Battaglia of CREA, Centro di Ricerca Cerealicoltura e Colture Industriali, for useful discussions about legislative aspects of the use of beneficial microorganisms in agriculture.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abaid Ullah, M., Hassan, M., Jamil, M., Brader, G., Shah, M., Sessitsch, A., et al. (2015). Plant growth promoting rhizobacteria: an alternate way to improve yield and quality of wheat (Triticum aestivum). *Int. J. Agric. Biol.* 17, 51–60.

Abbass, K., Qasim, M. Z., Song, H., Murshed, M., Mahmood, H., and Younis, I. (2022). A review of the global climate change impacts, adaptation, and sustainable mitigation measures. *Environ. Sci. pollut. Res.* 29, 42539–42559. doi: 10.1007/s11356-022-19718-6

Ahmed, T., Shahid, M., Noman, M., Hussain, S., Khan, M. A., Zubair, M., et al. (2019). "Plant growth-promoting rhizobacteria as biological tools for nutrient management and soil sustainability," in *Plant Growth Promoting Rhizobacteria for Agricultural Sustainability : From Theory to Practices*. Eds. A. Kumar and V. S. Meena (Springer, Singapore), 95–110. doi: 10.1007/978-981-13-7553-8_5

Alavaisha, E., Manzoni, S., and Lindborg, R. (2019). Different agricultural practices affect soil carbon, nitrogen and phosphorous in Kilombero -Tanzania. *J. Environ. Manage.* 234, 159–166. doi: 10.1016/j.jenvman.2018.12.039

Allan, E. (2014). Metagenomics: unrestricted access to microbial communities. *Virulence* 5, 397–398. doi: 10.4161/viru.28057

Ambriz, E., Báez-Pérez, A., Sánchez-Yáñez, J. M., Moutoglis, P., and Villegas, J. (2010). Fraxinus–Glomus–Pisolithus symbiosis: Plant growth and soil aggregation effects. *Pedobiologia* 53, 369–373. doi: 10.1016/j.pedobi.2010.07.001

Angon, P. B., Tahjib-Ul-Arif, M., Samin, S. I., Habiba, U., Hossain, M. A., and Brestic, M. (2022). How do plants respond to combined drought and salinity stress?-A systematic review. *Plants Basel Switz.* 11, 2884. doi: 10.3390/plants11212884

Arif, I., Batool, M., and Schenk, P. M. (2020). Plant microbiome engineering: expected benefits for improved crop growth and resilience. *Trends Biotechnol.* 38, 1385–1396. doi: 10.1016/j.tibtech.2020.04.015

Arora, N. K., and Mishra, J. (2016). Prospecting the roles of metabolites and additives in future bioformulations for sustainable agriculture. *Appl. Soil Ecol.* 107, 405–407. doi: 10.1016/j.apsoil.2016.05.020

Ausmees, N., Kobayashi, H., Deakin, W. J., Marie, C., Krishnan, H. B., Broughton, W. J., et al. (2004). Characterization of NopP, a type III secreted effector of rhizobium sp. Strain NGR234. *J. Bacteriol.* 186, 4774–4780. doi: 10.1128/JB.186.14.4774-4780.2004

Bago, B. (2000). Putative sites for nutrient uptake in arbuscular mycorrhizal fungi. *Plant Soil* 226, 263–274. doi: 10.1023/A:1026456818903

Balachandar, D., Sandhiya, G. S., Sugitha, T. C. K., and Kumar, K. (2006). Flavonoids and growth hormones influence endophytic colonization and in planta nitrogen fixation by a Diazotrophic Serratia sp. in rice. *World J. Microbiol. Biotechnol.* 22, 707–712. doi: 10.1007/s11274-005-9094-0

Banerjee, A., and Roychoudhury, A. (2018). "Chapter 19 - small heat shock proteins: structural assembly and functional responses against heat stress in plants," in *Plant Metabolites and Regulation Under Environmental Stress*. Eds. P. Ahmad, M. A. Ahanger, V. P. Singh, D. K. Tripathi, P. Alam and M. N. AlYemeni (Academic Press), 367–376. doi: 10.1016/B978-0-12-812689-9.00019-4

Bao, X., and Roossinck, M. J. (2013). "Chapter two - multiplexed interactions: viruses of endophytic fungi," in *Advances in Virus Research, Mycoviruses*. Ed. S. A. Ghabrial (Academic Press), 37–58. doi: 10.1016/B978-0-12-394315-6.00002-7

Baslam, M., Garmendia, I., and Goicoechea, N. (2011). Arbuscular mycorrhizal fungi (AMF) improved growth and nutritional quality of greenhouse-grown lettuce. *J. Agric. Food Chem.* 59, 5504–5515. doi: 10.1021/jf200501c

Becquer, A., Trap, J., Irshad, U., Ali, M. A., and Claude, P. (2014). From soil to plant, the journey of P through trophic relationships and ectomycorrhizal association. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00548

Benhiba, L., Fouad, M. O., Essahibi, A., Ghoulam, C., and Qaddoury, A. (2015). Arbuscular mycorrhizal symbiosis enhanced growth and antioxidant metabolism in date palm subjected to long-term drought. *Trees* 6, 1725–1733. doi: 10.1007/s00468-015-1253-9

Berg, G., Rybakova, D., Grube, M., and Köberl, M. (2016). The plant microbiome explored: implications for experimental botany. *J. Exp. Bot.* 67, 995–1002. doi: 10.1093/jxb/erv466

Berg, G., and Smalla, K. (2009). Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol. Ecol.* 68, 1–13. doi: 10.1111/fem.2009.68.issue-1

Bernal, P., Llamas, M. A., and Filloux, A. (2018). Type VI secretion systems in plant-associated bacteria. *Environ. Microbiol.* 20, 1–15. doi: 10.1111/1462-2920.13956

Bernard, G. C., Egnin, M., Bonsi, C., Bernard, G. C., Egnin, M., and Bonsi, C. (2017). "The impact of plant-parasitic nematodes on agriculture and methods of control," in *Nematology - Concepts, Diagnosis and Control* (IntechOpen). doi: 10.5772/intechopen.68958

Bever, J. D., Richardson, S. C., Lawrence, B. M., Holmes, J., and Watson, M. (2009). Preferential allocation to beneficial symbiont with spatial structure maintains mycorrhizal mutualism. *Ecol. Lett.* 12, 13–21. doi: 10.1111/j.1461-0248.2008.01254.x

Bhattacharjee, A., Velickovic, D., Wietsma, T. W., Bell, S. L., Jansson, J. K., Hofmockel, K. S., et al. (2020). Visualizing microbial community dynamics via a controllable soil environment. *mSystems* 5, e00645–e00619. doi: 10.1128/mSystems.00645-19

Bilal, S., Shahzad, R., Asaf, S., Imran, M., Al-Harrasi, A., and Lee, I.-J. (2023). Efficacy of endophytic SB10 and glycine betaine duo in alleviating phytotoxic impact of combined heat and salinity in Glycine max L. via regulation of redox homeostasis and physiological and molecular responses. *Environ. pollut.* 316, 120658. doi: 10.1016/j.envpol.2022.120658

Bluett, S. J., Thom, E. R., Clark, D. A., Macdonald, K. A., and Minneé, E. M. K. (2005a). Effects of perennial ryegrass infected with either AR1 or wild endophyte on dairy production in the Waikato. *N. Z. J. Agric. Res.* 48, 197–212. doi: 10.1080/00288233.2005.9513650

Bluett, S. J., Thom, E. R., Clark, D. A., and Waugh, C. D. (2005b). Effects of a novel ryegrass endophyte on pasture production, dairy cow milk production and calf liveweight gain. *Aust. J. Exp. Agric.* 45, 11–19. doi: 10.1071/EA03263

Bona, E., Lingua, G., Manassero, P., Cantamessa, S., Marsano, F., Todeschini, V., et al. (2015). AM fungi and PGP pseudomonads increase flowering, fruit production, and vitamin content in strawberry grown at low nitrogen and phosphorus levels. *Mycorrhiza* 25, 181–193. doi: 10.1007/s00572-014-0599-y

Bonfante, P., and Genre, A. (2008). Plants and arbuscular mycorrhizal fungi: an evolutionary-developmental perspective. *Trends Plant Sci.* 13, 492–498. doi: 10.1016/j.tplants.2008.07.001

Bonfante, P., and Genre, A. (2010). Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nat. Commun.* 1, 48. doi: 10.1038/ncomms1046

Boughton, B. A., and Thinagaran, D. (2018). Mass spectrometry imaging (MSI) for plant metabolomics. *Methods Mol. Biol. Clifton NJ* 1778, 241–252. doi: 10.1007/978-1-4939-7819-9_17

Bozsoki, Z., Gysel, K., Hansen, S. B., Lironi, D., Krönauer, C., Feng, F., et al. (2020). Ligand-recognizing motifs in plant LysM receptors are major determinants of specificity. *Science* 369, 663–670. doi: 10.1126/science.abb3377

Brader, G., Compant, S., Mitter, B., Trognitz, F., and Sessitsch, A. (2014). Metabolic potential of endophytic bacteria. *Curr. Opin. Biotechnol.* 27, 30–37. doi: 10.1016/j.copbio.2013.09.012

Bradshaw, C. J. A., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., et al. (2016). Massive yet grossly underestimated global costs of invasive insects. *Nat. Commun.* 7, 12986. doi: 10.1038/ncomms12986

Brotman, Y., Landau, U., Cuadros-Inostroza, Á., Takayuki, T., Fernie, A. R., Chet, I., et al. (2013). Trichoderma-plant root colonization: escaping early plant defense responses and activation of the antioxidant machinery for saline stress tolerance. *PloS Pathog.* 9, e1003221. doi: 10.1371/annotation/8b818c15-3fe0-4e56-9be2-e44fd1ed3fae

Bulgarelli, D., Garrido-Oter, R., Münch, P. C., Weiman, A., Dröge, J., Pan, Y., et al. (2015). Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microbe* 17, 392–403. doi: 10.1016/j.chom.2015.01.011

Bulgarelli, D., Schlaeppi, K., Spaepen, S., Ver Loren van Themaat, E., and Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annu. Rev. Plant Biol.* 64, 807–838. doi: 10.1146/annurev-arplant-050312-120106

Burleigh, S. H., Cavagnaro, T., and Jakobsen, I. (2002). Functional diversity of arbuscular mycorrhizas extends to the expression of plant genes involved in P nutrition. *J. Exp. Bot.* 53, 1593–1601. doi: 10.1093/jxb/erf013

Buscaill, P., and van der Hoorn, R. A. L. (2021). Defeated by the nines: nine extracellular strategies to avoid microbe-associated molecular patterns recognition in plants. *Plant Cell* 33, 2116–2130. doi: 10.1093/plcell/koab109

Cai, Q., He, B., Weiberg, A., Buck, A. H., and Jin, H. (2019). Small RNAs and extracellular vesicles: New mechanisms of cross-species communication and innovative tools for disease control. *PloS Pathog.* 15, e1008090. doi: 10.1371/journal.ppat.1008090

Case, R. J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W. F., and Kjelleberg, S. (2007). Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl. Environ. Microbiol.* 73, 278–288. doi: 10.1128/AEM.01177-06

Casieri, L., Ait Lahmidi, N., Doidy, J., Veneault-Fourrey, C., Migeon, A., Bonneau, L., et al. (2013). Biotrophic transportome in mutualistic plant-fungal interactions. *Mycorrhiza* 23, 597–625. doi: 10.1007/s00572-013-0496-9

Chagas, M. B., de, O., Prazeres dos Santos, I., Nascimento da Silva, L. C., Correia, M. T. d. S., Magali de Araújo, J., et al. (2017). Antimicrobial activity of cultivable endophytic fungi associated with hancornia speciosa gomes bark. *Open Microbiol. J.* 11, 179–188. doi: 10.2174/1874285801711010179

Chakraborti, S., Bera, K., Sadhukhan, S., and Dutta, P. (2022). Bio-priming of seeds: Plant stress management and its underlying cellular, biochemical and molecular mechanisms. *Plant Stress* 3, 100052. doi: 10.1016/j.stress.2021.100052

Chaparro, J. M., Sheflin, A. M., Manter, D. K., and Vivanco, J. M. (2012). Manipulating the soil microbiome to increase soil health and plant fertility. *Biol. Fertil. Soils* 48, 489–499. doi: 10.1007/s00374-012-0691-4

Chareesri, A., De Deyn, G. B., Sergeeva, L., Polthanee, A., and Kuyper, T. W. (2020). Increased arbuscular mycorrhizal fungal colonization reduces yield loss of rice (Oryza sativa L.) under drought. *Mycorrhiza* 30, 315–328. doi: 10.1007/s00572-020-00953-z

Chaturvedi, A., Cruz Corella, J., Robbins, C., Loha, A., Menin, L., Gasilova, N., et al. (2021). The methylome of the model arbuscular mycorrhizal fungus, Rhizophagus irregularis, shares characteristics with early diverging fungi and Dikarya. *Commun. Biol.* 4, 1–11. doi: 10.1038/s42003-021-02414-5

Chaudhary, P., Agri, U., Chaudhary, A., Kumar, A., and Kumar, G. (2022). Endophytes and their potential in biotic stress management and crop production. *Front. Microbiol.* 13. doi: 10.3389/fmicb.2022.933017

Chen, N., He, R., Chai, Q., Li, C., and Nan, Z. (2016). Transcriptomic analyses giving insights into molecular regulation mechanisms involved in cold tolerance by Epichloë endophyte in seed germination of Achnatherum inebrians. *Plant Growth Regul.* 80, 367–375. doi: 10.1007/s10725-016-0177-8

Chen, X.-L., Sun, M.-C., Chong, S.-L., Si, J.-P., and Wu, L.-S. (2021). Transcriptomic and metabolomic approaches deepen our knowledge of plant-endophyte interactions. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.700200

Chiu, C. H., and Paszkowski, U. (2021). How membrane receptors tread the fine balance between symbiosis and immunity signaling. *Proc. Natl. Acad. Sci.* 118, e2106567118. doi: 10.1073/pnas.2106567118

Choi, J., Roy Choudhury, A., Walitang, D. I., Lee, Y., and Sa, T. (2022). ACC deaminase-producing Brevibacterium linens RS16 enhances heat-stress tolerance of rice (Oryza sativa L.). *Physiol. Plant* 174, e13584. doi: 10.1111/ppl.13584

Classen, A. T., Sundqvist, M. K., Henning, J. A., Newman, G. S., Moore, J. A. M., Cregger, M. A., et al. (2015). Direct and indirect effects of climate change on soil microbial and soil microbial-plant interactions: What lies ahead? *Ecosphere* 6, art130. doi: 10.1890/ES15-00217.1

Colombo, E. M., Kunova, A., Pizzatti, C., Saracchi, M., Cortesi, P., and Pasquali, M. (2019). Selection of an Endophytic Streptomyces sp. Strain DEF09 From Wheat Roots as a Biocontrol Agent Against Fusarium graminearum. *Front. Microbiol.* 10. doi: 10.3389/fmicb.2019.02356

Compant, S., Samad, A., Faist, H., and Sessitsch, A. (2019). A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. *J. Adv. Res. Special Issue Plant Microbiome* 19, 29–37. doi: 10.1016/j.jare.2019.03.004

Compant, S., van der Heijden, M. G. A., and Sessitsch, A. (2010). Climate change effects on beneficial plant–microorganism interactions. *FEMS Microbiol. Ecol.* 73, 197–214. doi: 10.1111/j.1574-6941.2010.00900.x

Coombs, J. T., and Franco, C. M. M. (2003). Isolation and identification of actinobacteria from surface-sterilized wheat roots. *Appl. Environ. Microbiol.* 69, 5603–5608. doi: 10.1128/AEM.69.9.5603-5608.2003

Cord-Landwehr, S., Melcher, R. L. J., Kolkenbrock, S., and Moerschbacher, B. M. (2016). A chitin deacetylase from the endophytic fungus Pestalotiopsis sp. efficiently inactivates the elicitor activity of chitin oligomers in rice cells. *Sci. Rep.* 6, 38018. doi: 10.1038/srep38018

Cramer, G. R., Urano, K., Delrot, S., Pezzotti, M., and Shinozaki, K. (2011). Effects of abiotic stress on plants: a systems biology perspective. *BMC Plant Biol.* 11, 163. doi: 10.1186/1471-2229-11-163

Dallaire, A., Manley, B. F., Wilkens, M., Bista, I., Quan, C., Evangelisti, E., et al. (2021). Transcriptional activity and epigenetic regulation of transposable elements in the symbiotic fungus Rhizophagus irregularis. *Genome Res.* 31, 2290–2302. doi: 10.1101/gr.275752.121

Deng, S., Caddell, D. F., Xu, G., Dahlen, L., Washington, L., Yang, J., et al. (2021). Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *ISME J.* 15, 3181–3194. doi: 10.1038/s41396-021-00993-z

Dong, Y., Iniguez, A. L., and Triplett, E. W. (2003). Quantitative assessments of the host range and strain specificity of endophytic colonization by Klebsiella pneumoniae 342. *Plant Soil* 257, 49–59. doi: 10.1023/A:1026242814060

Driga, A. M., and Drigas, A. S. (2019). Climate change 101: how everyday activities contribute to the ever-growing issue. *Int. J. Recent Contrib. Eng. Sci. IT IJES* 7, 22–31. doi: 10.3991/ijes.v7i1.10031

Duran, P., Sobarzo, J., Jorquera, M., Azcón, R., Borie, F., Cornejo, P., et al. (2013). Enhanced selenium content in wheat grain by co-inoculation of selenobacteria and arbuscular mycorrhizal fungi: A preliminary study as a potential Se biofortification strategy. *J. Cereal Sci.* 57, 275–280. doi: 10.1016/j.jcs.2012.11.012

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., and Mearns, L. O. (2000). Climate extremes: observations, modeling, and impacts. *Science* 289, 2068–2074. doi: 10.1126/science.289.5487.2068

Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N. K., Bhatnagar, S., et al. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc. Natl. Acad. Sci. U.S.A.* 112, E911–E920. doi: 10.1073/pnas.1414592112

Eida, A. A., Alzubaidy, H. S., de Zélicourt, A., Synek, L., Alsharif, W., Lafi, F. F., et al. (2019). Phylogenetically diverse endophytic bacteria from desert plants induce transcriptional changes of tissue-specific ion transporters and salinity stress in Arabidopsis thaliana. *Plant Sci.* 280, 228–240. doi: 10.1016/j.plantsci.2018.12.002

Escudero-Martinez, C., and Bulgarelli, D. (2023). Engineering the crop microbiota through host genetics. *Annu. Rev. Phytopathol.* 61, 257–277. doi: 10.1146/annurev-phyto-021621-121447

Escudero-Martinez, C., Coulter, M., Alegria Terrazas, R., Foito, A., Kapadia, R., Pietrangelo, L., et al. (2022). Identifying plant genes shaping microbiota composition in the barley rhizosphere. *Nat. Commun.* 13, 3443. doi: 10.1038/s41467-022-31022-y

Evelin, H., Devi, T. S., Gupta, S., and Kapoor, R. (2019). Mitigation of salinity stress in plants by Arbuscular mycorrhizal symbiosis: current understanding and new challenges. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00470

Farooq, M., Wahid, A., Kobayashi, N., Fujita, D., and Basra, S. M. A. (2009). Plant drought stress: effects, mechanisms and management. *Agron. Sustain. Dev.* 29, 185–212. doi: 10.1051/agro:2008021

Favre-Godal, Q., Gourguillon, L., Lordel-Madeleine, S., Gindro, K., and Choisy, P. (2020). Orchids and their mycorrhizal fungi: an insufficiently explored relationship. *Mycorrhiza* 30, 5–22. doi: 10.1007/s00572-020-00934-2

Figueiredo, A. F., Boy, J., and Guggenberger, G. (2021). Common mycorrhizae network: A review of the theories and mechanisms behind underground interactions. *Front. Fungal Biol.* 2. doi: 10.3389/ffunb.2021.735299

Formey, D., Sallet, E., Lelandais-Brière, C., Ben, C., Bustos-Sanmamed, P., Niebel, A., et al. (2014). The small RNA diversity from Medicago truncatularoots under biotic interactions evidences the environmental plasticity of the miRNAome. *Genome Biol.* 15, 457. doi: 10.1186/s13059-014-0457-4

Fouad, M. O., Essahibi, A., Benhiba, L., and Qaddoury, A. (2014). Effectiveness of arbuscular mycorrhizal fungi in the protection of olive plants against oxidative stress induced by drought. *Span. J. Agric. Res.* 12, 763–771. doi: 10.5424/sjar/2014123-4815

Galic, M., Bilandzija, D., Percin, A., Sestak, I., Mesic, M., Blazinkov, M., et al. (2019). Effects of agricultural practices on carbon emission and soil health. *J. Sustain. Dev. Energy Water Environ. Syst.* 7, 539–552. doi: 10.13044/j.sdewes.d7.0271

Gamalero, E., and Glick, B. R. (2020). The use of plant growth-promoting bacteria to prevent nematode damage to plants. *Biology* 9, 381. doi: 10.3390/biology9110381

Geier, B., Sogin, E. M., Michellod, D., Janda, M., Kompauer, M., Spengler, B., et al. (2020). Spatial metabolomics of in *situ* host–microbe interactions at the micrometre scale. *Nat. Microbiol.* 5, 498–510. doi: 10.1038/s41564-019-0664-6

Geissler, C., Davidson, A., and Niesenbaum, R. A. (2023). The influence of climate warming on flowering phenology in relation to historical annual and seasonal temperatures and plant functional traits. *PeerJ* 11, e15188. doi: 10.7717/peerj.15188

Genre, A., Lanfranco, L., Perotto, S., and Bonfante, P. (2020). Unique and common traits in mycorrhizal symbioses. *Nat. Rev. Microbiol.* 18, 649–660. doi: 10.1038/s41579-020-0402-3

Ghaffari, M. R., Ghabooli, M., Khatabi, B., Hajirezaei, M. R., Schweizer, P., and Salekdeh, G. H. (2016). Metabolic and transcriptional response of central metabolism affected by root endophytic fungus Piriformospora indica under salinity in barley. *Plant Mol. Biol.* 90, 699–717. doi: 10.1007/s11103-016-0461-z

Ghaffari, M. R., Mirzaei, M., Ghabooli, M., Khatabi, B., Wu, Y., Zabet-Moghaddam, M., et al. (2019). Root endophytic fungus Piriformospora indica improves drought stress adaptation in barley by metabolic and proteomic reprogramming. *Environ. Exp. Bot.* 157, 197–210. doi: 10.1016/j.envexpbot.2018.10.002

Gough, C., and Cullimore, J. (2011). Lipo-chitooligosaccharide signaling in endosymbiotic plant-microbe interactions. *Mol. Plant-Microbe Interact. MPMI* 24, 867–878. doi: 10.1094/MPMI-01-11-0019

Govindarajulu, M., Pfeffer, P. E., Jin, H., Abubaker, J., Douds, D. D., Allen, J. W., et al. (2005). Nitrogen transfer in the arbuscular mycorrhizal symbiosis. *Nature* 435, 819–823. doi: 10.1038/nature03610

Granér, G., Persson, P., Meijer, J., and Alström, S. (2003). A study on microbial diversity in different cultivars of Brassica napus in relation to its wilt pathogen, Verticillium longisporum. *FEMS Microbiol. Lett.* 224, 269–276. doi: 10.1016/S0378-1097(03)00449-X

Green, E. R., and Mecsas, J. (2016). Bacterial secretion systems: an overview. *Microbiol. Spectr.* 4, 10.1128/microbiolspec.vmbf-0012–2015. doi: 10.1128/microbiolspec.VMBF-0012-2015

Großkinsky, D. K., Svensgaard, J., Christensen, S., and Roitsch, T. (2015). Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. *J. Exp. Bot.* 66, 5429–5440. doi: 10.1093/jxb/erv345

Grover, M., Ali, S., Sandhya, V., Rasul, A., and Venkateswarlu, B. (2011). Role of microorganisms in adaptation of agriculture crops to abiotic stresses. *World J. Microbiol. Biotechnol.* 27, 1231–1240. doi: 10.1007/s11274-010-0572-7

Gul Jan, F., Hamayun, M., Hussain, A., Jan, G., Iqbal, A., Khan, A., et al. (2019). An endophytic isolate of the fungus Yarrowia lipolytica produces metabolites that ameliorate the negative impact of salt stress on the physiology of maize. *BMC Microbiol.* 19, 3. doi: 10.1186/s12866-018-1374-6

Hacquard, S., Spaepen, S., Garrido-Oter, R., and Schulze-Lefert, P. (2017). Interplay between innate immunity and the plant microbiota. *Annu. Rev. Phytopathol.* 55, 565–589. doi: 10.1146/annurev-phyto-080516-035623

Haichar, F.el Z., Marol, C., Berge, O., Rangel-Castro, J. I., Prosser, J. I., Balesdent, J., et al. (2008). Plant host habitat and root exudates shape soil bacterial community structure. *ISME J.* 2, 1221–1230. doi: 10.1038/ismej.2008.80

Hajek, A. E., McManus, M. L., and Delalibera, I. (2007). A review of introductions of pathogens and nematodes for classical biological control of insects and mites. *Biol. Control* 41, 1–13. doi: 10.1016/j.biocontrol.2006.11.003

Hallmann, J. (2001). "Plant interactions with endophytic bacteria," in *Biot. Interact. Plant-Pathog. Assoc* (CABI Books), 87–119. doi: 10.1079/9780851995120.0087

Hamayun, M., Hussain, A., Khan, S. A., Kim, H.-Y., Khan, A. L., Waqas, M., et al. (2017). Gibberellins producing endophytic fungus porostereum spadiceum AGH786 rescues growth of salt affected soybean. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.00686

Hammond, F. N., Booth, C. A., Lamond, J. E., and Proverbs, D. G. (2012). "Introductory insights to climate change challenges," in *Solutions to Climate Change Challenges in the Built Environment* (John Wiley & Sons, Ltd), 1–9. doi: 10.1002/9781444354539.ch1

Hardoim, P. R., van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., et al. (2015). The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiol. Mol. Biol. Rev.* 79, 293–320. doi: 10.1128/MMBR.00050-14

Hart, M., Ehret, D. L., Krumbein, A., Leung, C., Murch, S., Turi, C., et al. (2015). Inoculation with arbuscular mycorrhizal fungi improves the nutritional value of tomatoes. *Mycorrhiza* 25, 359–376. doi: 10.1007/s00572-014-0617-0

Hassani, M. A., Durán, P., and Hacquard, S. (2018). Microbial interactions within the plant holobiont. *Microbiome* 6, 58. doi: 10.1186/s40168-018-0445-0

Hatfield, J. L., Antle, J., Garrett, K. A., Izaurralde, R. C., Mader, T., Marshall, E., et al. (2020). Indicators of climate change in agricultural systems. *Clim. Change* 163, 1719–1732. doi: 10.1007/s10584-018-2222-2

Hawkins, N. J., Bass, C., Dixon, A., and Neve, P. (2019). The evolutionary origins of pesticide resistance. *Biol. Rev.* 94, 135–155. doi: 10.1111/brv.12440

Hesse, U., Schöberlein, W., Wittenmayer, L., Förster, K., Warnstorff, K., Diepenbrock, W., et al. (2003). Effects of Neotyphodium endophytes on growth, reproduction and drought-stress tolerance of three Lolium perenne L. *genotypes. Grass Forage Sci.* 58, 407–415. doi: 10.1111/j.1365-2494.2003.00393.x

Hu, Y., Cheng, H., and Tao, S. (2016). The challenges and solutions for cadmium-contaminated rice in china: A critical review. *Environ. Int.* 92–93, 515–532. doi: 10.1016/j.envint.2016.04.042

Huang, C.-Y., Wang, H., Hu, P., Hamby, R., and Jin, H. (2019). Small RNAs - big players in plant-microbe interactions. *Cell Host Microbe* 26, 173–182. doi: 10.1016/j.chom.2019.07.021

Huang, S., Zhu, S., Kumar, P., and MacMicking, J. D. (2021). A phase-separated nuclear GBPL circuit controls immunity in plants. *Nature* 594, 424–429. doi: 10.1038/s41586-021-03572-6

Irshad, K., Shaheed Siddiqui, Z., Chen, J., Rao, Y., Hamna Ansari, H., Wajid, D., et al. (2023). Bio-priming with salt tolerant endophytes improved crop tolerance to salt stress via modulating photosystem II and antioxidant activities in a sub-optimal environment. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1082480

James, E. K., Olivares, F. L., Baldani, J. I., and Döbereiner, J. (1997). Herbaspirillum, an endophytic diazotroph colonizing vascular tissue 3Sorghum bicolor L. *Moench. J. Exp. Bot.* 48, 785–798. doi: 10.1093/jxb/48.3.785

Jan, S., and Parray, J. (2016). "Use of mycorrhiza as metal tolerance strategy in plants," in *Approaches to heavy metal tolerance in plants*. Eds. S. Jan and J. A. Parray (Singapore: Springer), 57–68. doi: 10.1007/978-981-10-1693-6_4

Jargeat, P., Rekangalt, D., Verner, M.-C., Gay, G., Debaud, J.-C., Marmeisse, R., et al. (2003). Characterisation and expression analysis of a nitrate transporter and nitrite reductase genes, two members of a gene cluster for nitrate assimilation from the symbiotic basidiomycete Hebeloma cylindrosporum. *Curr. Genet.* 43, 199–205. doi: 10.1007/s00294-003-0387-2

Jin, H., Pfeffer, P. E., Douds, D. D., Piotrowski, E., Lammers, P. J., and Shachar-Hill, Y. (2005). The uptake, metabolism, transport and transfer of nitrogen in an arbuscular mycorrhizal symbiosis. *New Phytol.* 168, 687–696. doi: 10.1111/j.1469-8137.2005.01536.x

Johnson, L. J., Johnson, R. D., Schardl, C. L., and Panaccione, D. G. (2003). Identification of differentially expressed genes in the mutualistic association of tall fescue with Neotyphodium coenophialum. *Physiol. Mol. Plant Pathol.* 63, 305–317. doi: 10.1016/j.pmpp.2004.04.001

Johnston-Monje, D., and Raizada, M. N. (2011). Conservation and diversity of seed associated endophytes in Zea across boundaries of evolution, ethnography and ecology. *PloS One* 6, e20396. doi: 10.1371/journal.pone.0020396

Jones, J., Gheysen, G., and Fenoll, C. (2011). *Genomics and Molecular Genetics of Plant-Nematode Interactions* (Dordrecht: Springer Netherlands). doi: 10.1007/978-94-007-0434-3

Kamle, M., Borah, R., Bora, H., Jaiswal, A. K., Singh, R. K., and Kumar, P. (2020). "Systemic acquired resistance (SAR) and induced systemic resistance (ISR): role and mechanism of action against phytopathogens," in *Fungal Biotechnology and Bioengineering, Fungal Biology*. Eds. A. E.-L. Hesham, R. S. Upadhyay, G. D. Sharma, C. Manoharachary and V. K. Gupta (Springer International Publishing, Cham), 457–470. doi: 10.1007/978-3-030-41870-0_20

Kamran, M., Imran, Q. M., Ahmed, M. B., Falak, N., Khatoon, A., and Yun, B.-W. (2022). Endophyte-mediated stress tolerance in plants: A sustainable strategy to enhance resilience and assist crop improvement. *Cells* 11, 3292. doi: 10.3390/cells11203292

Kasmir, J., Senthilkumar, S., Britto, S., and Michael, J. (2011). Identification of fungal endophytes from Orchidaceae members based on nrITS (Internal Transcribed Spacer) region. *Int. Res. J. Biotechnol.* 2 (6), 139–144.

Kaul, S., Gupta, S., Ahmed, M., and Dhar, M. K. (2012). Endophytic fungi from medicinal plants: a treasure hunt for bioactive metabolites. *Phytochem. Rev.* 11, 487–505. doi: 10.1007/s11101-012-9260-6

Kawaguchi, M., and Minamisawa, K. (2010). Plant–microbe communications for symbiosis. *Plant Cell Physiol.* 51, 1377–1380. doi: 10.1093/pcp/pcq125

Khalid, M., Hassani, D., Bilal, M., Asad, F., and Huang, D. (2017). Influence of biofertilizer containing beneficial fungi and rhizospheric bacteria on health promoting compounds and antioxidant activity of Spinacia oleracea L. *Bot. Stud.* 58, 35. doi: 10.1186/s40529-017-0189-3

Khare, E., Mishra, J., and Arora, N. K. (2018). Multifaceted interactions between endophytes and plant: developments and prospects. *Front. Microbiol.* 9. doi: 10.3389/fmicb.2018.02732

Khatoon, Z., Huang, S., Rafique, M., Fakhar, A., Kamran, M. A., and Santoyo, G. (2020). Unlocking the potential of plant growth-promoting rhizobacteria on soil health and the sustainability of agricultural systems. *J. Environ. Manage.* 273, 111118. doi: 10.1016/j.jenvman.2020.111118

Kiarie, S., Nyasani, J. O., Gohole, L. S., Maniania, N. K., and Subramanian, S. (2020). Impact of fungal endophyte colonization of maize (Zea mays L.) on induced resistance to thrips- and aphid-transmitted viruses. *Plants* 9, 416. doi: 10.3390/plants9040416

Kiers, E. T., Duhamel, M., Beesetty, Y., Mensah, J. A., Franken, O., Verbruggen, E., et al. (2011). Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* 333, 880–882. doi: 10.1126/science.1208473

Knauth, S., Hurek, T., Brar, D., and Reinhold-Hurek, B. (2005). Influence of different Oryza cultivars on expression of nifH gene pools in roots of rice. *Environ. Microbiol.* 7, 1725–1733. doi: 10.1111/j.1462-2920.2005.00841.x

Kord, H., Fakheri, B., Ghabooli, M., Solouki, M., Emamjomeh, A., Khatabi, B., et al. (2019). Salinity-associated microRNAs and their potential roles in mediating salt tolerance in rice colonized by the endophytic root fungus Piriformospora indica. *Funct. Integr. Genomics* 19, 659–672. doi: 10.1007/s10142-019-00671-6

Kovtunovych, G., Lar, O., Kamalova, S., Kordyum, V., Kleiner, D., and Kozyrovska, N. (1999). Correlation between pectate lyase activity and ability of diazotrophic Klebsiella oxytoca VN 13 to penetrate into plant tissues. *Plant Soil* 215, 1–6. doi: 10.1023/A:1004790122353

Krings, M., Taylor, T. N., Hass, H., Kerp, H., Dotzler, N., and Hermsen, E. J. (2007). Fungal endophytes in a 400-million-yr-old land plant: infection pathways, spatial distribution, and host responses. *New Phytol.* 174, 648–657. doi: 10.1111/j.1469-8137.2007.02008.x

Kumar, K. K., and Dara, S. K. (2021). Fungal and bacterial endophytes as microbial control agents for plant-parasitic nematodes. *Int. J. Environ. Res. Public. Health* 18, 4269. doi: 10.3390/ijerph18084269

Kusari, S., Singh, S., and Jayabaskaran, C. (2014). Biotechnological potential of plant-associated endophytic fungi: hope versus hype. *Trends Biotechnol.* 32, 297–303. doi: 10.1016/j.tibtech.2014.03.009

Lehnert, H., Serfling, A., Friedt, W., and Ordon, F. (2018). Genome-wide association studies reveal genomic regions associated with the response of wheat (Triticum aestivum L.) to mycorrhizae under drought stress conditions. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01728

Lekberg, Y., Hammer, E. C., and Olsson, P. A. (2010). Plants as resource islands and storage units–adopting the mycocentric view of arbuscular mycorrhizal networks. *FEMS Microbiol. Ecol.* 74, 336–345. doi: 10.1111/fem.2010.74.issue-2

Lery, L. M. S., Hemerly, A. S., Nogueira, E. M., von Krüger, W. M. A., and Bisch, P. M. (2011). Quantitative proteomic analysis of the interaction between the endophytic plant-growth-promoting bacterium Gluconacetobacter diazotrophicus and sugarcane. *Mol. Plant-Microbe Interact. MPMI* 24, 562–576. doi: 10.1094/MPMI-08-10-0178

Levy, A., Conway, J. M., Dangl, J. L., and Woyke, T. (2018). Elucidating bacterial gene functions in the plant microbiome. *Cell Host Microbe* 24, 475–485. doi: 10.1016/j.chom.2018.09.005

Li, Y., Dai, C., Hu, C., Liu, Z., and Kang, C. (2017). Global identification of alternative splicing via comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry. *Plant J.* 90, 164–176. doi: 10.1111/tpj.13462

Lin, B. B., Chappell, M. J., Vandermeer, J., Smith, G., Quintero, E., Bezner-Kerr, R., et al. (2011). Effects of industrial agriculture on climate change and the mitigation potential of small-scale agro-ecological farms. *CABI Rev.* 2011, 1–18. doi: 10.1079/PAVSNNR20116020

Lingua, G., Bona, E., Manassero, P., Marsano, F., Todeschini, V., Cantamessa, S., et al. (2013). Arbuscular mycorrhizal fungi and plant growth-promoting pseudomonads increases anthocyanin concentration in strawberry fruits (Fragaria x ananassa var. Selva) in conditions of reduced fertilization. *Int. J. Mol. Sci.* 14, 16207–16225. doi: 10.3390/ijms140816207

Lipiec, J., Doussan, C., Nosalewicz, A., and Kondracka, K. (2013). Effect of drought and heat stresses on plant growth and yield: a review. *Int. Agrophysics*. 27, 463–477. doi: 10.2478/intag-2013-0017

Liu, H., Carvalhais, L. C., Crawford, M., Singh, E., Dennis, P. G., Pieterse, C. M. J., et al. (2017). Inner plant values: diversity, colonization and benefits from endophytic bacteria. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.02552

Lohani, N., Jain, D., Singh, M. B., and Bhalla, P. L. (2020). Engineering multiple abiotic stress tolerance in canola, Brassica napus. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00003

López-Ráez, J. A., Shirasu, K., and Foo, E. (2017). Strigolactones in plant interactions with beneficial and detrimental organisms: the yin and yang. *Trends Plant Sci.* 22, 527–537. doi: 10.1016/j.tplants.2017.03.011

Luna, M. F., Aprea, J., Crespo, J. M., and Boiardi, J. L. (2012). Colonization and yield promotion of tomato by Gluconacetobacter diazotrophicus. *Appl. Soil Ecol. Microorganisms Sustain. Manage. Soil* 61, 225–229. doi: 10.1016/j.apsoil.2011.09.002

Lynch, J., Cain, M., Frame, D., and Pierrehumbert, R. (2021). Agriculture's contribution to climate change and role in mitigation is distinct from predominantly fossil CO2-emitting sectors. *Front. Sustain. Food Syst.* 4. doi: 10.3389/fsufs.2020.518039

Ma, Y., Dias, M. C., and Freitas, H. (2020). Drought and salinity stress responses and microbe-induced tolerance in plants. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.591911

Ma, Y., Rajkumar, M., Zhang, C., and Freitas, H. (2016). Beneficial role of bacterial endophytes in heavy metal phytoremediation. *J. Environ. Manage.* 174, 14–25. doi: 10.1016/j.jenvman.2016.02.047

Macabuhay, A., Arsova, B., Watt, M., Nagel, K. A., Lenz, H., Putz, A., et al. (2022). Plant growth promotion and heat stress amelioration in Arabidopsis inoculated with paraburkholderia phytofirmans PsJN rhizobacteria quantified with the growScreen-agar II phenotyping platform. *Plants* 11, 2927. doi: 10.3390/plants11212927

Mäder, P., Kaiser, F., Adholeya, A., Singh, R., Uppal, H. S., Sharma, A. K., et al. (2011). Inoculation of root microorganisms for sustainable wheat–rice and wheat–black gram rotations in India. *Soil Biol. Biochem.* 43, 609–619. doi: 10.1016/j.soilbio.2010.11.031

Mahajan, S., and Tuteja, N. (2005). Cold, salinity and drought stresses: an overview. *Arch. Biochem. Biophys.* 444, 139–158. doi: 10.1016/j.abb.2005.10.018

Malar C, M., Wang, Y., Stajich, J. E., Kokkoris, V., Villeneuve-Laroche, M., Yildirir, G., et al. (2022). Early branching arbuscular mycorrhizal fungus Paraglomus occultum carries a small and repeat-poor genome compared to relatives in the Glomeromycotina. *Microb. Genomics* 8, 810. doi: 10.1099/mgen.0.000810

Malinowski, D. P., and Belesky, D. P. (2006). Ecological importance of Neotyphodium spp. grass endophytes in agroecosystems. *Grassl. Sci.* 52, 1–14. doi: 10.1111/j.1744-697X.2006.00041.x

Martin, F., and Nehls, U. (2009). Harnessing ectomycorrhizal genomics for ecological insights. *Curr. Opin. Plant Biol.* 12, 508–515. doi: 10.1016/j.pbi.2009.05.007

Martínez-Hidalgo, P., García, J. M., and Pozo, M. J. (2015). Induced systemic resistance against Botrytis cinerea by Micromonospora strains isolated from root nodules. *Front. Microbiol.* 6. doi: 10.3389/fmicb.2015.00922

Martínez-Medina, A., Fernandez, I., Lok, G. B., Pozo, M. J., Pieterse, C. M. J., and Van Wees, S. C. M. (2017). Shifting from priming of salicylic acid- to jasmonic acid-regulated defences by Trichoderma protects tomato against the root knot nematode Meloidogyne incognita. *New Phytol.* 213, 1363–1377. doi: 10.1111/nph.14251

Martín-Rodríguez, J.Á., Ocampo, J. A., Molinero-Rosales, N., Tarkowská, D., Ruíz-Rivero, O., and García-Garrido, J. M. (2015). Role of gibberellins during arbuscular mycorrhizal formation in tomato: new insights revealed by endogenous quantification and genetic analysis of their metabolism in mycorrhizal roots. *Physiol. Plant* 154, 66–81. doi: 10.1111/ppl.12274

Mathur, S., Tomar, R. S., and Jajoo, A. (2019). Arbuscular mycorrhizal fungi (AMF) protects photosynthetic apparatus of wheat under drought stress. *Photosynth. Res.* 139, 227–238. doi: 10.1007/s11120-018-0538-4

Masuda, K., Abouleila, Y., Ali, A., Yanagida, T., and Masujima, T. (2018). Live single-cell mass spectrometry (LSC-MS) for plant metabolomics. *Methods Mol. Biol. Clifton NJ* 1778, 269–282. doi: 10.1007/978-1-4939-7819-9_19

McLaughlin, A., and Mineau, P. (1995). The impact of agricultural practices on biodiversity. *Agric. Ecosyst. Environ.* 55, 201–212. doi: 10.1016/0167-8809(95)00609-V

Mendes, R., and Raaijmakers, J. M. (2015). Cross-kingdom similarities in microbiome functions. *ISME J.* 9, 1905–1907. doi: 10.1038/ismej.2015.7

Miliute, I., Buzaite, O., Baniulis, D., and Stanys, V. (2015). Bacterial endophytes in agricultural crops and their role in stress tolerance: A review. *Zemdirb.-Agric* 102, 465–478. doi: 10.13080/z-a.2015.102.060

Mishra, B. K., Meena, K. K., Dubey, P. N., Aishwath, O. P., Kant, K., Sorty, A. M., et al. (2016). Influence on yield and quality of fennel (Foeniculum vulgare mill.) grown under semi-arid saline soil, due to application of native phosphate solubilizing rhizobacterial isolates. *Ecol. Eng.* 97, 327–333. doi: 10.1016/j.ecoleng.2016.10.034

Mohsenifard, E., Ghabooli, M., Mehri, N., and Bakhshi, B. (2017). Regulation of miR159 and miR396 mediated by Piriformospora indica confer drought tolerance in rice. *J. Plant Mol. Breed.* 5, 10–18. doi: 10.22058/jpmb.2017.60864.1129

Montero, H., Lee, T., Pucker, B., Ferreras-Garrucho, G., Oldroyd, G., Brockington, S. F., et al. (2021). A mycorrhiza-associated receptor-like kinase with an ancient origin in the green lineage. *Proc. Natl. Acad. Sci.* 118, e2105281118. doi: 10.1073/pnas.2105281118

Moore, W. M., Chan, C., Ishikawa, T., Rennie, E. A., Wipf, H. M.-L., Benites, V., et al. (2021). Reprogramming sphingolipid glycosylation is required for endosymbiont persistence in Medicago truncatula. *Curr. Biol.* 31, 2374–2385.e4. doi: 10.1016/j.cub.2021.03.067

Muthu Narayanan, M., Ahmad, N., Shivanand, P., and Metali, F. (2022). The role of endophytes in combating fungal- and bacterial-induced stress in plants. *Mol. Basel Switz.* 27, 6549. doi: 10.3390/molecules27196549

Nazarov, P. A., Baleev, D. N., Ivanova, M. I., Sokolova, L. M., and Karakozova, M. V. (2020). Infectious plant diseases: etiology, current status, problems and prospects in plant protection. *Acta Naturae* 12, 46–59. doi: 10.32607/actanaturae.11026

Nehls, U., and Plassard, C. (2018). Nitrogen and phosphate metabolism in ectomycorrhizas. *New Phytol.* 220, 1047–1058. doi: 10.1111/nph.15257

Nerva, L., Sandrini, M., Moffa, L., Velasco, R., Balestrini, R., and Chitarra, W. (2022). Breeding toward improved ecological plant-microbiome interactions. *Trends Plant Sci.* 27, 1134–1143. doi: 10.1016/j.tplants.2022.06.004

Newman, M.-A., Sundelin, T., Nielsen, J. T., and Erbs, G. (2013). MAMP (microbe-associated molecular pattern) triggered immunity in plants. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00139

Nguema-Ona, E., Vicré-Gibouin, M., Cannesan, M.-A., and Driouich, A. (2013). Arabinogalactan proteins in root-microbe interactions. *Trends Plant Sci.* 18, 440–449. doi: 10.1016/j.tplants.2013.03.006

Nostadt, R., Hilbert, M., Nizam, S., Rovenich, H., Wawra, S., Martin, J., et al. (2020). A secreted fungal histidine- and alanine-rich protein regulates metal ion homeostasis and oxidative stress. *New Phytol.* 227, 1174–1188. doi: 10.1111/nph.16606

Okazaki, S., Kaneko, T., Sato, S., and Saeki, K. (2013). Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. *Proc. Natl. Acad. Sci.* 110, 17131–17136. doi: 10.1073/pnas.1302360110

Okazaki, S., Tittabutr, P., Teulet, A., Thouin, J., Fardoux, J., Chaintreuil, C., et al. (2016). Rhizobium–legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *ISME J.* 10, 64–74. doi: 10.1038/ismej.2015.103

Oyserman, B. O., Flores, S. S., Griffioen, T., Pan, X., van der Wijk, E., Pronk, L., et al. (2022). Disentangling the genetic basis of rhizosphere microbiome assembly in tomato. *Nat. Commun.* 13, 3228. doi: 10.1038/s41467-022-30849-9

Padilla-Gálvez, N., Luengo-Uribe, P., Mancilla, S., Maurin, A., Torres, C., Ruiz, P., et al. (2021). Antagonistic activity of endophytic actinobacteria from native potatoes (Solanum tuberosum subsp. tuberosum L.) against Pectobacterium carotovorum subsp. carotovorum and Pectobacterium atrosepticum. *BMC Microbiol.* 21, 335. doi: 10.1186/s12866-021-02393-x

Palaniyandi, S. A., Damodharan, K., Yang, S. H., and Suh, J. W. (2014). Streptomyces sp. strain PGPA39 alleviates salt stress and promotes growth of 'Micro tom' tomato plants. *J. Appl. Microbiol.* 117, 766–773. doi: 10.1111/jam.12563

Palmieri, D., Vitale, S., Lima, G., Di Pietro, A., and Turrà, D. (2020). A bacterial endophyte exploits chemotropism of a fungal pathogen for plant colonization. *Nat. Commun.* 11, 5264. doi: 10.1038/s41467-020-18994-5

Panaccione, D. G., Beaulieu, W. T., and Cook, D. (2014). Bioactive alkaloids in vertically transmitted fungal endophytes. *Funct. Ecol.* 28, 299–314. doi: 10.1111/1365-2435.12076

Pandey, V., Ansari, M. W., Tula, S., Yadav, S., Sahoo, R. K., Shukla, N., et al. (2016). Dose-dependent response of Trichoderma harzianum in improving drought tolerance in rice genotypes. *Planta* 243, 1251–1264. doi: 10.1007/s00425-016-2482-x

Pandey, P. K., Samanta, R., and Yadav, R. N. S. (2019). Inside the plant: addressing bacterial endophytes in biotic stress alleviation. *Arch. Microbiol.* 201, 415–429. doi: 10.1007/s00203-019-01642-y

Park, Y.-G., Mun, B.-G., Kang, S.-M., Hussain, A., Shahzad, R., Seo, C.-W., et al. (2017). Bacillus aryabhattai SRB02 tolerates oxidative and nitrosative stress and promotes the growth of soybean by modulating the production of phytohormones. *PloS One* 12, e0173203. doi: 10.1371/journal.pone.0173203

Parniske, M. (2008). Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat. Rev. Microbiol.* 6, 763–775. doi: 10.1038/nrmicro1987

Partida-Martínez, L. P., and Heil, M. (2011). The microbe-free plant: fact or artifact? *Front. Plant Sci.* 2. doi: 10.3389/fpls.2011.00100

Pellegrino, E., and Bedini, S. (2014). Enhancing ecosystem services in sustainable agriculture: Biofertilization and biofortification of chickpea (Cicer arietinum L.) by arbuscular mycorrhizal fungi. *Soil Biol. Biochem.* 68, 429–439. doi: 10.1016/j.soilbio.2013.09.030

Pendergrass, A. G., Meehl, G. A., Pulwarty, R., Hobbins, M., Hoell, A., AghaKouchak, A., et al. (2020). Flash droughts present a new challenge for subseasonal-to-seasonal prediction. *Nat. Clim. Change* 10, 191–199. doi: 10.1038/s41558-020-0709-0

Pentimone, I., Lebrón, R., Hackenberg, M., Rosso, L. C., Colagiero, M., Nigro, F., et al. (2018). Identification of tomato miRNAs responsive to root colonization by endophytic Pochonia chlamydosporia. *Appl. Microbiol. Biotechnol.* 102, 907–919. doi: 10.1007/s00253-017-8608-7

Pérez-Jaramillo, J. E., Carrión, V. J., de Hollander, M., and Raaijmakers, J. M. (2018). The wild side of plant microbiomes. *Microbiome* 6, 143. doi: 10.1186/s40168-018-0519-z

Piromyou, P., Songwattana, P., Greetatorn, T., Okubo, T., Kakizaki, K. C., Prakamhang, J., et al. (2015). The type III secretion system (T3SS) is a determinant for rice-endophyte colonization by non-photosynthetic Bradyrhizobium. *Microbes Environ.* 30, 291–300. doi: 10.1264/jsme2.ME15080

Plett, J. M., and Martin, F. M. (2018). Know your enemy, embrace your friend: using omics to understand how plants respond differently to pathogenic and mutualistic microorganisms. *Plant J.* 93, 729–746. doi: 10.1111/tpj.13802

Porter, S. S., and Sachs, J. L. (2020). Agriculture and the disruption of plant–microbial symbiosis. *Trends Ecol. Evol.* 35, 426–439. doi: 10.1016/j.tree.2020.01.006

Poveda, J., Abril-Urias, P., and Escobar, C. (2020). Biological control of plant-parasitic nematodes by filamentous fungi inducers of resistance: Trichoderma, Mycorrhizal and Endophytic fungi. *Front. Microbiol.* 11. doi: 10.3389/fmicb.2020.00992

Pulavarty, A., Egan, A., Karpinska, A., Horgan, K., and Kakouli-Duarte, T. (2021). Plant parasitic nematodes: A review on their behaviour, host interaction, management approaches and their occurrence in two sites in the republic of Ireland. *Plants* 10, 2352. doi: 10.3390/plants10112352

Qawasmeh, A., Obied, H. K., Raman, A., and Wheatley, W. (2012). Influence of fungal endophyte infection on phenolic content and antioxidant activity in grasses: interaction between Lolium perenne and different strains of Neotyphodium lolii. *J. Agric. Food Chem.* 60, 3381–3388. doi: 10.1021/jf204105k

Qin, S., Feng, W.-W., Zhang, Y.-J., Wang, T.-T., Xiong, Y.-W., and Xing, K. (2018). Diversity of Bacterial Microbiota of Coastal Halophyte Limonium sinense and Amelioration of Salinity Stress Damage by Symbiotic Plant Growth-Promoting Actinobacterium Glutamicibacter halophytocola KLBMP 5180. *Appl. Environ. Microbiol.* 84, e01533–e01518. doi: 10.1128/AEM.01533-18

Raaijmakers, J. M., Vlami, M., and de Souza, J. T. (2002). Antibiotic production by bacterial biocontrol agents. *Antonie Van Leeuwenhoek* 81, 537–547. doi: 10.1023/A:1020501420831

Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Grønlund, M., et al. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* 425, 585–592. doi: 10.1038/nature02039

Rais, A., Jabeen, Z., Shair, F., Hafeez, F. Y., and Hassan, M. N. (2017). Bacillus spp., a bio-control agent enhances the activity of antioxidant defense enzymes in rice against Pyricularia oryzae. *PloS One* 12, e0187412. doi: 10.1371/journal.pone.0187412

Ramesh, A., Sharma, S. K., Sharma, M. P., Yadav, N., and Joshi, O. P. (2014). Inoculation of zinc solubilizing Bacillus aryabhattai strains for improved growth, mobilization and biofortification of zinc in soybean and wheat cultivated in Vertisols of central India. *Appl. Soil Ecol.* 73, 87–96. doi: 10.1016/j.apsoil.2013.08.009

Ramírez-Flores, M. R., Bello-Bello, E., Rellán-Álvarez, R., Sawers, R. J. H., and Olalde-Portugal, V. (2019). Inoculation with the mycorrhizal fungus Rhizophagus irregularis modulates the relationship between root growth and nutrient content in maize (Zea mays ssp. mays L.). *Plant Direct* 3, e00192. doi: 10.1002/pld3.192

Rana, A., Joshi, M., Prasanna, R., Shivay, Y. S., and Nain, L. (2012). Biofortification of wheat through inoculation of plant growth promoting rhizobacteria and cyanobacteria. *Eur. J. Soil Biol.* 50, 118–126. doi: 10.1016/j.ejsobi.2012.01.005

Rashid, S., Charles, T. C., and Glick, B. R. (2012). Isolation and characterization of new plant growth-promoting bacterial endophytes. *Appl. Soil Ecol. Microorganisms Sustain. Manage. Soil* 61, 217–224. doi: 10.1016/j.apsoil.2011.09.011

Rathinasabapathi, B. (2000). Metabolic engineering for stress tolerance: installing osmoprotectant synthesis pathways. *Ann. Bot.* 86, 709–716. doi: 10.1006/anbo.2000.1254

Redecker, D., Kodner, R., and Graham, L. E. (2000). Glomalean fungi from the Ordovician. *Science* 289, 1920–1921. doi: 10.1126/science.289.5486.1920

Reinhardt, D., Roux, C., Corradi, N., and Pietro, A. D. (2021). Lineage-specific genes and cryptic sex: parallels and differences between Arbuscular mycorrhizal fungi and fungal pathogens. *Trends Plant Sci.* 26, 111–123. doi: 10.1016/j.tplants.2020.09.006

Reinhold-Hurek, B., and Hurek, T. (1998). Life in grasses: diazotrophic endophytes. *Trends Microbiol.* 6, 139–144. doi: 10.1016/S0966-842X(98)01229-3

Reinhold-Hurek, B., and Hurek, T. (2011). Living inside plants: bacterial endophytes. *Curr. Opin. Plant Biol.* 14, 435–443. doi: 10.1016/j.pbi.2011.04.004

Ren, B., Wang, X., Duan, J., and Ma, J. (2019). Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation. *Science* 365, 919–922. doi: 10.1126/science.aav8907

Rodriguez, R., and Durán, P. (2020). Natural holobiome engineering by using native extreme microbiome to counteract the climate change effects. *Front. Bioeng. Biotechnol.* 8. doi: 10.3389/fbioe.2020.00568

Rodriguez, R. J., Henson, J., Van Volkenburgh, E., Hoy, M., Wright, L., Beckwith, F., et al. (2008). Stress tolerance in plants via habitat-adapted symbiosis. *ISME J.* 2, 404–416. doi: 10.1038/ismej.2007.106

Rolli, E., Marasco, R., Vigani, G., Ettoumi, B., Mapelli, F., Deangelis, M. L., et al. (2015). Improved plant resistance to drought is promoted by the root-associated microbiome as a water stress-dependent trait. *Environ. Microbiol.* 17, 316–331. doi: 10.1111/1462-2920.12439

Rowan, D. D. (1993). Lolitrems, peramine and paxilline: Mycotoxins of the ryegrass/endophyte interaction. *Agric. Ecosyst. Environ. Acremonium/Grass Interact.* 44, 103–122. doi: 10.1016/0167-8809(93)90041-M

Rozpądek, P., Domka, A. M., Nosek, M., Ważny, R., Jędrzejczyk, R. J., Wiciarz, M., et al. (2018). The Role of Strigolactone in the Cross-Talk Between Arabidopsis thaliana and the Endophytic Fungus Mucor sp. *Front. Microbiol.* 9. doi: 10.3389/fmicb.2018.00441

Ruelland, E., and Zachowski, A. (2010). How plants sense temperature. *Environ. Exp. Bot.* 69, 225–232. doi: 10.1016/j.envexpbot.2010.05.011

Sahebani, N., and Hadavi, N. (2008). Biological control of the root-knot nematode Meloidogyne javanica by Trichoderma harzianum. *Soil Biol. Biochem.* 40, 2016–2020. doi: 10.1016/j.soilbio.2008.03.011

Samal, I., Bhoi, T. K., Majhi, P. K., Murmu, S., Pradhan, A. K., Kumar, D., et al. (2023). Combatting insects mediated biotic stress through plant associated endophytic entomopathogenic fungi in horticultural crops. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1098673

Sandrini, M., Nerva, L., Sillo, F., Balestrini, R., Balestrini, R., Chitarra, W., et al. (2022). Abiotic stress and belowground microbiome: the potential of omics approaches. *Int. J. Mol. Sci.* 23, 1091–1091. doi: 10.3390/ijms23031091

Santoyo, G., Hernández-Pacheco, C., Hernández-Salmerón, J., and Hernández-León, R. (2017). The role of abiotic factors modulating the plant-microbe-soil interactions: toward sustainable agriculture. A review. *Span. J. Agric. Res.* 15, e03R01. doi: 10.5424/sjar/2017151-9990

Santoyo, G., Moreno-Hagelsieb, G., del Carmen Orozco-Mosqueda, Ma., and Glick, B. R. (2016). Plant growth-promoting bacterial endophytes. *Microbiol. Res.* 183, 92–99. doi: 10.1016/j.micres.2015.11.008

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., et al. (2020). The global methane budget 2000–2017. *Earth Syst. Sci. Data* 12, 1561–1623. doi: 10.5194/essd-12-1561-2020

Schadt, E. E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. *Hum. Mol. Genet.* 19, R227–R240. doi: 10.1093/hmg/ddq416

Scholthof, K.-B. G., Adkins, S., Czosnek, H., Palukaitis, P., Jacquot, E., Hohn, T., et al. (2011). Top 10 plant viruses in molecular plant pathology. *Mol. Plant Pathol.* 12, 938–954. doi: 10.1111/j.1364-3703.2011.00752.x

Schüβler, A., Schwarzott, D., and Walker, C. (2001). A new fungal phylum, the Glomeromycota: phylogeny and evolution* *Dedicated to Manfred Kluge (Technische Universität Darmstadt) on the occasion of his retirement. *Mycol. Res.* 105, 1413–1421. doi: 10.1017/S0953756201005196

Seleiman, M. F., Al-Suhaibani, N., Ali, N., Akmal, M., Alotaibi, M., Refay, Y., et al. (2021). Drought stress impacts on plants and different approaches to alleviate its adverse effects. *Plants* 10, 259. doi: 10.3390/plants10020259

Shaffique, S., Khan, M. A., Wani, S. H., Pande, A., Imran, M., Kang, S.-M., et al. (2022). A review on the role of endophytes and plant growth promoting rhizobacteria in mitigating heat stress in plants. *Microorganisms* 10, 1286. doi: 10.3390/microorganisms10071286

Shahzad, A., Ullah, S., Dar, A. A., Sardar, M. F., Mehmood, T., Tufail, M. A., et al. (2021). Nexus on climate change: agriculture and possible solution to cope future climate change stresses. *Environ. Sci. pollut. Res. Int.* 28, 14211–14232. doi: 10.1007/s11356-021-12649-8

Sharma, S., Aneja, M. K., Mayer, J., Schloter, M., and Munch, J. C. (2004). RNA fingerprinting of microbial community in the rhizosphere soil of grain legumes. *FEMS Microbiol. Lett.* 240, 181–186. doi: 10.1016/j.femsle.2004.09.026

Sharma, M., Sudheer, S., Usmani, Z., Rani, R., and Gupta, P. (2020). Deciphering the omics of plant-microbe interaction: perspectives and new insights. *Curr. Genomics* 21, 343–362. doi: 10.2174/1389202921999200515140420

Sheibani-Tezerji, R., Rattei, T., Sessitsch, A., Trognitz, F., and Mitter, B. (2015). Transcriptome profiling of the endophyte burkholderia phytofirmans psJN indicates sensing of the plant environment and drought stress. *mBio* 6, e00621–e00615. doi: 10.1128/mBio.00621-15

Shi, T., Ma, J., Wu, X., Ju, T., Lin, X., Zhang, Y., et al. (2018). Inventories of heavy metal inputs and outputs to and from agricultural soils: A review. *Ecotoxicol. Environ. Saf.* 164, 118–124. doi: 10.1016/j.ecoenv.2018.08.016

Shi, J., Wang, X., and Wang, E. (2023). Mycorrhizal symbiosis in plant growth and stress adaptation: from genes to ecosystems. *Annu. Rev. Plant Biol.* 74, 569–607. doi: 10.1146/annurev-arplant-061722-090342

Silvestri, A., Fiorilli, V., Miozzi, L., Accotto, G. P., Turina, M., and Lanfranco, L. (2019). In silico analysis of fungal small RNA accumulation reveals putative plant mRNA targets in the symbiosis between an arbuscular mycorrhizal fungus and its host plant. *BMC Genomics* 20, 169. doi: 10.1186/s12864-019-5561-0

S. N. Singh (Ed.) (2000). *Trace Gas Emissions and Plants* (Dordrecht: Springer Netherlands). doi: 10.1007/978-94-017-3571-1

Singh, D., Geat, N., Rajawat, M. V. S., Prasanna, R., Kar, A., Singh, A. M., et al. (2018). Prospecting endophytes from different Fe or Zn accumulating wheat genotypes for their influence as inoculants on plant growth, yield, and micronutrient content. *Ann. Microbiol.* 68, 815–833. doi: 10.1007/s13213-018-1388-1

Singh, D., and Prasanna, R. (2020). Potential of microbes in the biofortification of Zn and Fe in dietary food grains. *A review. Agron. Sustain. Dev.* 40, 15. doi: 10.1007/s13593-020-00619-2

Singh, D., Rajawat, M. V. S., Kaushik, R., Prasanna, R., and Saxena, A. K. (2017). Beneficial role of endophytes in biofortification of Zn in wheat genotypes varying in nutrient use efficiency grown in soils sufficient and deficient in Zn. *Plant Soil* 416, 107–116. doi: 10.1007/s11104-017-3189-x

Singh, A., and Roychoudhury, A. (2021). Gene regulation at transcriptional and post-transcriptional levels to combat salt stress in plants. *Physiol. Plant* 173, 1556–1572. doi: 10.1111/ppl.13502

Singh, V. K., Shukla, A. K., and Singh, A. K. (2019). "Chapter 7 - Impact of climate change on plant–microbe interactions under agroecosystems," in *Climate Change and Agricultural Ecosystems*. Eds. K. K. Choudhary, A. Kumar and A. K. Singh (Woodhead Publishing), 153–179. doi: 10.1016/B978-0-12-816483-9.00007-4

Slepetiene, A., Volungevicius, J., Jurgutis, L., Liaudanskiene, I., Amaleviciute-Volunge, K., Slepetys, J., et al. (2020). The potential of digestate as a biofertilizer in eroded soils of Lithuania. *Waste Manage.* 102, 441–451. doi: 10.1016/j.wasman.2019.11.008

Soudzilovskaia, N. A., van Bodegom, P. M., Terrer, C., Zelfde, M.van't E., McCallum, I., Luke McCormack, M., et al. (2019). Global mycorrhizal plant distribution linked to terrestrial carbon stocks. *Nat. Commun.* 10, 5077. doi: 10.1038/s41467-019-13019-2

Sperschneider, J., Yildirir, G., Rizzi, Y. S., Malar C, M., Mayrand Nicol, A., Sorwar, E., et al. (2023). Arbuscular mycorrhizal fungi heterokaryons have two nuclear populations with distinct roles in host–plant interactions. *Nat. Microbiol.* 8, 2142–2153. doi: 10.1038/s41564-023-01495-8

Stępniewska, Z., and Kuźniar, A. (2013). Endophytic microorganisms–promising applications in bioremediation of greenhouse gases. *Appl. Microbiol. Biotechnol.* 97, 9589–9596. doi: 10.1007/s00253-013-5235-9

Stranska, M., Dzuman, Z., Prusova, N., Behner, A., Kolouchova, I., Lovecka, P., et al. (2022). Fungal endophytes of vitis vinifera—Plant growth promoters or potentially toxinogenic agents? *Toxins* 14, 66. doi: 10.3390/toxins14020066

Stuart, E. K., and Plett, K. L. (2019). Digging deeper: in search of the mechanisms of carbon and nitrogen exchange in ectomycorrhizal symbioses. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01658

Su, L., Shen, Z., Ruan, Y., Tao, C., Chao, Y., Li, R., et al. (2017). Isolation of Antagonistic Endophytes from Banana Roots against Meloidogyne javanica and Their Effects on Soil Nematode Community. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.02070

Subramanian, K., Balakrishnan, N., and Natesan, S. (2013). Mycorrhizal symbiosis to increase the grain micronutrient content in maize. *Aust. J. Crop Sci.* 7, 900-910.

Surjan, A., Kudo, S., and Uitto, J. I. (2016). "Risk and vulnerability," in *Sustainable Development and Disaster Risk Reduction, Disaster Risk Reduction*. Eds. J. I. Uitto and R. Shaw (Springer Japan, Tokyo), 37–55. doi: 10.1007/978-4-431-55078-5_3

Sykes, M. T. (2009). "Climate change impacts: vegetation," in *Encyclopedia of Life Sciences* (ELS). (Chichester: John Wiley & Sons, Ltd). doi: 10.1002/9780470015902.a0021227

Tanaka, Y., and Yano, K. (2005). Nitrogen delivery to maize via mycorrhizal hyphae depends on the form of N supplied. *Plant Cell Environ.* 28, 1247–1254. doi: 10.1111/j.1365-3040.2005.01360.x

Tang, M.-J., Lu, F., Yang, Y., Sun, K., Zhu, Q., Xu, F.-J., et al. (2022). Benefits of endophytic fungus phomopsis liquidambaris inoculation for improving mineral nutrition, quality, and yield of rice grains under low nitrogen and phosphorus condition. *J. Plant Growth Regul.* 41, 2499–2513. doi: 10.1007/s00344-021-10462-8

Tang, D., Wang, G., and Zhou, J.-M. (2017). Receptor kinases in plant-pathogen interactions: more than pattern recognition. *Plant Cell* 29, 618–637. doi: 10.1105/tpc.16.00891

Tao, Y., Li, C., Liu, Y., Xu, C., Okabe, S., Matsushita, N., et al. (2023). Identification of microRNAs involved in ectomycorrhizal formation in Populus tomentosa. *Tree Physiol.* 43, 2012–2030. doi: 10.1093/treephys/tpad102

Teixeira, P. J. P. L., Colaianni, N. R., Law, T. F., Conway, J. M., Gilbert, S., Li, H., et al. (2021). Specific modulation of the root immune system by a community of commensal bacteria. *Proc. Natl. Acad. Sci.* 118, e2100678118. doi: 10.1073/pnas.2100678118

Teste, F. P., Jones, M. D., and Dickie, I. A. (2020). Dual-mycorrhizal plants: their ecology and relevance. *New Phytol* 225, 1835–1851. doi: 10.1111/nph.16190

Theocharis, A., Bordiec, S., Fernandez, O., Paquis, S., Dhondt-Cordelier, S., Baillieul, F., et al. (2012). Burkholderia phytofirmans PsJN Primes Vitis vinifera L. and Confers a Better Tolerance to Low Nonfreezing Temperatures. *Mol. Plant-Microbe Interactions®* 25, 241–249. doi: 10.1094/MPMI-05-11-0124

Trdá, L., Boutrot, F., Claverie, J., Brulé, D., Dorey, S., and Poinssot, B. (2015). Perception of pathogenic or beneficial bacteria and their evasion of host immunity: pattern recognition receptors in the frontline. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00219

Trenberth, K. E., Dai, A., van der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., et al. (2014). Global warming and changes in drought. *Nat. Clim. Change* 4, 17–22. doi: 10.1038/nclimate2067

Ullah, A., Mushtaq, H., Ali, H., Munis, M. F. H., Javed, M. T., and Chaudhary, H. J. (2015). Diazotrophs-assisted phytoremediation of heavy metals: a novel approach. *Environ. Sci. pollut. Res. Int.* 22, 2505–2514. doi: 10.1007/s11356-014-3699-5

Vaid, S., Kumar, B., Sharma, A., Shukla, A., and Srivastava, P. (2014). Effect of zinc solubilizing bacteria on growth promotion and zinc nutrition of rice. *J. Soil Sci. Plant Nutr.* 14, 889–910. doi: 10.4067/S0718-95162014005000071

Valente, J., Gerin, F., Le Gouis, J., Moënne-Loccoz, Y., and Prigent–Combaret, C. (2020). Ancient wheat varieties have a higher ability to interact with plant growth-promoting rhizobacteria. *Plant Cell Environ.* 43, 246–260. doi: 10.1111/pce.13652

van den Hoogen, J., Geisen, S., Wall, D. H., Wardle, D. A., Traunspurger, W., de Goede, R. G. M., et al. (2020). A global database of soil nematode abundance and functional group composition. *Sci. Data* 7, 103. doi: 10.1038/s41597-020-0437-3

Vandenkoornhuyse, P., Quaiser, A., Duhamel, M., Le Van, A., and Dufresne, A. (2015). The importance of the microbiome of the plant holobiont. *New Phytol.* 206, 1196–1206. doi: 10.1111/nph.13312

van der Heijden, M. G. A., Klironomos, J. N., Ursic, M., Moutoglis, P., Streitwolf-Engel, R., Boller, T., et al. (1998). Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* 396, 69–72. doi: 10.1038/23932

van der Heijden, M. G. A., Martin, F. M., Selosse, M.-A., and Sanders, I. R. (2015). Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytol.* 205, 1406–1423. doi: 10.1111/nph.13288

Venice, F., Ghignone, S., Salvioli di Fossalunga, A., Amselem, J., Novero, M., Xianan, X., et al. (2020). At the nexus of three kingdoms: the genome of the mycorrhizal fungus

Gigaspora margarita provides insights into plant, endobacterial and fungal interactions. *Environ. Microbiol.* 22, 122–141. doi: 10.1111/1462-2920.14827

Verbruggen, E., van der Heijden, M. G. A., Weedon, J. T., Kowalchuk, G. A., and Röling, W. F. M. (2012). Community assembly, species richness and nestedness of arbuscular mycorrhizal fungi in agricultural soils. *Mol. Ecol.* 21, 2341–2353. doi: 10.1111/j.1365-294X.2012.05534.x

Verma, H., Kumar, D., Kumar, V., Kumari, M., Singh, S. K., Sharma, V. K., et al. (2021). The potential application of endophytes in management of stress from drought and salinity in crop plants. *Microorganisms* 9, 1729. doi: 10.3390/microorganisms9081729

Verma, P. K., Verma, S., and Pandey, N. (2022). Root system architecture in rice: impacts of genes, phytohormones and root microbiota. *3 Biotech.* 12, 239. doi: 10.1007/s13205-022-03299-9

Vyas, P., and Singh, B. (2023). Plant growth-promoting endophytes from Stevia rebaudiana identified to possess bio-control potential against maize sheath blight pathogen Rhizoctonia solani. *Eur. J. Plant Pathol.* doi: 10.1007/s10658-023-02785-x

Walters, W. A., Jin, Z., Youngblut, N., Wallace, J. G., Sutter, J., Zhang, W., et al. (2018). Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proc. Natl. Acad. Sci. U.S.A.* 115, 7368–7373. doi: 10.1073/pnas.1800918115

Wang, M., Eyre, A. W., Thon, M. R., Oh, Y., and Dean, R. A. (2020). Dynamic changes in the microbiome of rice during shoot and root growth derived from seeds. *Front. Microbiol.* 11. doi: 10.3389/fmicb.2020.559728

Wang, Y., Yang, X., Zhang, X., Dong, L., Zhang, J., Wei, Y., et al. (2014). Improved Plant Growth and Zn Accumulation in Grains of Rice (Oryza sativa L.) by Inoculation of Endophytic Microbes Isolated from a Zn Hyperaccumulator, Sedum alfredii H. *J. Agric. Food Chem.* 62, 1783–1791. doi: 10.1021/jf404152u

Wang, X., Zhang, X., Liu, L., Xiang, M., Wang, W., Sun, X., et al. (2015). Genomic and transcriptomic analysis of the endophytic fungus Pestalotiopsis fici reveals its lifestyle and high potential for synthesis of natural products. *BMC Genomics* 16, 28. doi: 10.1186/s12864-014-1190-9

Wani, Z. A., Ashraf, N., Mohiuddin, T., and Riyaz-Ul-Hassan, S. (2015). Plant-endophyte symbiosis, an ecological perspective. *Appl. Microbiol. Biotechnol.* 99, 2955–2965. doi: 10.1007/s00253-015-6487-3

Watts-Williams, S. J., Emmett, B. D., Levesque-Tremblay, V., MacLean, A. M., Sun, X., Satterlee, J. W., et al. (2019). Diverse Sorghum bicolor accessions show marked variation in growth and transcriptional responses to arbuscular mycorrhizal fungi. *Plant Cell Environ.* 42, 1758–1774. doi: 10.1111/pce.13509

Webster, G., Jain, V., Davey, M. R., Gough, C., Vasse, J., Dénarié, J., et al. (1998). The flavonoid naringenin stimulates the intercellular colonization of wheat roots by Azorhizobium caulinodans. *Plant Cell Environ.* 21, 373–383. doi: 10.1046/j.1365-3040.1998.00278.x

Weinert, N., Piceno, Y., Ding, G.-C., Meincke, R., Heuer, H., Berg, G., et al. (2011). PhyloChip hybridization uncovered an enormous bacterial diversity in the rhizosphere of different potato cultivars: many common and few cultivar-dependent taxa. *FEMS Microbiol. Ecol.* 75, 497–506. doi: 10.1111/fem.2011.75.issue-3

Werner, G. D. A., and Kiers, E. T. (2015). Partner selection in the mycorrhizal mutualism. *New Phytol.* 205, 1437–1442. doi: 10.1111/nph.13113

Wilkinson, H. H., Siegel, M. R., Blankenship, J. D., Mallory, A. C., Bush, L. P., and Schardl, C. L. (2000). Contribution of fungal loline alkaloids to protection from aphids in a grass-endophyte mutualism. *Mol. Plant-Microbe Interact. MPMI* 13, 1027–1033. doi: 10.1094/MPMI.2000.13.10.1027

Wu, P., Wu, Y., Liu, C.-C., Liu, L.-W., Ma, F.-F., Wu, X.-Y., et al. (2016). Identification of Arbuscular mycorrhiza (AM)-responsive microRNAs in tomato. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00429

Xiong, C., Zhu, Y.-G., Wang, J.-T., Singh, B., Han, L.-L., Shen, J.-P., et al. (2021). Host selection shapes crop microbiome assembly and network complexity. *New Phytol.* 229, 1091–1104. doi: 10.1111/nph.16890

Yadav, P., Jaiswal, D. K., and Sinha, R. K. (2021). "7 - Climate change: Impact on agricultural production and sustainable mitigation," in *Global Climate Change*. Eds. S. Singh, P. Singh, S. Rangabhashiyam and K. K. Srivastava (Elsevier), 151–174. doi: 10.1016/B978-0-12-822928-6.00010-1

Yan, L., and Khan, R. A. A. (2021). Biological control of bacterial wilt in tomato through the metabolites produced by the biocontrol fungus, Trichoderma harzianum. *Egypt. J. Biol. Pest Control* 31, 5. doi: 10.1186/s41938-020-00351-9

Yan, L., Zhu, J., Zhao, X., Shi, J., Jiang, C., and Shao, D. (2019). Beneficial effects of endophytic fungi colonization on plants. *Appl. Microbiol. Biotechnol.* 103, 3327–3340. doi: 10.1007/s00253-019-09713-2

Yang, J., Kloepper, J. W., and Ryu, C.-M. (2009). Rhizosphere bacteria help plants tolerate abiotic stress. *Trends Plant Sci.* 14, 1–4. doi: 10.1016/j.tplants.2008.10.004

Yildirir, G., Sperschneider, J., Malar C, M., Chen, E. C. H., Iwasaki, W., Cornell, C., et al. (2022). Long reads and Hi-C sequencing illuminate the two-compartment genome of the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. *New Phytol.* 233, 1097–1107. doi: 10.1111/nph.17842

Yin, R., Xia, K., and Xu, X. (2023). Spatial transcriptomics drives a new era in plant research. *Plant J.* 116, 1571–1581. doi: 10.1111/tpj.16437

Zeidler, D., Zähringer, U., Gerber, I., Dubery, I., Hartung, T., Bors, W., et al. (2004). Innate immunity in Arabidopsis thaliana: lipopolysaccharides activate nitric oxide synthase (NOS) and induce defense genes. *Proc. Natl. Acad. Sci. U.S.A.* 101, 15811–15816. doi: 10.1073/pnas.0404536101

Zhang, C., He, J., Dai, H., Wang, G., Zhang, X., Wang, C., et al. (2021). Discriminating symbiosis and immunity signals by receptor competition in rice. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2023738118. doi: 10.1073/pnas.2023738118

Zhang, J., Liu, Y.-X., Zhang, N., Hu, B., Jin, T., Xu, H., et al. (2019). NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. *Nat. Biotechnol.* 37, 676–684. doi: 10.1038/s41587-019-0104-4

Zhang, G., Sun, M., Wang, J., Lei, M., Li, C., Zhao, D., et al. (2019). PacBio full-length cDNA sequencing integrated with RNA-seq reads drastically improves the discovery of splicing transcripts in rice. *Plant J. Cell Mol. Biol.* 97, 296–305. doi: 10.1111/tpj.14120

Zhang, W., Wang, J., Xu, L., Wang, A., Huang, L., Du, H., et al. (2018). Drought stress responses in maize are diminished by Piriformospora indica. *Plant Signal. Behav.* 13, e1414121. doi: 10.1080/15592324.2017.1414121

Zhang, H., Zhu, J., Gong, Z., and Zhu, J.-K. (2022). Abiotic stress responses in plants. *Nat. Rev. Genet.* 23, 104–119. doi: 10.1038/s41576-021-00413-0

Zhou, F., Emonet, A., Dénervaud Tendon, V., Marhavy, P., Wu, D., Lahaye, T., et al. (2020). Co-incidence of damage and microbial patterns controls localized immune responses in roots. *Cell* 180, 440–453.e18. doi: 10.1016/j.cell.2020.01.013

Zhou, M., Lou, J., Li, Y.-K., Wang, Y.-D., Zhou, K., Ji, B.-K., et al. (2015). Butyrolactones from the Endophytic Fungus *Aspergillus versicolor* and their Anti-Tobacco Mosaic Virus Activity. *J. Braz. Chem. Soc* 26, 545–549. doi: 10.5935/0103-5053.20150008

Zhou, X., Liu, X., Zhao, J., Guan, F., Yao, D., Wu, N., et al. (2021). The endophytic bacterium bacillus koreensis 181–22 promotes rice growth and alleviates cadmium stress under cadmium exposure. *Appl. Microbiol. Biotechnol.* 105, 8517–8529. doi: 10.1007/s00253-021-11613-3

Zubair, M., Shakir, M., Ali, Q., Rani, N., Fatima, N., Farooq, S., et al. (2016). Rhizobacteria and phytoremediation of heavy metals. *Environ. Technol. Rev.* 5, 112–119. doi: 10.1080/21622515.2016.1259358

agronomy

MDPI

Article

Rice Regeneration in a Genebank: 21 Years of Data

Francesca Sansoni [1,*], Lorenzo Sena [1,2], Virginia Pozzi [1], Marco Canella [1] and Patrizia Vaccino [1,*]

[1] Council for Agricultural Research and Economics, Research Center for Cereal and Industrial Crops (CREA-CI), Strada Statale 11 per Torino km 2,5, 13100 Vercelli, Italy; lorenzo.sena@unimore.it (L.S.); virginia.pozzi@crea.gov.it (V.P.); marco.canella@crea.gov.it (M.C.)
[2] Department of Life Sciences, University of Modena and Reggio Emilia, 42122 Reggio Emilia, Italy
* Correspondence: francesca.sansoni@crea.gov.it (F.S.); patrizia.vaccino@crea.gov.it (P.V.)

Abstract: Genebanks, other than their pivotal role as diversity conservation repositories, regenerate part of their collection every year to maintain their material in optimal conditions. During regeneration cycles, morpho-physiological data are collected, contributing to the creation of large datasets that offer a valuable resource of information. In Italy, rice cultivation has been documented since the second half of the 15th century, and nowadays, Italy contributes more than 50% of the total European rice production. The ex situ collection of rice (mainly *Oryza sativa* L. subgroup *japonica*) held at the Research Center for Cereal and Industrial Crops (CREA-CI) of Vercelli is quite unique in Italy and its establishment dates back to the beginning of the 20th century. The collection is hereby presented through the analysis of 21 years of historic data, from 2001 to 2022, in 17 different locations in Northern Italy, for a total of 6592 entries, 677 genotypes analyzed and 9 phenotypic traits under investigation. An R script has been developed to analyze the dataset. The BLUEs calculation, heritability, PCA and correlation with weather data provided a comprehensive overview of the germplasm stored in the genebank. The great variability and phenotypic diversity were assessed, key aspects from the perspective of breeding programs. This work starts a re-evaluation of historic data, historic cultivars, and represents the first step toward the shift of the genebank to a bio-digital resource center.

Keywords: *Oryza sativa*; rice; genebank; historical data; R language

Citation: Sansoni, F.; Sena, L.; Pozzi, V.; Canella, M.; Vaccino, P. Rice Regeneration in a Genebank: 21 Years of Data. *Agronomy* **2024**, *14*, 1379. https://doi.org/10.3390/agronomy14071379

Academic Editors: Roberto Pilu, Michela Landoni and Jianbo Wang

Received: 30 April 2024
Revised: 18 June 2024
Accepted: 23 June 2024
Published: 26 June 2024

1. Introduction

Rice (*Oryza sativa* L.) is one of the most important staple crops for more than half of the world's population [1]. The details concerning the history of domestication of rice are still debated, but most researchers agree on three independent domestication events that took place about 8000 years ago. In southern China and northern India, the *japonica* and *indica* subspecies were domesticated from *O. rufipogon* and *O. nivara*, with later crossing and introgression of the two subgroups [2–4]. Around 3000 years ago, African farmers domesticated *O. glaberrima* starting from *O. barthii* [5].

O. sativa has been extensively studied because of its importance in agriculture, but also as a model species for cereals, due to its small diploid genome (389 Mb), whose sequence was firstly obtained for the cultivar Nipponbare and is constantly updated [6,7].

Italy is the first producer of rice in Europe [8], with 1,236,960 tons produced on 218,420 hectares in 2022 [9]; in the same year, it was the 14th exporter in the world, with 12,193 tons exported [9]. Rice cultivation has been documented in Italy since the second half of the fifteenth century, particularly in the swampy areas along the Po river [10]. Over time, it became well established in the northern region between Lombardy and Piedmont, where it became fundamental to the local economy and contributed to strongly modifying the landscape, mainly due to the extensive system of canals created to bring water from the nearby Alps, to ensure a steady supply of water for irrigation.

The first information on the germplasm used refers to a variety called Nostrale (literally "local"), probably a mixture of different ecotypes, as the most cultivated in the area in the 19th century. At the beginning of the 20th century, in order to face rice blast (*Pyricularia oryzae*), which caused dramatic epidemics, new varieties were introduced from Asia and subjected to mass selection, as confirmed by studies on the genetic closeness between Italian rice varieties and Northern Chinese cultivars [11]. Then, starting from 1925, crossing and selection drove the creation of new varieties, focusing on shortening the growth cycle and reducing plant height and blast susceptibility [12,13]. Rice breeding has also been influenced by the market, which has changed over time: from mainly the round rice type, breeding moved to "Long A" type with a high amylose content, for "risotto", and in recent years, to "Long A" for parboiling and "Long B" varieties appreciated for Asian cuisine [12,13].

The Research Centre in Vercelli was founded in 1908 as "Stazione sperimentale di risicoltura" ("Rice Research Station") and quickly became a center of excellence for rice research and breeding, being a pioneer in Italy in several fields, e.g., performing the first hybridization between two rice varieties (1925), and the first experiments of mutation through white light (1934) [10,13].

At the beginning of the 20th century, most of the research carried out in the Centre of Vercelli was based on importing rice varieties from abroad, assessing their adaptability to local conditions and eventually using them for breeding purposes. An example is represented by the varieties "Vary Lava A" and "Vary Lava 51A", acquired in 1925 from Madagascar and extensively used for breeding in the following years. From the 1930s onwards, many successful cultivars have been developed in our Centre. Among them, Roma and Vialone Nano, still appreciated for their culinary characteristics.

The rice collection held at the CREA-CI Genebank (ITA383) consists of 701 *O. sativa* genotypes, mainly ssp. *japonica*, both Italian and international, 2 accessions of the ancestor *O. rufipogon* Griff., and 5 accessions of *O. glaberrima* Steud. Most of the accessions of *O. sativa* in the collection are Italian, while the most represented international varieties are American, Spanish, Portuguese, Chinese, and Egyptian. The oldest variety in the collection is Bertone: according to *Giornale di Risicoltura* (a journal dating back to 1909 with all the latest news and discoveries on rice obtained in the Vercelli Research Centre), it first appeared in 1829, and its cultivation was widespread in the first part of the twentieth century. Originario (also known as "Chinese Originario") was imported from China or Japan in the early 1900s and had great success in the Vercelli area: it was the most cultivated variety until the 1960s, as stated in many technical reports, and it has been used in breeding programs for many years. Ostiglia, Lencino, and Ranghino, although less widespread than Bertone, were cultivated in Italy before 1900, representing the oldest cultivars in the collection.

O. rufipogon was proven to carry alternative alleles that modify the kernel shape, plant architecture, and yield [14] making it an interesting starting point for breeding programs. Introgression lines (ILs) of *O. rufipogon* in the genetic background of *O. sativa* cv Vialone Nano have recently been developed in our Centre and could represent valuable elements in the search for potentially useful genes and QTLs.

O. glaberrima is a good source of valuable genes that can be used to improve the resistance of *O. sativa* against pathogens, such as nematodes, insects, and viruses, as well as factors like drought stress [15]. This species also represents an interesting source of new alleles for salt and dehydration tolerance [16], panicle architecture [17,18], and photosynthetic activity under heat stress [19].

The maintenance and characterization of the collection are the main focus of the RGV-FAO Program, funded by the Italian Ministry of Agriculture, Food Sovereignty and Forestry (Masaf) in order to implement the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) [20]. During the regeneration cycles, the materials are characterized at the morpho-physiological level with international descriptors according to the "Standard Evaluation System for Rice" (www.irri.org) and the national guidelines for in situ, on-farm, and ex situ conservation of plant, animal, and microbial biodiversity [21],

allowing the collection of a huge amount of historical data. In addition, many of the accessions have been characterized at the genomic level (genotyped), thus allowing the deep investigation of the loci responsible for important agronomic traits, e.g., resistance to biotic and abiotic stress, and for quality traits [12,22,23].

This work, through an overview of the historical data collected during years of regeneration and experiments, aims to describe the rice collection stored at CREA-CI Vercelli to infer its variability and diversity, fundamental characteristics for breeding programs that aim to overcome the climate change challenges we are doomed to face, a strategy suggested for the first time less than 10 years ago [24].

2. Materials and Methods

2.1. Materials

The collection stores a huge variability in terms of both the origin of the accessions, covering all the continents except Antarctica (Table 1), and the historical period of their release (Table 2).

Table 1. Country of origin of the rice accessions held in the collection.

Country of Origin	Number of Accessions	Percentage
Italy	311	43.99
United States	60	8.49
Spain	53	7.50
Portugal	40	5.66
China	29	4.10
Egypt	24	3.39
Argentina	21	2.97
France	18	2.55
Japan	18	2.55
Philippines	14	1.98
Bulgaria	12	1.70
India	13	1.84
Brazil	10	1.41
Greece	10	1.41
Australia	9	1.27
Colombia	7	0.99
Hungary	6	0.85
South Korea	4	0.57
Other	28	3.96
Unknown	20	2.83

Table 2. Release period of the rice accessions held in the collection.

Release Period	Number of Accessions	Percentage
Before 1900	4	0.56
1900–1927	33	4.66
1928–1950	28	3.96
1951–1970	81	11.45
1971–1990	119	16.83
1991–2005	198	28.00
2005–2022	121	17.11
Unknown	123	17.39

For the present work, the following parameters were analyzed: flowering time, maturity time, days between flowering and maturity (flo-mat), plant height, Thousand Kernel Weight (TKW), yield, and infection severity/susceptibility (in-field conditions) to the pathogens *Pyricularia oryzae* (rice blast), *Magnaporthe grisea* (neck rice blast), and *Bipolaris oryzae* (brown spot).

In total, the curated dataset comprises 6592 entries, 674 genotypes in a time span of 21 years, from 2002 to 2023. The phenotypic traits did not have the same number of data, with the flowering and maturity times being the most measured, having more than 6000 entries, while the TKW had the lowest number of datapoints (just over 2000). Almost half of the collection is represented by Italian varieties, with a good representation of cultivars originating from the USA, Spain, and China. The complete list of the rice cultivars used in this study can be found in the Supplementary Materials, Table S1.

The trial locations are 17, all in the region between Lombardy and Piedmont in Northern Italy, which is the main rice-growing area in the country (Figure 1).

Figure 1. The locations (blue dots) of the different trials considered in the present work. Some of the experiments considered were performed multiple times in the same location. In the upper left section, the same locations (in red) on a smaller scale.

2.2. Agronomic Information and Phenotypic Data Collection

The materials were sown on dry fields at a seedling rate of 450 seeds per square meter, in single plots of eight rows, 20 cm wide and 2.5 m long, with two to four replications, between April and May, and harvested between September and October, depending on the specific environmental conditions of each year. The agronomic practices followed the standard techniques of the area: permanent flooding condition, chemical control of weeds, and chemical control of pests and diseases when necessary. Fertilization was fractionated into three treatments: before sowing (24/0/29 NPK, 260 kg/ha), at tillering (24/0/29 NPK, 260 kg/ha), and at stem elongation (urea, 40 kg/ha).

The phenotypic traits analyzed were measured according to the Plant Trait Ontology [25,26]: total plant height (TO:0000207), excluding the panicle, recorded in the field at complete maturity; flowering time (TO:0002616), defined as the number of days since sowing when the panicles are fully flowering; maturity time (TO:0000933), defined as the number of days since sowing when the panicles are fully mature; *Pyricularia oryzae*, *Bipolaris oryzae*, and *Magnaporthe grisea* infections (all measured as 0 to 9, where 0 is no infection and 9 is severe infection); yield (in tons/hectare), measured on paddy at 14% humidity;

and TKW (TO:0000533), recorded as the average weight of two 100-paddy rice samples, measured in grams and multiplied by 10.

2.3. Statistical Analyses

The analyses were performed using the R programming language, version 4.3.1 [27]. Since the data were collected over the span of 21 years, and not all the accessions had the same number of replications during the experimental trials, and the analysis was carried out under the assumption it was a longitudinal study. As previous works [28,29] suggest, a Linear Mixed Model was used with the package *nlme*, which employs the Laird–Ware form [30–32] of the mixed model shown in Equation (1):

$$Y_{ij} = X\beta_i + Zu_j + e_{ij} \quad (1)$$

Y_{ij} is the observed phenotype at the ith year and at the jth accession; β_i is the vector of the fixed effect coefficient for the year ith; u_j is the effect of the jth random variable (the accession, in this work also called the "genotype" or "cultivar"); e_{ij} is the error for the observation at the ith year and at the jth accession; and X and Z are matrixes for the effects. The equation has been performed individually for each trait under investigation.

Following the code available in Philipp et al. [28], the outliers have been pinpointed and eliminated from the dataset. The code has been slightly changed since, in this work, the R package nlme was used instead of asreml. The workflow was divided in two steps: (1) year effect and year-specific error variance were obtained from Equation (1) and used to calculate the coefficient of variation, and (2) then Equation (1) was adjusted based on the data just obtained, by assuming accession as the fixed effect and year as the random effect. The Bonferroni–Holm test was then applied to scan for outliers. The normality of the data was tested with the package nortest [33] and the Anderson–Darling test for normality. Skewness and kurtosis were calculated with the package psych [34,35].

The Best Linear Unbiased Estimates (BLUEs) of the phenotypic data, after outlier elimination, were calculated exploiting the R package polyqtlR [36].

The R package sommer [37] was used for the calculation of heritability, following Cullis' heritability model shown in Equation (2):

$$H^2_{\text{Cullis}} = 1 - \frac{v_{\Delta}^{\text{BLUP}}}{2{*}\sigma^2_g} \quad (2)$$

where v_{Δ}^{BLUP} is the mean variance of difference between genotypic BLUPs, while σ^2_g is the genetic variance of the accession. Principal Component Analysis (PCA) was performed using the R packages factoextra and FactoMineR [38].

2.4. Acquisition and Preparation of Bioclimatic Data

Bioclimatic data were collected from ClimatologyLab (https://www.climatologylab.org, accessed on 1 February 2024). From the available datasets, TerraClimate was selected [39]. For this study, only three climatic parameters were selected: minimum temperature (Tmin, in °C), maximum temperature (Tmax, in °C), and precipitation (Ppt, in mm). The climatic data were downloaded as raster files in .TIF format.

To produce the digital map and extract the required data, the software QGIS, version 3.36 "Maidenhead" (https://qgis.org, accessed on 1 February 2024), was utilized. For this work, a project was created in which all the rasters for the selected bioclimatic variables were collected and divided into groups based on the type of variable. The ED50/UTM zone 32N (authority ID: EPSG: 23032) Coordinate Reference System (CRS) was selected for this work.

For each bioclimatic variable, one layer for each year spanning from 2002 to 2023 was loaded into the project. A total of 66 raster layers for the bioclimatic data were added to the map. Experimental sites were included in the QGIS project using a dataset in .csv format. For each site, the latitude and longitude coordinates were provided. Due to the availability of the experimental sites, but not the precise location coordinates in degrees,

the latitude and longitude were set at the nearest town or village, which was used as a reference location for the projects considered in this study. After loading the dataset for the locations, a new points vector layer was added to the digital map. These points were used to collect Tmin, Tmax, and Ppt for each considered location and for each year from 2002 to 2023. Before extracting the desired data, all the loaded layers were reprojected to the project CRS, using the "Warp" option (for rasters) and "Reproject layer" (for vectors) in QGIS. The ED50/UTM zone 32 CRS was selected. The Point sampling tool plugin was used to extract the climate data from the obtained digital map. With this plugin, a GeoPackage (.gpkg) file was obtained for each bioclimatic dataset, represented as a new point vector layer on the map—one layer for each Tmax, Tmin and Ppt. These new data points contain both location data and information relative to the given parameter. The final step to obtain datasets about climate data consisted of extracting the attribute table for each Tmax, Tmin and Ppt GeoPackage file in a .csv file. For this purpose, the MMQGIS plugin was used to extract the desired data from the layers. To reduce the amount of raw data, only those relative to the growing season (from April to October) were selected for each year.

3. Results

The complete dataset, which includes more than 11,000 entries, was manually curated to eliminate obvious mistakes and select, out of the 209 different experiments, only those that could be compared to each other and were performed under standard agronomic management conditions, as described in the Materials and Methods. The final curated dataset comprises 6592 entries, with 9 phenotypic traits analyzed over a span of 21 years, from 2002 to 2023, in 17 different locations in Lombardy and Piedmont, both located in Northern Italy.

The analyses showed great variability among the different cultivars and the different traits. As can be seen in Table 3 (columns three and five), all the traits, except for the yield, showed some outliers, ranging from 3 datapoints excluded for flowering-maturity days up to 42 datapoints excluded for *Magnaporthe grisea* infection. For each trait, the heritability and number of experimental years were calculated before and after outlier removal. The Anderson–Darling normality test, as well as the calculation of skewness and kurtosis, was conducted after outlier removal.

Table 3. Descriptive statistics of the original and outlier-corrected historical data. [a] before outlier removal; [b] after outlier removal. *Pyricularia oryzae* is the pathogen responsible for rice blast, *Magnaporthe grisea* for neck rice blast and *Bipolaris oryzae* for brown spot. Norm.test = Anderson–Darling normality test; TKW = Thousand Kernel Weight.

Trait	N. of Years [a,b]	N. Total Entries [a]	Heritability [a]	N. Total Entries [b]	Heritability [b]	Norm.test (*p*-Value) [b]	Skewness [b]	Kurtosis [b]
Flowering time	18	6317	0.75	6293	0.86	2.2×10^{-16}	0.64	0.92
Maturity time	19	6371	0.54	6358	0.71	2.2×10^{-16}	0.09	1.43
Flowering-Maturity days	18	6194	0.46	6191	0.61	1.2×10^{-13}	−0.03	0.57
Plant height	19	5531	0.85	5511	0.94	2.2×10^{-16}	0.65	0.7
Magnaporthe infection	15	3902	0.23	3860	0.39	2.2×10^{-16}	1.59	2.65
Bipolaris infection	15	2982	0.08	2966	0.41	2.2×10^{-16}	0.7	0.43
Pyricularia infection	17	4224	0.23	4202	0.39	2.2×10^{-16}	1.31	2.23
Yield	16	2393	0.77	2393	0.77	2.2×10^{-16}	−0.39	−0.37
TKW	5	2283	0.89	2216	0.96	2.2×10^{-16}	0.41	−0.12

Removing the outliers is a critical step when a huge amount of data collected over a large timeline is analyzed. Heritability was calculated for all the traits, both before and after the removal of the outliers (Table 3). In all cases (except yield), the parameter improved after filtration, especially for brown spot (*B. oryzae*) infection (from 0.08 to 0.41). Among all the traits analyzed, pathogen infections showed the lowest heritability, confirming how

this trait is heavily influenced by both the environment, such as the weather conditions and location (some regions are more heavily affected by the diseases), and the field treatments. This is also confirmed by the results reported in Table S2 (Supplementary Materials): *Bipolaris* infection is shown to be influenced by some of the experimental sites but not by the experimental year.

After elimination of the outliers, the Anderson–Darling normality test carried out on the entire dataset did not suggest the normal distribution of the data (Table 3 and Figure 2, lower graphs), as confirmed by the skewness and kurtosis, with multiple peaks. The flowering and maturity time showed the same tendency to have a lower peak in the middle.

The flowering time, maturity time, and flowering-maturity days (Figure 2a–c, upper graphs) visibly fluctuated in different years, while other parameters such as the plant height (Figure 2d) were more stable during the years, as expected. This last result is confirmed by the data shown in Table S2 (Supplementary Materials) where, as for the previously described *Bipolaris oryzae* infection, this parameter is not affected by the years of experiment as it is by the different environments. The yield (Figure 2e) seemed to be more affected by the location (Figure S1 and Table S2) than by the year of the experiment.

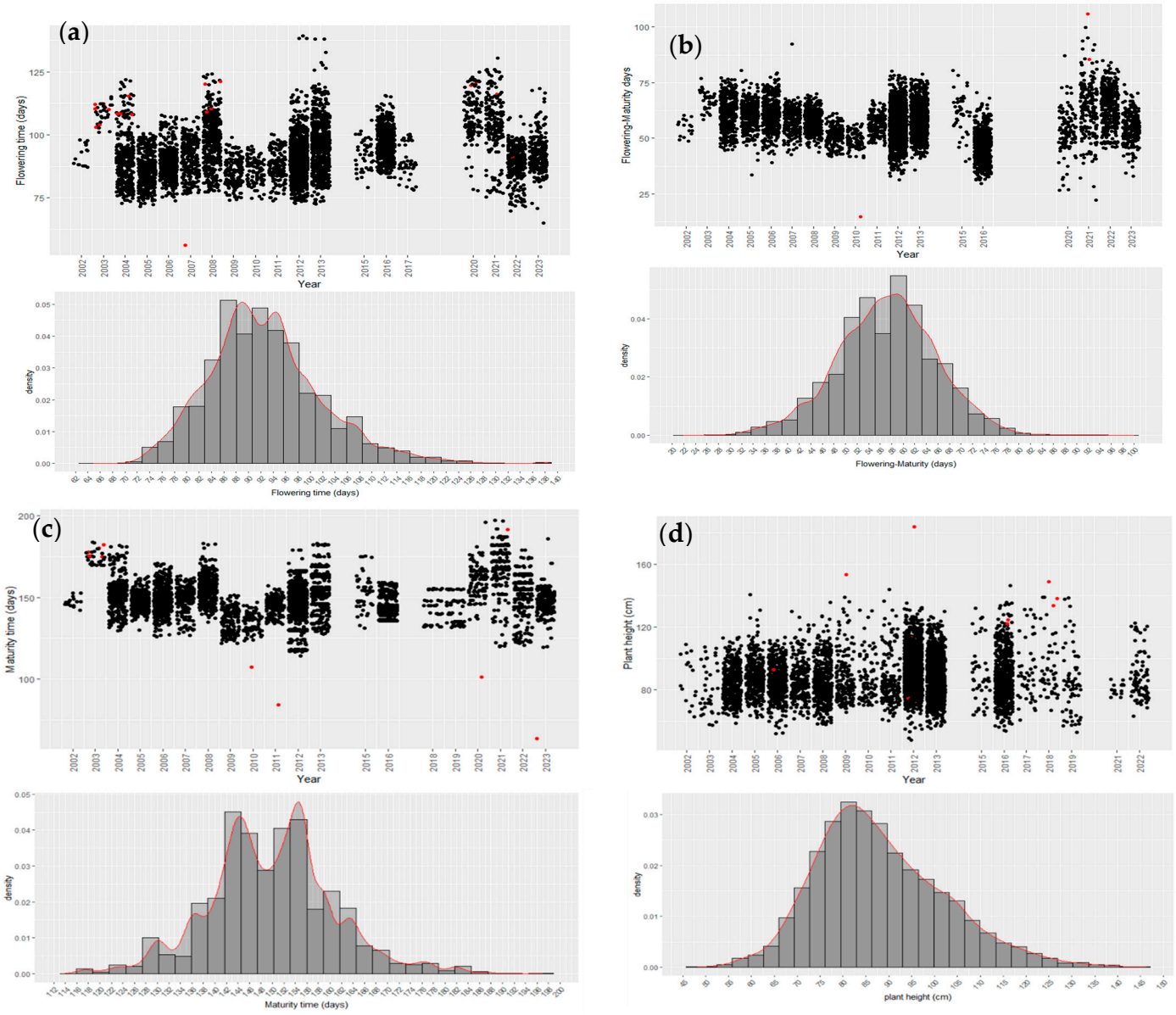


Figure 2. *Cont.*

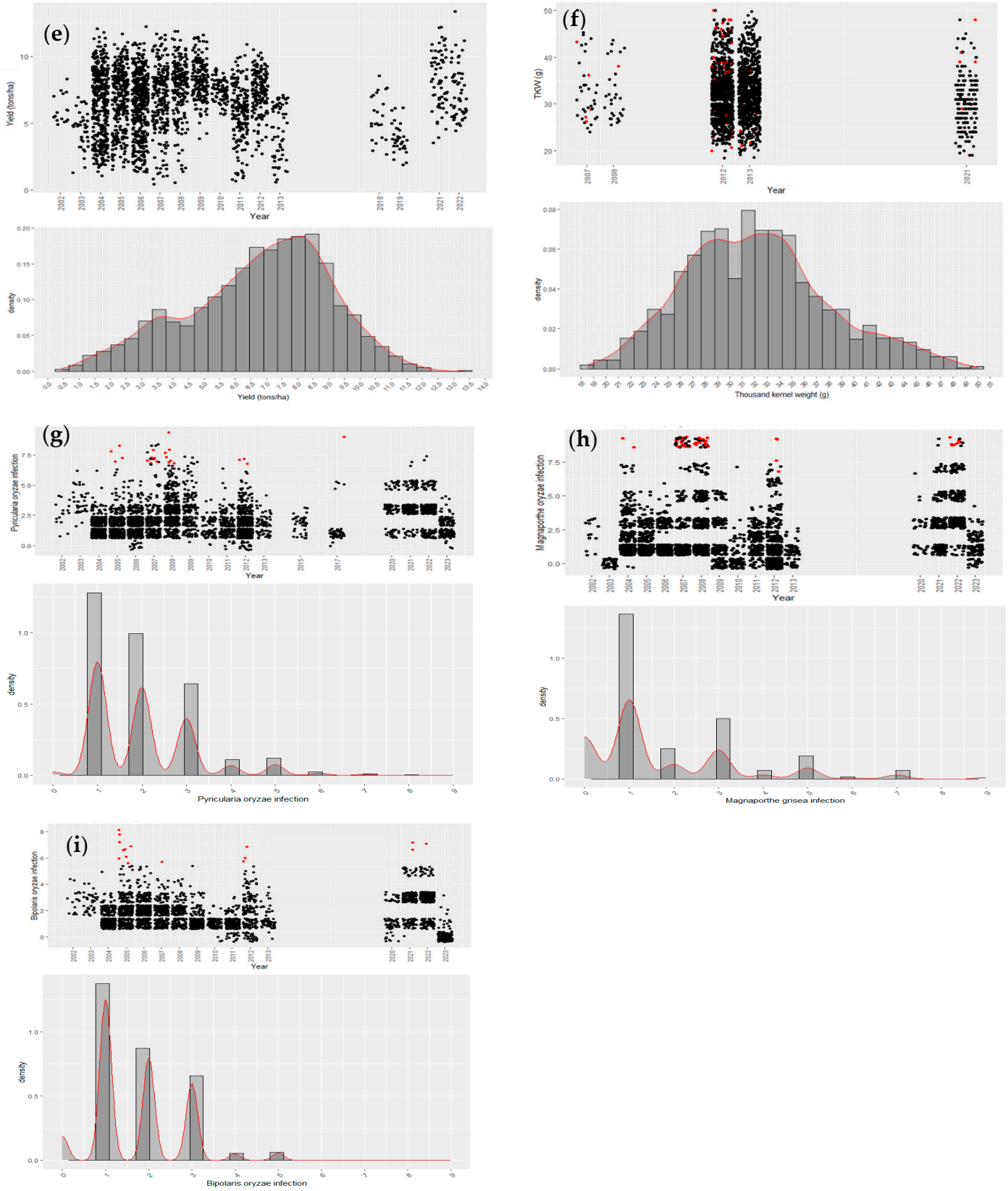


Figure 2. Graphical representation of the values for the traits analyzed in 21 years of experiments. For each trait (**a**–**i**) in the upper graph, the values of the trait for each experimental season are represented by dots (with outliers in red). In the lower graph, the frequency (histogram) and density (red line) distribution after outlier removal are reported. The non-normal distribution of the data is evident. (**a**–**f**) Agronomic and post-harvesting measurements, and (**g**–**i**) fungal infection severity.

Principal Component Analysis (PCA) was used to visualize the relationship among the traits and their contribution to the overall variability (Figure 3). The results were calculated using the complete dataset, after removing outliers. The reproductive stages of the plants (flowering, maturity, and flowering-maturity times) generated most of the variation in the dataset in the first two dimensions. In the first dimension, the diseases under investigation accounted for between 15% and 18% of the total variability. The PCA also includes weather information extracted for each year in each location. Since all the locations are quite close to each other (Figure 1), there are no big differences between the environment's weather conditions, but it is possible to appreciate the great variation in the total rain, minimum and maximum temperatures during the years (Figure S2): from 2002 to 2023, a slight decrease in precipitation associated with an increase in temperature (both minimum and maximum) can be observed. There have been many peaks of high temperature; for example, in 2022 (average maximum temperature in August of 33.08 °C), 2006 (32.41 °C), and 2010 (31.09 °C). The minimums of the precipitation along the rice-growing season (April to October) were recorded in 2003 (315 mm), 2004 (333 mm), 2017 (364 mm), and 2022 (376 mm).

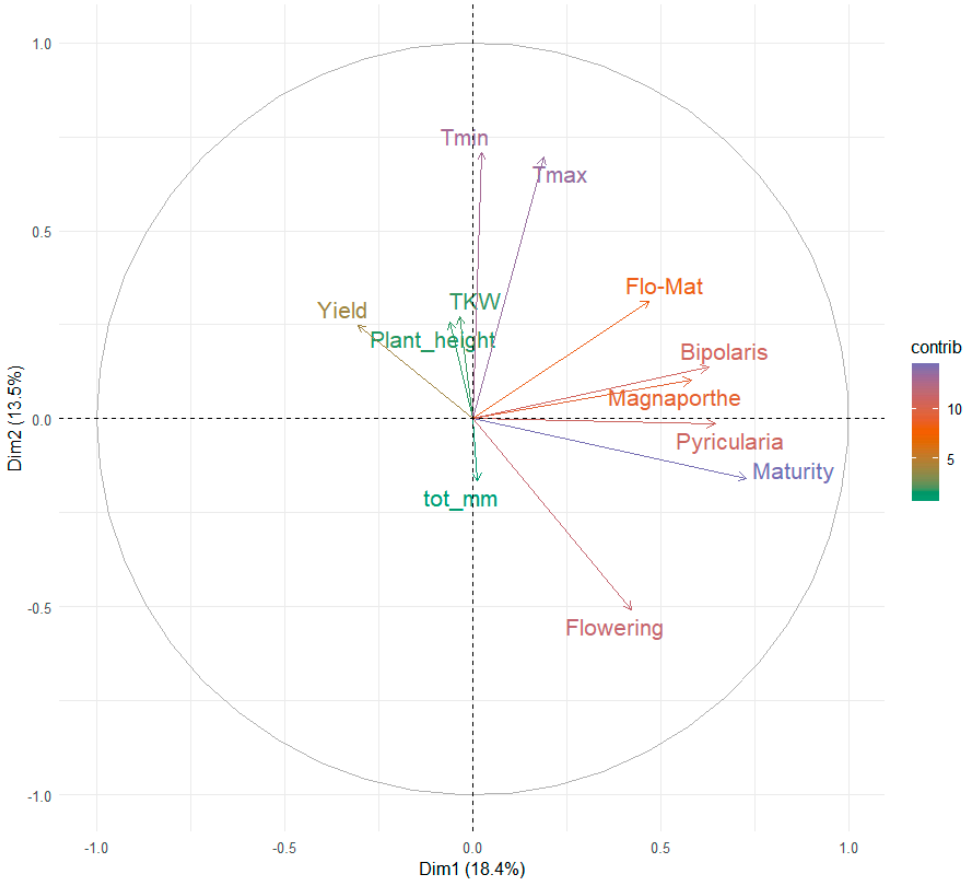


Figure 3. Percentage of variance for each trait and weather variable in the first two dimensions of the Principal Component Analysis.

The correlation matrix was calculated using the weather data and from the complete dataset after removing outliers (Figure 4). The matrix did not show any statistically significant relationships, but some trends can be observed. For instance, there is a negative relationship between maturity and total rainfall of the season, as well as between TKW and plant height. Additionally, the yield and total rainfall are negatively correlated (−0.41).

Figure 4. Pearson correlation matrix among the variables and the weather data of each year and location of the trials.

After the BLUEs calculation, a statistical description of the varieties under investigation was performed (Table 4). The variance is particularly high for the plant height, reproductive stages, yield, and yield-related traits (TKW), while the mode showed results more comparable to modern commercially available cultivars.

Table 4. Statistical description of the rice accessions held in the collection over the 21 years of experiments (BLUEs data). St. Dev. = standard deviation; σ^2 = variance.

Trait	Mean	Median	Min	Max	Mode	St. Dev.	σ^2
Flowering time	95.52	94.57	65.90	139.55	90.53	8.27	68.5
Maturity time	152.54	152.84	126.92	183.89	154.87	7.81	61.1
Flowering-Maturity days	57.68	57.85	16.77	94.77	60.77	7.21	52.1
Plant height	90.73	88.41	58.79	137.56	88.76	14.2	203.65
Magnaporthe infection	1.71	1.56	0.01	6.50	0.50	1.10	1.22
Bipolaris infection	1.91	2.11	0.10	5.00	2.11	0.86	0.74
Pyricularia infection	2.20	2.10	0.10	6.10	2.24	0.74	0.58
Yield	6.51	6.67	1.76	10.74	1.76	1.35	1.84
TKW	31.46	31.46	19.06	47.39	31.06	5.56	30.94

4. Discussion

Historical data are gaining increasing interest in the scientific community for their potential [40–42]. These studies have already proven fundamental for understanding the importance of genebanks [28], their data, and the information extracted from them, especially if associated with genomic data [43].

From this perspective, this work aims to present the collection held in the CREA genebank in Vercelli, exploiting historical data collected during over 20 years of field regeneration. To provide a proper description and inference of the general characteristics

of each cultivar in the collection, outliers were calculated and eliminated. Despite this step, the Anderson–Darling normality test did not suggest the normal distribution of the data. The diseases' distribution and the density of the data are very oddly shaped, but these results are expected since they are non-continuous data. Regarding flowering and maturity days, it is possible to see a double high peak with a lower central peak: indeed, the collection is mainly represented by early or late varieties, with few varieties having a medium growing cycle.

Even though the distributions of the data were not normal, the heritability after outlier filtering was quite high for the flowering time (0.86), plant height (0.94), and TKW (0.96). These results are expected since these traits are under strong genetic control [44]. The flowering time, connected with the circadian clock and photoperiod sensitivity of the plant, is one of the first aspects domestication put pressure on, since it was fundamental for adapting the plant to different environments [3,45,46]. *Oryza sativa* is a short-day plant, but it was induced to mutate into a long-day plant to complete its life cycle in European conditions and be cultivated in the region.

The heritability was medium–high for the yield (0.77). This parameter is ruled by many genes and loci, but it is heavily influenced by environmental conditions, as inferred from similar work [41] and confirmed in this study: there is not a detectable influence of the agronomic year, rather there is a quite remarkable influence by the different environments, where the genotype influence, or rather the differences in performance of the different cultivars are quite strong (Table S2).

Regarding the plant height, great variability was observed among the rice cultivars, with the tendency, driven by the breeding selection, to shorten it over time to avoid lodging. Ostiglia and Lencino (both cultivars developed in the 1800s) are about 110 cm in height; among the oldest varieties, Bombon (reaching more than 130 cm), Gigante Vercelli (literally "the giant of Vercelli", 128 cm), and Fortuna (126 cm), bred between 1900 and 1930, were the tallest. Generally, the shortest varieties have been bred in the last 50 years, even if the decrease in stature is not as dramatic in rice (in the first two dimensions of the PCA plant height count for 1.60% of the total variability) as in other cereals such as wheat, for example [47]. An interesting exception is represented by the cultivar Nano, developed in 1912, which strikingly showed an average height of 55 cm during the years of the experiments.

The pathogens infection had a peak in the early 2000s: the years 2002 and 2007 had the highest amount of rain and lowest mean temperature during the rice-growing season. Rice cultivation in Europe is traditionally performed under constantly flooded conditions, as in these experiments; therefore, rain should not directly affect yield. However, rain events are often connected to higher humidity and lower temperature, as can be inferred from the results presented in Figure 3, which are highly connected with the incidence of pathogen infections that can consequently impact yield. Fungal susceptibility showed the lowest heritability in the dataset, 0.39 for *Magnaporthe grisea* and *Pyricularia oryzae*, and 0.41 for *Bipolaris oryzae*. Many genes are well-known to be involved in building resistance against illnesses [48–50], but much of the plant infection rate seems to be dependent on the agronomic season, the weather, and the field treatments. The results shown in Table S2 make it possible to infer the same conclusion: rice blast and rice neck blast (*P. oryzae* and *M. grisea*, respectively) are influenced by the experimental year. In general, all three infectious diseases addressed in this work show how some locations are more prone to the illnesses: Borgolavezzaro, Ronsecco, Rovasenda, Valle Lomellina and Vercelli. A major part of the data were collected in Vercelli, so this could explain why this location has shown to be correlated with most traits. *Bipolaris* infection is not affected by the year but mostly by the different environments: this illness is more virulent in certain sites, also depending on the field treatment (i.e., pesticides) applied and, in this scenario, is influenced by the genotypes that have different resistance to the pathogen.

Yield and plant height show a pattern similar to *Bipolaris*: there is no influence by the agronomic year, but mainly by the environment. Some locations are more fertile naturally

and in this context it is possible to better appreciate the influence of the different genotypes, as also confirmed by the high heritability scores for these two parameters.

Dividing the cultivars according to their release period (Figure S3, Supplementary Materials), a good overview of the collection and its diversity can be appreciated. Ostiglia (purple dot), selected from Asian material imported in the 19th century, is related to a long flowering time; Ranghino (pink dot), from 1887, is quite susceptible to rice blast, neck blast, and brown spot. The first news regarding rice blast (*P. oryzae*) infections in Italy was reported in the beginning of 1900, and since then, Italian breeders have worked to produce new resistant varieties [48,49,51]. The first variety reported to be resistant to rice blast is Precoce Rossi, developed in 1953. Analyzing the correlation between the continent of origin and the phenotypic variability for each trait, it is possible to see how the European group seems the most affected by the illnesses (Figure S4): this group consists mostly of Italian varieties, many of which are quite old, not yet selected for disease resistance. Furthermore, this pathogen is known for its rapid evolution, which jeopardizes efforts to build plant resistance [52,53]. Many accessions from South America, North Africa (Egypt), and Oceania (Australia) are primarily breeding lines of modern origin and generally show lower infection rates, as illustrated in the PCA in Figure S4 (Supplementary Materials).

The statistical description of the varieties based on the BLUEs values (Table 4) provides a good overview of the collection. The variance is high, while the mode and the average are more in line with the values recorded for currently commercially grown rice cultivars. This confirms the great variability of the collection, mainly due to the performance of the old varieties. Further variability could be ascribed to the influence of environmental conditions (Figure S1 and Table S2 in Supplementary Materials), especially for the reproductive stages, plant height, and yield. Regarding the pathogen infections, the collection showed a medium–low average value (Table 4). A possible explanation could be that almost half of the collection is composed by varieties developed in the last 40 years, when breeding was heavily focused on finding solutions against pathogens.

The PCA highlights how infectious diseases, maturation time, flowering time, and temperature are the inputs that provide most of the variability in our dataset. As expected, yield is negatively correlated with late flowering and maturity and infectious disease severity. Although the TKW does not significantly contribute to the overall variability, it can be related to the yield, a fundamental trait of agronomic interest. Yield is proven to be severely affected by infectious diseases that can harm the plants in the field. It is estimated that up to 30% of the global annual rice production is lost due to plant pathogens [19].

The Pearson correlation matrix does not provide any statistically significant correlation, but some trends are visible: between plant height and TKW, and between maturity time and total rain in the agronomic season. The latter trend could be due to the decrease in temperature caused by heavy rains, which increases the time needed for completing the life cycle.

5. Conclusions

The aim of this work was to shed light on the potential of our genebank, its diversity, and the breeding treasure it can represent, especially when historic and easy-to-access data are curated. The collection was found to exhibit great variability for the traits analyzed, the period of development (ranging from 1829 to 2020), the genetic diversity, and the phenotypic performance in field. Some experimental sites have proven to be more productive and less prone to some illnesses, while the influence of the genetic background was always quite strong: the most resistant varieties are the most recent ones (within the last 40 years). The most productive varieties (more than 8 tons/ha) are Baldo, the best among the Italian varieties, and Handao 297 (a Chinese cultivar), the best performing within all the collection.

This study is a valuable source for future multi-environment analysis and, strengthened by the multitude of traits recorded, for multi-trait and selection models analysis. Furthermore, this approach could form the basis for developing a common protocol for

genebank data analysis, making it easier and faster for other research centers, even with other plants under investigation, to perform similar analyses.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/agronomy14071379/s1. Table S1: List of the 674 rice accessions used in this study, identification code, name, country of origin (coded by the ISO country code), species and release period. Table S2: Results of the Linear Mixed Model used for outlier correction. All the calculations are performed before outlier removal on the entire raw dataset. In the upper part of the table are shown the results of the Linear Mixed Model used in the outlier calculation where the year of experiment is used as the fixed effect and the genotype as the random effect. In the lower part of the table are presented the results of the Linear Mixed Model where the experiment location is used as the fixed effect and the genotype as the random effect. *Bipolaris* infection was not measured in all the environments: missing data are indicated with "No data" label in the table. Thousand Kernel Weight (TKW) was measured just in one location (Vercelli), hence the Linear Mixed Model with sites as the fixed effect was not performed. The random variable is written in italics. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Figure S1. Graphical representation of the row data for each trait collected in each experimental site (upper) and of the regression for the Linear Mixed Model (Equation (1) in the main text) with year as fixed effect (lower). All the calculations were performed before outlier correction. The barplots show, for each trait, the values for each experimental site. For these visual representations, only genotypes sown in at least 3 different environments were considered. The results of the Linear Mixed Model are reported in Table S2. The TKW barplot is not shown since all the data were collected in only one location, Vercelli. Figure S2. The graphs show the total rain in the agronomic season (in mm), maximum temperature and minimum temperature (both in Celsius), average of all the locations considered, for the 21 years of experiments. Figure S3. Percentage of variance for each trait in the first two dimensions of the PCA, where the cultivars have been grouped by the period of development. Figure S4. Percentage of variance for each trait in the first two dimensions of the PCA, where the cultivars have been grouped by the continent of origin.

Author Contributions: Conceptualization, methodology, formal analysis and data curation, F.S.; writing—original draft preparation, F.S. and L.S.; writing—review and editing, F.S., L.S., M.C., V.P. and P.V.; visualization, F.S. and L.S.; supervision, P.V.; funding acquisition, P.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Agriculture, Food Sovereignty and Forestry (MASAF) within the RGV FAO program (grant D.M._n._50045/2023).

Data Availability Statement: The data, the R code developed for the analyses and the code for wheatear data downloading and management are available upon request.

Acknowledgments: We would like to add a special thanks to Aurora Cattaneo (CREA-DC, Vercelli) for her help and support in finding information about the cultivars in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gutaker, R.M.; Groen, S.C.; Bellis, E.S.; Choi, J.Y.; Pires, I.S.; Bocinsky, R.K.; Slayton, E.R.; Wilkins, O.; Castillo, C.C.; Negrão, S.; et al. Genomic History and Ecology of the Geographic Spread of Rice. *Nat. Plants* **2020**, *6*, 492–502. [CrossRef]
2. Wang, W.; Mauleon, R.; Hu, Z.; Chebotarov, D.; Tai, S.; Wu, Z.; Li, M.; Zheng, T.; Fuentes, R.R.; Zhang, F.; et al. Genomic Variation in 3010 Diverse Accessions of Asian Cultivated Rice. *Nature* **2018**, *557*, 43–49. [CrossRef]
3. Jing, C.Y.; Zhang, F.M.; Wang, X.H.; Wang, M.X.; Zhou, L.; Cai, Z.; Han, J.D.; Geng, M.F.; Yu, W.H.; Jiao, Z.H.; et al. Multiple Domestications of Asian Rice. *Nat. Plants* **2023**, *9*, 1221–1235. [CrossRef]
4. Gross, B.L.; Zhao, Z. Archaeological and Genetic Insights into the Origins of Domesticated Rice. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 6190–6197. [CrossRef]
5. Fornasiero, A.; Wing, R.A.; Ronald, P. Rice Domestication. *Curr. Biol.* **2022**, *32*, R20–R24. [CrossRef] [PubMed]
6. Matsumoto, T.; Wu, J.; Kanamori, H.; Katayose, Y.; Fujisawa, M.; Namiki, N.; Mizuno, H.; Yamamoto, K.; Antonio, B.A.; Baba, T.; et al. The Map-Based Sequence of the Rice Genome. *Nature* **2005**, *436*, 793–800. [CrossRef]
7. Wang, C.; Han, B. Twenty Years of Rice Genomics Research: From Sequencing and Functional Genomics to Quantitative Genomics. *Mol. Plant* **2022**, *15*, 593–619. [CrossRef] [PubMed]
8. Muthayya, S.; Sugimoto, J.D.; Montgomery, S.; Maberly, G.F. An Overview of Global Rice Production, Supply, Trade, and Consumption. *Ann. N. Y. Acad. Sci.* **2014**, *1324*, 7–14. [CrossRef]

9. FAOSTAT. Available online: www.fao.org (accessed on 27 March 2024).
10. Tamborini, L.; Cavigiolo, S.; Greppi, D.; Lupotto, E. *Le Varietà di Riso Coltivate in Italia Nella Prima Metà Del XX Secolo*; Dow AgroSciences: Midland, TX, USA, 2008; pp. 1–135.
11. Cai, X.; Fan, J.; Jiang, Z.; Basso, B.; Sala, F.; Spada, A.; Grassi, F.; Lu, B.R. The Puzzle of Italian Rice Origin and Evolution: Determining Genetic Divergence and Affinity of Rice Germplasm from Italy and Asia. *PLoS ONE* **2013**, *8*, e80351. [CrossRef]
12. Spada, A.; Mantegazza, R.; Biloni, M.; Caporali, E.; Sala, F. Italian Rice Varieties: Historical Data, Molecular Markers and Pedigrees to Reveal Their Genetic Relationships. *Plant Breed.* **2004**, *123*, 105–111. [CrossRef]
13. Mongiano, G.; Titone, P.; Tamborini, L.; Pilu, R.; Bregaglio, S. Evolutionary Trends and Phylogenetic Association of Key Morphological Traits in the Italian Rice Varietal Landscape. *Sci. Rep.* **2018**, *8*, 13612. [CrossRef]
14. Wang, Y.; Liang, X.; Gong, G.; Zhang, G.; Zhao, H.; Zheng, Z.; Wang, C.; Zhu, H.; Huang, J.; Li, Z.; et al. qGLF5 from Oryza rufipogon Griff. improves kernel shape, plant architecture, and yield in rice. *Theor. Appl. Genet.* **2023**, *136*, 225. [CrossRef]
15. Wambugu, P.W.; Ndjiondjop, M.; Henry, R. Advances in Molecular Genetics and Genomics of African Rice (*Oryza glaberrima* Steud). *Plants* **2019**, *8*, 376. [CrossRef]
16. Prodjinoto, H.; Irakoze, W.; Gandonou, C.; Quinet, M.; Lutts, S. Comparison between the Impact of Osmotic and NaCl Treatments on the Expression of Genes Coding for Ion Transporters in *Oryza glaberrima* Steud. *PLoS ONE* **2023**, *18*, e0290752. [CrossRef] [PubMed]
17. Adam, H.; Gutiérrez, A.; Couderc, M.; Sabot, F.; Ntakirutimana, F.; Serret, J.; Orjuela, J.; Tregear, J.; Jouannic, S.; Lorieux, M. Genomic Introgressions from African Rice (*Oryza glaberrima*) in Asian Rice (*O. sativa*) Lead to the Identification of Key QTLs for Panicle Architecture. *BMC Genom.* **2023**, *24*, 587. [CrossRef]
18. Ntakirutimana, F.; Tranchant-Dubreuil, C.; Cubry, P.; Chougule, K.; Zhang, J.; Wing, R.A.; Adam, H.; Lorieux, M.; Jouannic, S. Genome-Wide Association Analysis Identifies Natural Allelic Variants Associated with Panicle Architecture Variation in African Rice, *Oryza glaberrima* Steud. *G3 Genes Genomes Genet.* **2023**, *13*, jkad174. [CrossRef]
19. Robson, J.K.; Ferguson, J.N.; McAusland, L.; Atkinson, J.A.; Tranchant-Dubreuil, C.; Cubry, P.; Sabot, F.; Wells, D.M.; Price, A.H.; Wilson, Z.A.; et al. Chlorophyll Fluorescence-Based High-Throughput Phenotyping Facilitates the Genetic Dissection of Photosynthetic Heat Tolerance in African (*Oryza glaberrima*) and Asian (*Oryza sativa*) Rice. *J. Exp. Bot.* **2023**, *74*, 5181–5197. [CrossRef] [PubMed]
20. De Salvador, F.R.; Fideghelli, C.; Engel, P.; Giorgioni, M.; Palombi, M.A. Planta-Res, The Italian National Network on Plant Genetic Resources for Food and Agriculture. *Acta Hortic.* **2014**, *1032*, 215–219. [CrossRef]
21. Available online: https://www.reterurale.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/9580 (accessed on 1 April 2024).
22. Faivre-Rampant, O.; Bruschi, G.; Abbruscato, P.; Cavigiolo, S.; Picco, A.M.; Borgo, L.; Lupotto, E.; Piffanelli, P. Assessment of Genetic Diversity in Italian Rice Germplasm Related to Agronomic Traits and Blast Resistance (*Magnaporthe oryzae*). *Mol. Breed.* **2011**, *27*, 233–246. [CrossRef]
23. Mantegazza, R.; Biloni, M.; Grassi, F.; Basso, B.; Lu, B.R.; Cai, X.X.; Sala, F.; Spada, A. Temporal Trends of Variation in Italian Rice Germplasm over the Past Two Centuries Revealed by AFLP and SSR Markers. *Crop Sci.* **2008**, *48*, 1832–1840. [CrossRef]
24. Keilwagen, J.; Kilian, B.; Özkan, H.; Babben, S.; Perovic, D.; Mayer, K.F.X.; Walther, A.; Poskar, C.H.; Ordon, F.; Eversole, K.; et al. Separating the Wheat from the Chaff—A Strategy to Utilize Plant Genetic Resources from Ex Situ Genebanks. *Sci. Rep.* **2014**, *4*, 5231. [CrossRef]
25. Aleksander, S.A.; Balhoff, J.; Carbon, S.; Cherry, J.M.; Drabkin, H.J.; Ebert, D.; Feuermann, M.; Gaudet, P.; Harris, N.L.; Hill, D.P.; et al. The Gene Ontology Knowledgebase in 2023. *Genetics* **2023**, *224*, iyad031. [CrossRef] [PubMed]
26. Cooper, L.; Elser, J.; Laporte, M.-A.; Arnaud, E.; Jaiswal, P. Planteome 2024 Update: Reference Ontologies and Knowledgebase for Plant Biology. *Nucleic Acids Res.* **2024**, *52*, D1548–D1555. [CrossRef] [PubMed]
27. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
28. Philipp, N.; Weise, S.; Oppermann, M.; Börner, A.; Keilwagen, J.; Kilian, B.; Arend, D.; Zhao, Y.; Graner, A.; Reif, J.C.; et al. Historical Phenotypic Data from Seven Decades of Seed Regeneration in a Wheat Ex Situ Collection. *Sci. Data* **2019**, *6*, 137. [CrossRef] [PubMed]
29. Hao, X.; Yujun, S.; Xinjie, W.; Jin, W.; Yao, F. Linear Mixed-Effects Models to Describe Individual Tree Crown Width for China-Fir in Fujian Province, Southeast China. *PLoS ONE* **2015**, *10*, e0122257. [CrossRef] [PubMed]
30. Laird, N.M.; Ware, J.H. Random-Effects Models for Longitudinal Data. *Biometrics* **1982**, *38*, 963–974. [CrossRef] [PubMed]
31. José, C. *Pinheiro and Douglas M. Bates Mixed-Effects Models in S and S-PLUS*; Springer: New York, NY, USA, 2000; ISBN 0-387-98957-9.
32. Pinheiro, J.; Bates, D.; R Core Team. Nlme: Linear and Nonlinear Mixed Effects Models. 2023. Available online: https://cran.r-project.org/web/packages/nlme/nlme.pdf (accessed on 1 February 2024).
33. Gross, J.; Ligges, U. Nortest: Tests for Normality. 2015. Available online: https://cran.r-project.org/web/packages/nortest/nortest.pdf (accessed on 1 February 2024).
34. Ravelle, W. Psych: Procedures for Psychological, Psychometric, and Personality Research. 2024. Available online: https://cran.r-project.org/web/packages/psych/index.html (accessed on 1 February 2024).
35. Henderson, A.R. Testing Experimental Data for Univariate Normality. *Clin. Chim. Acta* **2006**, *366*, 112–129. [CrossRef] [PubMed]

36. Bourke, P.M.; Voorrips, R.E.; Hackett, C.A.; Van Geest, G.; Willemsen, J.H.; Arens, P.; Smulders, M.J.M.; Visser, R.G.F.; Maliepaard, C. Detecting Quantitative Trait Loci and Exploring Chromosomal Pairing in Autopolyploids Using PolyqtlR. *Bioinformatics* **2021**, *37*, 3822–3829. [CrossRef] [PubMed]
37. Giovanny, C.-P. Genome Assisted Prediction of Quantitative Traits Using the R Package *sommer*. *PLoS ONE* **2016**, *11*, e0156744. [CrossRef]
38. Lê, S.; Josse, J.; Husson, F. FactoMineR: An R. Package for Multivariate Analysis. *J. Stat. Softw.* **2008**, *25*, 1–18. [CrossRef]
39. Abatzoglou, J.T.; Dobrowski, S.Z.; Parks, S.A.; Hegewisch, K.C. TerraClimate, a High-Resolution Global Dataset of Monthly Climate and Climatic Water Balance from 1958–2015. *Sci. Data* **2018**, *5*, 170191. [CrossRef] [PubMed]
40. Matsushita, K.; Onogi, A.; Yonemaru, J. NARO Historical Phenotype Dataset from Rice Breeding. *Breed. Sci.* **2024**, *74*, 114–123. [CrossRef]
41. Rebollo, I.; Scheffel, S.; Blanco, P.; Molina, F.; Martínez, S.; Carracelas, G.; Aguilar, I.; Pérez de Vida, F.; Rosas, J.E. Consolidating 23 Years of Historical Data from an Irrigated Subtropical Rice Breeding Program in Uruguay. *Crop Sci.* **2023**, *63*, 1300–1315. [CrossRef]
42. Philipp, N.; Weise, S.; Oppermann, M.; Börner, A.; Graner, A.; Keilwagen, J.; Kilian, B.; Zhao, Y.; Reif, J.C.; Schulthess, A.W. Leveraging the Use of Historical Data Gathered during Seed Regeneration of an Ex Situ Genebank Collection of Wheat. *Front. Plant Sci.* **2018**, *9*, 609. [CrossRef] [PubMed]
43. Aubry, S. Genebanking Plant Genetic Resources in the Postgenomic Era. *Agric. Hum. Values* **2023**, *40*, 961–971. [CrossRef]
44. Zhao, H.; Fu, Y.; Zhang, G.; Luo, Y.; Yang, W.; Liang, X.; Yin, L.; Zheng, Z.; Wang, Y.; Li, Z.; et al. GS6.1 Controls Kernel Size and Plant Architecture in Rice. *Planta* **2023**, *258*, 42. [CrossRef] [PubMed]
45. Song, Y.H.; Shim, J.S.; Kinmonth-Schultz, H.A.; Imaizumi, T. Photoperiodic Flowering: Time Measurement Mechanisms in Leaves. *Annu. Rev. Plant Biol.* **2015**, *66*, 441–464. [CrossRef] [PubMed]
46. Sun, C.; Chen, D.; Fang, J.; Wang, P.; Deng, X.; Chu, C. Understanding the Genetic and Epigenetic Architecture in Complex Network of Rice Flowering Pathways. *Protein Cell* **2014**, *5*, 889–898. [CrossRef]
47. Ormoli, L.; Costa, C.; Negri, S.; Perenzin, M.; Vaccino, P. Diversity Trends in Bread Wheat in Italy during the 20th Century Assessed by Traditional and Multivariate Approaches. *Sci. Rep.* **2015**, *5*, 8574. [CrossRef]
48. Volante, A.; Tondelli, A.; Desiderio, F.; Abbruscato, P.; Menin, B.; Biselli, C.; Casella, L.; Singh, N.; McCouch, S.R.; Tharreau, D.; et al. Genome Wide Association Studies for Japonica Rice Resistance to Blast in Field and Controlled Conditions. *Rice* **2020**, *13*, 71. [CrossRef]
49. Zampieri, E.; Volante, A.; Marè, C.; Orasen, G.; Desiderio, F.; Biselli, C.; Canella, M.; Carmagnola, L.; Milazzo, J.; Adreit, H.; et al. Marker-Assisted Pyramiding of Blast-Resistance Genes in a Japonica Elite Rice Cultivar through Forward and Background Selection. *Plants* **2023**, *12*, 757. [CrossRef] [PubMed]
50. Volante, A.; Tondelli, A.; Aragona, M.; Valente, M.T.; Biselli, C.; Desiderio, F.; Bagnaresi, P.; Matic, S.; Gullino, M.L.; Infantino, A.; et al. Identification of Bakanae Disease Resistance Loci in Japonica Rice through Genome Wide Association Study. *Rice* **2017**, *10*, 29. [CrossRef] [PubMed]
51. Titone, P.; Mongiano, G.; Tamborini, L. Resistance to Neck Blast Caused by *Pyricularia oryzae* in Italian Rice Cultivars. *Eur. J. Plant Pathol.* **2015**, *142*, 49. [CrossRef]
52. Oliveira-Garcia, E.; Valent, B. How Eukaryotic Filamentous Pathogens Evade Plant Recognition. *Curr. Opin. Microbiol.* **2015**, *26*, 92–101. [CrossRef]
53. Valent, B.; Khang, C.H. Recent Advances in Rice Blast Effector Research. *Curr. Opin. Plant Biol.* **2010**, *13*, 434–441. [CrossRef]

Tesi di dottorato di ricerca co-finanziata nell'ambito del Programma Operativo Nazionale Ricerca e Innovazione 2014-2020 (CCI 2014IT16M2OP005), risorse FSE REACT-EU, Azione IV.4 “Dottorati e contratti di ricerca su tematiche dell’innovazione” e Azione IV.5 “Dottorati su tematiche Green”.