e-ISSN 2965-7342

Documentos

141

Jaguariúna, SP / May, 2025

Exploiting the Genetic Sequencing of the Soil Microbiome in Agriculture

Embrapa
Meio Ambiente

Brazilian Agricultural Research Corporation
Embrapa Environment
Ministry of Agriculture, Livestock and Food Supply

e-ISSN 2965-7342

Documentos 141

May, 2025

Exploiting the Genetic Sequencing of the Soil Microbiome in Agriculture

Rodrigo Mendes
Lucas William Mendes
Fernando Dini Andreote

Embrapa Environment
Jaguariúna, SP
2025

Embrapa Environment
Rodovia SP 34, Km 127,5
Tanquinho Velho
www.embrapa.br/meio-ambiente
www.embrapa.br/fale-conosco/sac

Local publishing committee

President
Janaína Paula Marques Tanure

Executive secretary
Anderson Soares Pereira

Members
Janaína Paula Marques Tanure, Robson Rolland Monticelli Barizon, Alfredo José Barreto Luiz, Fagoni Fayer Calegario, Marcos Eliseu Losekann, Vera Lúcia Ferracini, Victor Paulo Marques Simão, Julio Ferraz de Queiroz, Márcia Regina Assalin, Maria de Cléofas Faggion Alencar e Sonia Claudia do Nascimento de Queiroz

Executive edition
Anderson Soares Pereira

Text revision
Reinaldo Rodrigues

Bibliographic standardization
Victor Paulo Marques Simão (CRB-8/5139)

Graphic design
Leandro Sousa Fazio

Layout
Silvana Cristina Teixeira

Cover
Rodrigo Mendes

Digital publishing: PDF

International Cataloging-in-Publication Data (CIP)
Embrapa Environment

Mendes, Rodrigo.

Exploiting the genetic sequencing of the soil microbiome in agriculture / Rodrigues Mendes, Lucas William Mendes, Fernando Dini Andreote. – Jaguariúna : Embrapa Meio Ambiente, 2025.

PDF (15 p.) : il. color. – (Documentos / Embrapa Meio Ambiente, e-ISSN 2965-7342 ; 141)

Text originally published in portuguese.

Translation of: Explorando o sequenciamento genético do microbioma do solo na agricultura.

1. Microbiome. 2. Soil. 3. Genetic sequencing. I. Mendes, Lucas William. II. Andreote, Fernando Dini. III. Série.

CDD (21. ed.) 574.88

Maria de Cléofas Faggion Alencar (CRB-8/1658)

Authors

Rodrigo Mendes
Agricultural engineer, doctor in Genetics and Plant Improvement, researcher at Embrapa Environment, Jaguariúna, São Paulo, Brazil

Lucas William Mendes
Biologist, doctor in Sciences, professor at the Center for Nuclear Energy in Agriculture, University of São Paulo, Piracicaba, SP, Brazil

Fernando Dini Andreote
Agricultural engineer, doctor in Genetics and Plant Improvement, full professor at the Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, São Paulo, Brazil

Foreword

The health of agricultural soils stems from the interaction among their physical, chemical, and biological properties. While the chemical and physical dimensions have traditionally been managed in agricultural systems, the biological component, including soil microorganisms, has received growing attention in recent years.

The concept of One Health, established in 2008 by the World Health Organization (WHO), the World Organisation for Animal Health (OIE), and the Food and Agriculture Organization (FAO), highlights the interconnection between human, animal, plant, and environmental health. In this context, the microbiome stands out as an essential link, connecting and influencing the health of living beings and the environment.

Since soil is an important reservoir of microbial diversity, it can be affirmed that healthy soils form the foundation for the health of plants, animals, and people. Thus, understanding and managing the soil microbiome are crucial to fostering sustainable agriculture and, consequently, advancing the principles of One Health.
In alignment with the mission of Embrapa Meio Ambiente, which seeks to integrate agricultural production with sustainability in all its dimensions, managing the soil microbiome directly contributes to achieving Sustainable Development Goal (SDG) 2: Zero hunger - to end hunger, achieve food security and improved nutrition and promote sustainable agriculture.

In this text, the authors provide an overview of the rapid advancements in molecular biology techniques, focusing on genetic sequencing, and discuss how this tool can be applied in production systems. This text can also be seen as a guide to effectively analyzing and fully leveraging genetic sequencing data of the soil microbiome in agriculture.

Ana Paula Contador Packer
General Head
Embrapa Environment

Sumary

Introduction

The emergence of advanced genetic sequencing techniques revolutionized several areas of science in the late 20th and early 21st centuries. The major milestone was the completion of the sequencing of the human genome in 2003 (Schmutz et al., 2004). The Human Genome Project (Lander et al., 2001) emerged in the early 1980s from two fundamental perceptions. Firstly, that the ability to know the genome could significantly accelerate biomedical research, allowing researchers to approach problems in a comprehensive and unbiased manner. Secondly, the development of this project would require an enormous effort on the part of the scientific community to build up research infrastructure, the like of which has never been seen before in the biological sciences. Interestingly, the first multicellular organism to have its genome completely sequenced was not man, but a small free-living nematode, Caenorhabditis elegans, which makes up the soil microbiome (Genome...,1998). In Brazil, inspired by the two fundamental insights that drove the Human Genome Project, in the same period, an initiative coordinated by the São Paulo State Research Foundation (FAPESP), with support from the National Council for Scientific and Technological Development (CNPq) and Fundecitrus, was launched for the first sequencing of a plant pathogen. The complete sequencing of the genome of the pathogenic bacterium Xylella fastidiosa, which causes Citrus Variegated Chlorosis (CVC), was published in 2000 (Simpson et al., 2000), which then allowed a detailed comparison of the complete genomes of pathogens from animals and plants.

After the first complete genomes were published, various research groups were set up around the world and important advances were made in the evolution and availability of sequencing technologies. An important indicator that reflects the popularization of the application of genetic sequencing in various areas of science, including agriculture, is the cost of sequencing. In 2001, when the sequencing of the human genome was completed, the estimated cost for complete sequencing was US$ 95.3 million. In 2022, the cost of sequencing a complete human genome was US$ 525 (DNA..., 2024). This impressive cost reduction is justified by the emergence of “second-generation DNA sequencing”, which encompasses several massively parallel sequencing technologies with ultra-throughput, scalability and speed.

In agriculture, the physical, chemical and biological dimensions of the soil are intrinsically linked and determine the development and health of cultivated plants. With regard to the biological component of soils, our view of their functional composition has for a long time remained restricted to groups of organisms accessible by cultivation methods, which access less than 1% of the microbiological biodiversity of soils (Pham; Kim, 2012; Anthony et al., 2023). In this context, the application of sequencing techniques has been fundamental in elucidating the complex interactions between the soil microbiome and plants. However, most studies of the soil microbiome using second-generation sequencing are still descriptive. Translating theoretical knowledge into practical application in sustainable agriculture requires the development of approaches and concepts that provide a deeper understanding of the mechanisms involved in these microbiome interactions. In this text, we will put into perspective the potential and limitations of exploiting genetic sequencing data from the soil microbiome in agriculture. By critically examining these issues, we aim to contribute to the transition from a merely theoretical understanding to practical strategies that drive agriculture towards more sustainable and effective pattern.

Genetic sequencing of DNA (or RNA) obtained from soils in agricultural production systems allows investigation of microbial interactions in the soil, rhizosphere and other plant-associated environments. Researchers can study the effects of agricultural practices, including crop rotation and the use of fertilizers and bio-inputs on soil microbial communities. They can also understand the interaction between the soil microbiome and

crop plants, and target sustainable agricultural practices that promote biodiversity and/or microbial activity, resulting in improvements in agricultural productivity, promoted by processes such as nutrient cycling and other processes related to plant health.

Soil DNA or RNA sequencing approaches and objectives

EThere are two main approaches to globally assessing the composition and functionality of soil microbial communities by isolating and analyzing their genetic material. The first approach accesses conserved ribosomal DNA marker genes, which are used primarily to determine the identity (taxonomy) of the microorganisms in a community, giving rise to a technique called amplicon sequencing or metataxonomics. The second, accesses virtually all the DNA (or RNA) segments obtained from a soil sample, making it possible to analyze both taxonomic marker genes and genes related to specific functions present in the microbiome. This approach is called metagenomics, when it is carried out using DNA, or metatranscriptomics, when it is carried out using RNA. The term metagenomics is often used in a broad but technically erroneous way when it includes the approach of sequencing amplicons. These methods allow us to study the genetic and functional diversity of microorganisms present in the soil without the need to grow them in culture media in the laboratory, exponentially increasing our effectiveness in analyzing these communities. The analytical strategies of each approach are illustrated in Figure 1 and described in the following sections.

Soil DNA and RNA extraction

The first step in studying the soil microbiome is to extract DNA and/or RNA from the samples, thus allowing access to the genetic material of the microorganisms present and, subsequently, detailed analysis of the microbial diversity and functions present in the environment to be studied. Different extraction methods are used, each with its advantages and disadvantages. In general, the process involves rupturing the microbial cells present in the soil sample, followed by purification of the DNA or RNA. The choice of extraction method can significantly affect the results obtained, influencing the efficiency with which the genetic material is recovered and the representativeness of the microbiome studied.

In addition, it is important to consider the specific characteristics of each soil, such as the presence of organic and inorganic materials that can interfere with the extraction and quality of the genetic material recovered. Methods that take into account the complexity and variability of the soil, such as extraction with physical and chemical destabilizing agents followed by careful purification, are often used to obtain more accurate results. The effective extraction of DNA and RNA from soil not only allows detailed analysis of microbial composition and function, but also contributes to a better understanding of the ecological and biogechemical processes taking place in this important environmental compartment.

It is worth noting that there are technical differences between extracting DNA and RNA from soil samples, based mainly on the molecular characteristics and specific extraction techniques required for each type of nucleic acid. For DNA, the main challenge lies in the efficient removal of compounds that can inhibit subsequent processes, such as the PCR reaction, such as polyphenols and humic compounds present in the soil. Extracting RNA from soil presents additional challenges due to its more unstable nature and the presence of ribonuclease enzymes (RNases), which can rapidly degrade RNA. RNA extraction methods often include the addition of stabilizing agents, such as phenol or guanidine thiocyanate, to inactivate the RNases and preserve the RNA. In addition, RNA extraction often requires additional steps, such as the removal of residual DNA by treatment with DNAse and the use of silica columns or specific resins to purify the RNA. Currently, several commercial kits are available on the market, offering fast and efficient strategies for extracting DNA and RNA from soil samples.

Metataxonomics (amplicon sequencing)

In this approach, marker ribosomal genes are used as targets, which are present in all organisms in the target study group. These genes, known as essential or housekeeping genes, perform fundamental functions for cell survival. Ribosomal genomic regions are commonly chosen for this type of study because they present conserved and varied regions, which allow for the precise identification of specific groups of organisms, such as bacteria, fungi and other eukaryotes, and the discrimination of members of these groups with more detailed taxonomic resolution.

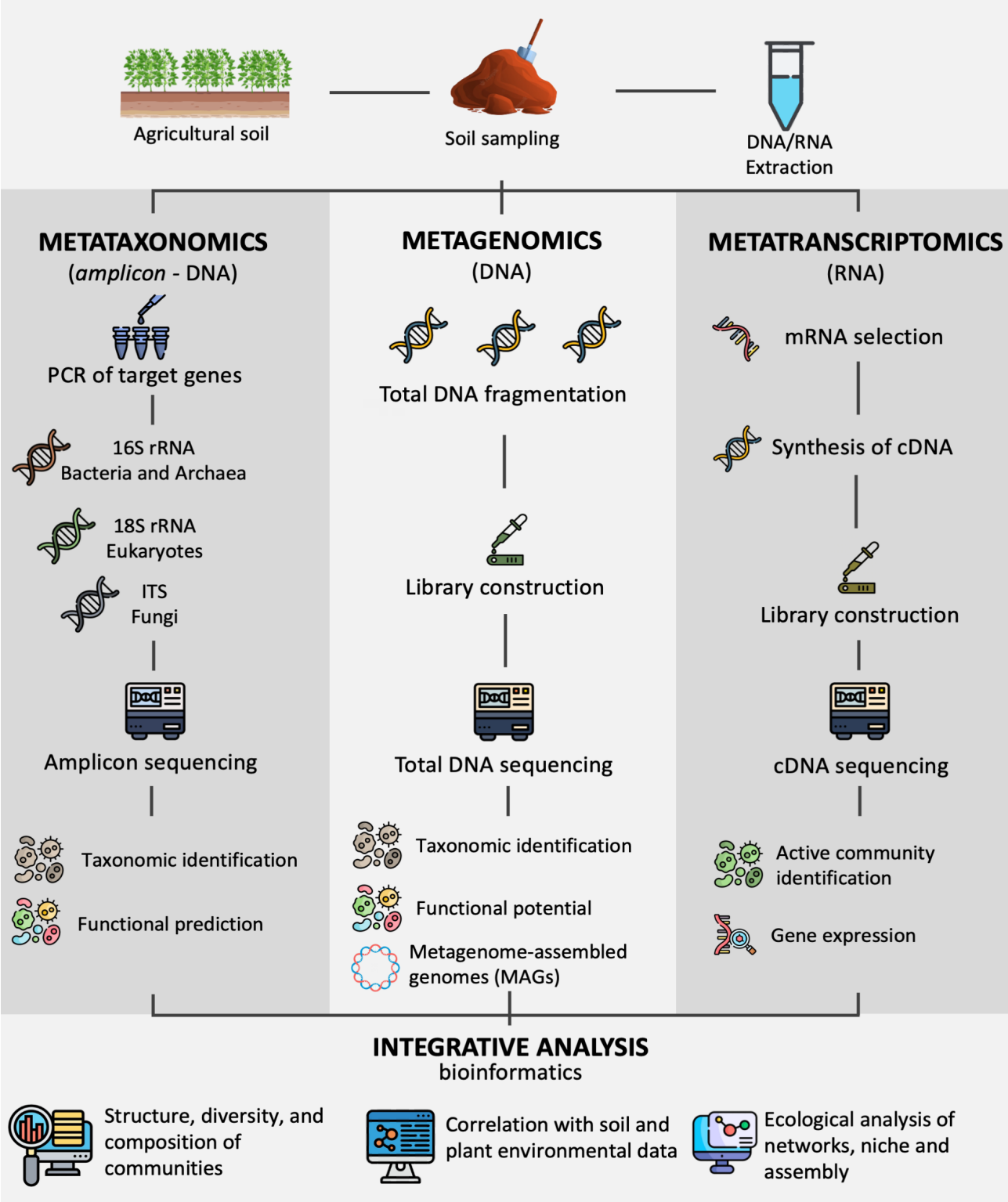

Figure 1. General outline of analytical approaches to study soil microbiome in agricultural systems.

The variable portions of the gene can be used for identification and phylogeny studies. Amplicon sequencing involves amplification by Polymerase Chain Reaction (PCR) using specific primers according to the target gene, 16S rRNA (bacteria and archaea), ITS (fungi) and 18S rRNA (other eukaryotes). The PCR product is then used for massive sequencing, generating thousands of DNA sequences, which are subjected to analysis with bioinformatics tools.
After processing the quality of the sequences obtained, those of low quality are removed, and clusters of high-quality sequences are used to carry out diversity analyses, identification and comparison of communities. The most popular databases for taxonomic assignment are Silva, GreenGenes, Ribosomal Database Project (RDP) for bacteria and archaea, and some specific ones such as UNITE for fungi and PR2 for protists. It is important to note that, due to the limited size of the sequenced DNA fragment, the classification of the sequences rarely allows for a description down to species level of the microorganisms present in soils. The taxonomic determination of the sequences obtained can vary from domain to genus level. In addition to taxonomic affiliation, the sequences are also analyzed for similarity, which generates empirical taxonomic group classifications, from which it is possible to statistically estimate the number of species present in a sample, or other diversity parameters inherent to it. Innovations in this application seek to predict functionalities from sequencing data, as well as to draw negative or positive correlations between microbial groups present in the same sample, and to determine microbial groups that are essential to the organization of the microbial community under study. All of these functionalities have their efficiency related to the data set used, where the development of bioinformatics and artificial intelligence, such as machine learning, plays a fundamental role.

Metagenomics (DNA sequencing) and Metatranscriptomics (RNA sequencing)

In this approach, total DNA or RNA is isolated and purified from soil samples and then processed, fragmented and used for sequencing. This allows access to all the genetic material recovered from the soil, and not just marker genes, as in the metataxonomic approach, allowing the identification of bacteria, fungi and other organisms. It also provides information on the biological functions and metabolic potential of these microorganisms. When the analysis is based on DNA, the information on functional genes represents the functional potential of the microbiome in the soil sample evaluated. When the analysis is based on RNA, the sequencing data results from the genes actually expressed by the soil microbiome at the time of sampling. This approach provides a more comprehensive view of microbial diversity compared to traditional culture methods, allowing an understanding of microbial ecology, metabolic functions and potential environmental impacts on the soil. However, these analyses are more laborious, more costly and more complex to interpret. The databases commonly used for functional assignment of metagenomic data are MG-RAST, KEGG, IMG/M, SEED, Uniprot, and pFAM. Other databases are dedicated to the annotation of specific metabolic pathways such as NcycDB, dbCAN and Resfam, respectively for the nitrogen cycle, carbohydrates and antibiotic resistance genes. Also, from the data generated from metagenomics, there is the possibility of assembling genomes of organisms present in the samples, known as metagenome-assembled genomes (MAGs). This computational technique has advanced and allows us to access genomic information from non-cultivated organisms, which gives us the ability to connect microbial functions in the soil with the taxonomic classification of microbial groups that have been poorly studied.

Analytical roadmap for genetic sequencing data

To analyze genetic sequencing data from the microbiome, a comprehensive roadmap involves several steps and different bioinformatics tools. Initially, it is essential to pre-process the sequences, including quality checks, removal of adapters and low-quality bases. After filtering the sequences, it is important to conduct analyses of the diversity, structure and composition of microbial communities, assigning sequences to taxonomic categories, calculating diversity metrics and comparing community structure between different samples and treatments. The visualizations of these inferences are quite diverse, ranging from numbers to dispersions of samples by similarity in community composition. Functional analysis can be carried out either by

prediction, from ribosomal genes, or inference by sequences obtained from metagenomic and metatranscriptomic approaches, predicting metabolic functions and identifying specific genes of interest. To understand differences between groups of samples, statistical tests are applied, with correction for multiple comparisons if necessary. Integration with metadata is crucial to link microbiome data with relevant environmental information. Visualizing the data using graphs and figures helps to visually represent the results. After interpreting the results in the light of the study's objectives and existing literature, it is possible to discuss the biological implications and suggest future studies. Finally, it is essential to document in detail all the methods, parameters and results obtained in order to prepare reports or scientific articles. An analytical roadmap for sequencing data is shown in Figure 2 and described below:

1) Pre-processing the sequencing data: check the quality of the raw sequencing data, for example, using programs such as FastQC; remove adapters and low-quality sequenced bases, for example, using tools such as Trimmomatic or Cutadapt; and assemble sequences if applicable.

2) Analysis of diversity, structure and composition of the microbiome: analyses to assign sequences to taxonomic categories using specific databases, using software such as QIIME and mothur; calculate diversity metrics such as richness and alpha diversity; compare the structure of the microbial communities between the different treatments by means of scatter plots, for example, using multivariate analyses such as PCA, PCoA, NMDS, RDA, and CCA; identify the species present and their relative abundance, for example, using databases such as RDP, Silva, and Greengenes; and assemble genomes (MAGs) from metagenome data.

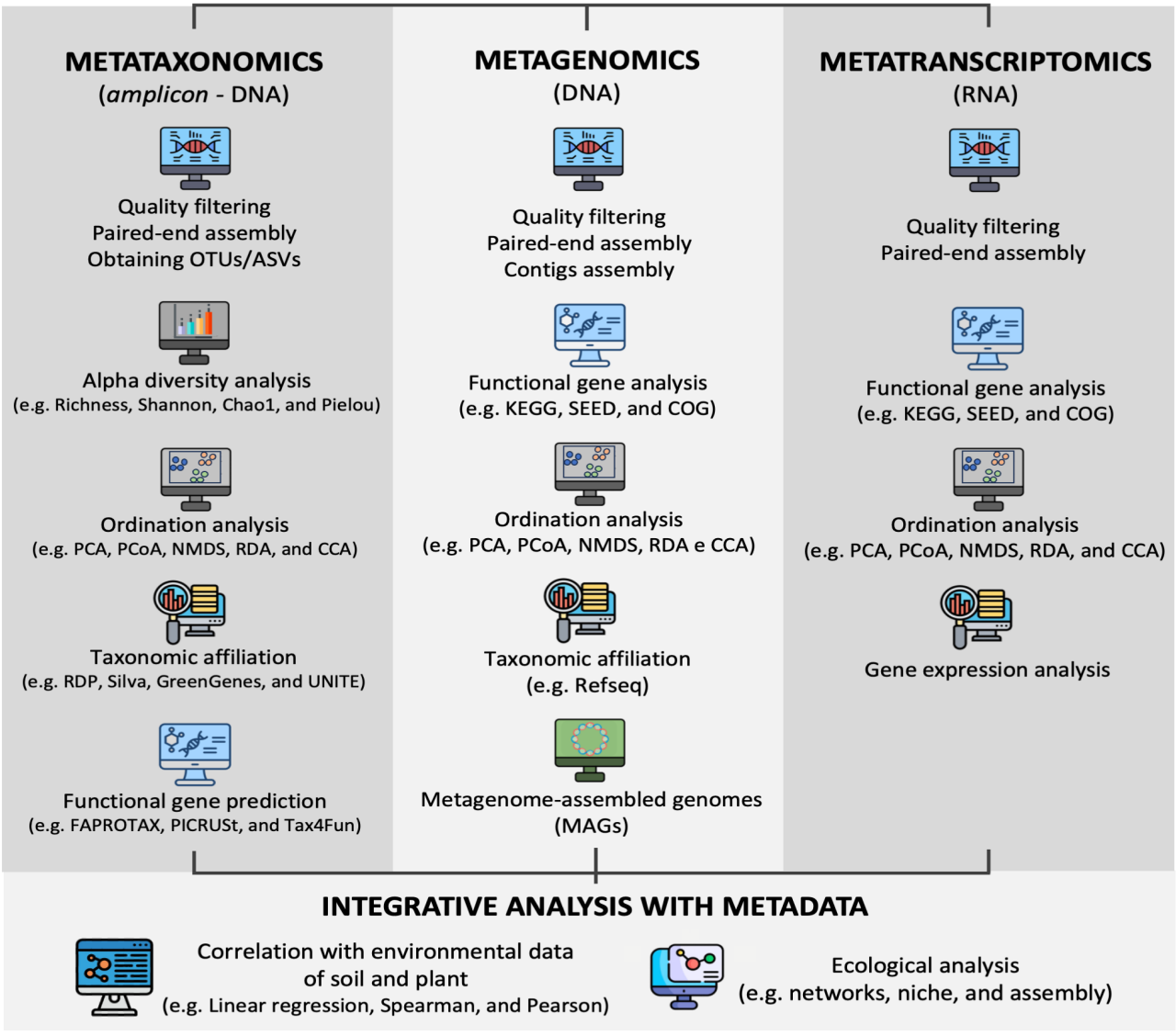

Figure 2. General outline of analytical approaches to study soil microbiome in agricultural systems.

3) Functional analysis of the microbiome: perform functional prediction from amplicon data, for example, using databases such as FAPROTAX, Tax4Fun2, PICRUSt2 and FUNGuild, and perform functional inference identifying genes of interest from metagenomic data, for example using databases such as KEGG, SEED and COG

4) Differential analysis for comparing groups: statistical tests to identify significant differences in the taxonomic or functional composition of microorganisms between groups of samples using parametric and non-parametric tests, for example, using t-test, ANOVA, DESeq2, edgeR, ALDEx2, ANCOM, LEfSe, and Random Forest.

5) Integration of metadata: association of microbiome sequencing data with environmental information or other relevant metadata to investigate correlations or patterns, for example, with soil chemical analysis data, disease occurrence and crop yield. Correlation analyses can be carried out using statistical methods such as linear regression, Spearman's correlation, Pearson's correlation, among others.

6) Data visualization and interpretation: creation of graphs and figures to visually represent the results of the analyses, such as bar charts, Venn diagrams, heatmaps, network analyses, among others; interpretation of the results obtained in the light of the objectives of the study and the existing literature; discussion of the biological implications of the results observed; and management recommendations based on the study findings.

Opportunities and challenges of microbiome sequencing applied to agricultural systems

The activity of the soil microbiome plays a determining role in plant productivity in agricultural systems (Tkacz; Poole, 2015). In a metagenomic sequencing study comparing areas of low and high agricultural productivity, it was shown that differences in productivity are associated with the composition of the microbiome (Chang et al., 2017). The Metagenome-Wide Association Study (MWAS), combined with machine learning prediction, identified specific groups of microorganisms associated with high productivity and revealed that the abundance of microbiome members associated with nitrogen transformations are determinants of high productivity (Chang et al., 2017). In a metataxonomic study of the bacterial community, the effect of microbial diversity on soybean productivity and the rate of infection by the gall nematode (Meloidogyne spp.) was compared. The results showed that soils with greater bacterial diversity had a lower rate of infection by the pathogen and higher productivity (Barros et al., 2022), highlighting the importance of soil microbial diversity for plant health and growth. Thus, considering the essentiality of the microbiome for plants, genetic sequencing analysis of the soil microbiome in agricultural systems has the potential to predict the health and productivity of agricultural systems.

There is an effort by the scientific community to organize a robust system of indicators to monitor soil health. Given that the microbiome contains information on the biological, chemical and physical state of the soil, it can serve as an integrated measure of soil health. Wilhelm et al. (2022) demonstrated the potential of using soil microbiome data to predict soil health accurately and efficiently. By analyzing a 16S rRNA gene amplicon sequencing dataset, the researchers developed machine learning models capable of predicting various soil health metrics, including biological, chemical and physical properties. The results indicate that the composition of the soil microbiome can serve as a reliable indicator of soil quality, allowing for the creation of more accurate and accessible diagnostic tools to monitor the health of agroecosystems (Wilhelm et al., 2022). In another study carried out to predict disease occurrence in six crops with high accuracy, the authors investigated the relationship between the distribution of microorganisms in the soil and the occurrence of Fusarium wilt. Using complex data analysis techniques, they identified that the presence of certain types of fungi in the soil was strongly correlated with the appearance of the disease. The method identified 45 bacterial OTUs (Operational Taxonomic Units) and 40 fungal OTUs that categorized the state of soil health with an accuracy of over 80%. The conclusion of the study suggests that analyzing the composition of the soil microbial community can be a useful tool for predicting and controlling this disease in agricultural crops (Yuan et al., 2020). Sequencing, using primers for the ITS region and primers targeting arbuscular mycorrhizal fungi, was used to predict the success of the fungus inoculation in 54 maize fields distributed in Switzerland. The soil microbiome indicators, together with some soil parameters, were able to predict 86% of the

variation in corn growth in response to inoculation. The abundance of pathogenic fungi, rather than the availability of nutrients, was the best predictor of the inoculant's success. This predictive capacity extends the potential of using the microbiome as a tool for sustainable agricultural management (Lutz et al., 2023).

In terms of pathogen detection and its implications, there are several particularities that must be considered when using genetic sequencing techniques to predict disease occurrence. The potential of genetic sequencing to identify microbial groups, including plant pathogens, is undeniable. However, when we retrieve genetic material from complex samples, the sequenced fragments, whether by metataxonomics or metagenomics, do not usually provide a resolution to infer the pathogenic potential of the organisms detected. This occurs because, for many groups with this potential, infective capacity is highly specific, either to a species or variation within the species (for example, forma specialis) or related to one or a few genes present in strains with a higher infection potential. In this case, complementary molecular analyses, such as quantitative PCR targeting specific pathogens, can be useful for the accurate identification and quantification of the pathogen in soil samples.

Challenges that remain for using microbiome information to predict soil health or productivity include increasing the depth of sequencing, i.e. obtaining a greater number of sequences per soil sample; increasing resolution in taxonomic classification; and increasing the number of samples that represent the variation and diversity of soils with different levels of health (Wilhelm et al., 2022). Additionally, prediction models based on microbiome sequencing will perform better as they are customized and adapted considering soil type, biogeography and regional differences in the soil microbiome (Gschwend et al., 2021) and the production system, due to the legacy effect of the plants grown in the system (Schmid et al., 2019).

Concluding remarks

Genetic sequencing of the soil microbiome represents a promising tool for sustainable agricultural management, helping to predict soil health, disease occurrence and productivity. However, there are still challenges in standardizing and optimizing these techniques so that they can be widely applied. Combining sequencing data with artificial intelligence and machine learning can accelerate this transition, enabling a more efficient and sustainable approach to agriculture. The evolution of sequencing technologies, coupled with the advancement of analytical tools based on artificial intelligence, has the potential to radically transform agricultural management. In the future, these technologies are expected to become increasingly accessible and personalized, adapting to different types of soil and crop systems. Furthermore, developing more robust databases and region-specific predictive models will enhance the effective utilization of the soil microbiome, supporting a more sustainable and productive agriculture.

References

ANTHONY, M. A.; BENDER, S. F.; VAN DER HEIJDEN, M. G. A. Enumerating soil biodiversity. **Proceedings of the National Academy of Sciences of the United States of America**, v. 120, n. 33, p. e2304663120, Aug. 2023. DOI: https://doi.org/10.1073/pnas.2304663120.

BARROS, F.; PEDRINHO, A.; MENDES, L. W.; FREITAS, C. C. G.; ANDREOTE, F. D. Interactions between soil bacterial diversity and plant-parasitic nematodes in soybean plants. **Applied and Environmental Microbiology**, v. 88, n. 17, p. e0096322, Sep. 2022. DOI: https://doi.org/10.1128/aem.00963-22.

CHANG, H. X.; HAUDENSHIELD, J. S.; BOWEN, C. R.; HARTMAN, G. L. Metagenome-Wide Association Study and Machine Learning Prediction of Bulk Soil Microbiome and Crop Productivity. **Frontiers in Microbiology**, v. 8, p. 519, Apr. 2017. DOI: https://doi.org/10.3389/fmicb.2017.00519.

DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). **National Human Genome Research Institute**. Disponível em: http://www.genome.gov/sequencingcostsdata. Acesso em: 18 nov. 2024.

GENOME sequence of the nematode *C. elegans*: a platform for investigating biology. **Science**, v. 282, n. 5396, p. 2012-2018, Dec. 1998. DOI: https://doi.org/10.1126/science.282.5396.2012.

GSCHWEND, F.; HARTMANN, M.; HUG, A. S.; ENKERLI, J.; GUBLER, A.; FREY, B.; MEULI, R. G.; WIDMER, F.

Long-term stability of soil bacterial and fungal community structures revealed in their abundant and rare fractions. **Molecular Ecology**, v. 30, n. 17, p. 4305-4320, Sept. 2021. DOI: https://doi.org/10.1111/mec.16036.

LANDER, E. S.; LINTON, L. M.; BIRREN, B.; NUSBAUM, C.; ZODY, M. C.; BALDWIN, J.; DEVON, K.; DEWAR, K.; DOYLE, M.; FITZHUGH, W.; FUNKE, R.; GAGE, D.; HARRIS, K.; HEAFORD, A.; HOWLAND, J.; KANN, L.; LEHOCZKY, J.; LEVINE, R.; MCEWAN, P.; MCKERNAN, K.; MELDRIM, J.; MESIROV, J. P.; MIRANDA, C.; MORRIS, W.; NAYLOR, J.; RAYMOND, C.; ROSETTI, M.; SANTOS, R.; SHERIDAN, A.; SOUGNEZ, C.; STANGE-THOMANN, Y.; STOJANOVIC, N.; SUBRAMANIAN, A.; WYMAN, D.; ROGERS, J.; SULSTON, J.; AINSCOUGH, R.; BECK, S.; BENTLEY, D.; BURTON, J.; CLEE, C.; CARTER, N.; COULSON, A.; DEADMAN, R.; DELOUKAS, P.; DUNHAM, A.; DUNHAM, I.; DURBIN, R.; FRENCH, L.; GRAFHAM, D.; GREGORY, S.; HUBBARD, T.; HUMPHRAY, S.; HUNT, A.; JONES, M.; LLOYD, C.; MCMURRAY, A.; MATTHEWS, L.; MERCER, S.; MILNE, S.; MULLIKIN, J.C.; MUNGALL, A.; PLUMB. R.; ROSS. M.; SHOWNKEEN, R.; SIMS, S.; WATERSTON, R. H.; WILSON, R. K.; HILLIER, L. W.; MCPHERSON, J. D.; MARRA, M. A.; MARDIS, E. R.; FULTON, L. A.; CHINWALLA, A. T.; PEPIN, K. H.; GISH, W. R.; CHISSOE, S. L.; WENDL, M. C.; DELEHAUNTY, K. D.; MINER, T. L.; DELEHAUNTY, A.; KRAMER, J. B.; COOK, L. L.; FULTON, R. S.; JOHNSON, D. L.; MINX, P. J.; CLIFTON, S. W.; HAWKINS, T.; BRANSCOMB, E.; PREDKI, P.; RICHARDSON, P.; WENNING, S.; SLEZAK, T.; DOGGETT, N.; CHENG, J. F.; OLSEN, A.; LUCAS, S.; ELKIN, C.; UBERBACHER, E.; FRAZIER, M.; GIBBS, R. A.; MUZNY, D. M.; SCHERER, S. E.; BOUCK, J. B.; SODERGREN, E. J.; WORLEY, K. C.; RIVES, C. M.; GORRELL, J. H.; METZKER, M. L.; NAYLOR, S. L.; KUCHERLAPATI, R. S.; NELSON, D. L.; WEINSTOCK, G. M.; SAKAKI, Y.; FUJIYAMA, A.; HATTORI, M.; YADA, T.; TOYODA, A.; ITOH, T.; KAWAGOE, C.; WATANABE, H.; TOTOKI, Y.; TAYLOR, T.; WEISSENBACH, J.; HEILIG, R.; SAURIN, W.; ARTIGUENAVE, F.; BROTTIER, P.; BRULS, T.; PELLETIER, E.; ROBERT, C.; WINCKER, P.; SMITH, D. R.; DOUCETTE-STAMM, L.; RUBENFIELD, M.; WEINSTOCK, K.; LEE, H. M.; DUBOIS, J.; ROSENTHAL, A.; PLATZER, M.; NYAKATURA, G.; TAUDIEN, S.; RUMP, A.; YANG, H.; YU, J.; WANG, J.; HUANG, G.; GU, J.; HOOD, L.; ROWEN, L.; MADAN, A.; QIN, S.; DAVIS, R. W.; FEDERSPIEL, N. A.; ABOLA, A. P.; PROCTOR, M. J.; MYERS, R. M.; SCHMUTZ, J.; DICKSON, M.; GRIMWOOD, J.; COX, D. R.; OLSON, M. V.; KAUL, R.; RAYMOND, C.; SHIMIZU, N.; KAWASAKI, K.; MINOSHIMA, S.; EVANS, G. A.; ATHANASIOU, M.; SCHULTZ, R.; ROE, B. A.; CHEN, F.; PAN, H.; RAMSER, J.; LEHRACH, H.; REINHARDT, R.; MCCOMBIE, W. R.; DE LA BASTIDE, M.; DEDHIA, N.; BLÖCKER, H.; HORNISCHER, K.; NORDSIEK, G.; AGARWALA, R.; ARAVIND, L.; BAILEY, J. A.; BATEMAN, A.; BATZOGLOU, S.; BIRNEY, E.; BORK, P.; BROWN, D. G.; BURGE, C. B.; CERUTTI, L.; CHEN, H. C.; CHURCH, D.; CLAMP, M.; COPLEY, R. R.; DOERKS, T.; EDDY, S. R.; EICHLER, E. E.; FUREY, T. S.; GALAGAN, J.; GILBERT, J. G.; HARMON, C.; HAYASHIZAKI, Y.; HAUSSLER, D.; HERMJAKOB, H.; HOKAMP, K.; JANG, W.; JOHNSON, L. S.; JONES, T. A.; KASIF, S.; KASPRYZK, A.; KENNEDY, S.; KENT, W. J.; KITTS, P.; KOONIN, E. V.; KORF, I.; KULP, D.; LANCET, D.; LOWE, T. M.; MCLYSAGHT, A.; MIKKELSEN, T.; MORAN, J. V.; MULDER, N.; POLLARA, V. J.; PONTING, C. P.; SCHULER, G.; SCHULTZ, J.; SLATER, G.; SMIT, A. F.; STUPKA, E.; SZUSTAKOWKI, J.; THIERRY-MIEG, D.; THIERRY-MIEG, J.; WAGNER, L.; WALLIS, J.; WHEELER, R.; WILLIAMS, A.; WOLF, Y. I.; WOLFE, K. H.; YANG, S. P.; YEH, R. F.; COLLINS, F.; GUYER, M. S.; PETERSON, J.; FELSENFELD, A.; WETTERSTRAND, K. A.; PATRINOS, A.; MORGAN, M. J.; DE JONG, P.; CATANESE, J. J.; OSOEGAWA, K.; SHIZUYA, H.; CHOI, S.; CHEN, Y. J.; SZUSTAKOWKI, J.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860-921, Feb. 2001. DOI: https://doi.org/10.1038/35057062.

LUTZ, S.; BODENHAUSEN, N.; HESS, J.; VALZANO-HELD, A.; WAELCHLI, J.; DESLANDES-HÉROLD, G.; SCHLAEPPI, K.; VAN DER HEIJDEN, M. G. A. Soil microbiome indicators can predict crop growth response to large-scale inoculation with arbuscular mycorrhizal fungi. **Nature Microbiology**, v. 8, n. 12, p. 2277-2289, Dec. 2023. DOI: https://doi.org/10.1038/s41564-023-01520-w.

PHAM, V. H.; KIM, J. Cultivation of unculturable soil bacteria. **Trends in Biotechnology**, v. 30, n. 9, p. 475-484, Sept. 2012. DOI: https://doi.org/10.1016/j.tibtech.2012.05.007.

SCHMID, M. W.; HAHL, T.; VAN MOORSEL, S. J.; WAGG, C.; DENY, G. B. de; SCHMID, B. Feedbacks of plant identity and diversity on the diversity and community composition of rhizosphere microbiomes from a long-term biodiversity experiment. **Molecular Ecology**, v. 28, n. 4, p. 863-878, Feb. 2019. DOI: https://doi.org/10.1111/mec.14987.

SCHMUTZ, J.; WHEELER, J.; GRIMWOOD, J.; DICKSON, M.; YANG, J.; CAOILE, C.; BAJOREK, E.; BLACK, S.; CHAN, Y. M.; DENYS, M.; ESCOBAR, J.; FLOWERS, D.; FOTOPULOS, D.; GARCIA, C.; GOMEZ,

M.; GONZALES, E.; HAYDU, L.; LOPEZ, F.; RAMIREZ. L.; RETTERER, J.; RODRIGUEZ, A.; ROGERS, S.; SALAZAR, A.; TSAI, M.; MYERS, R. M. Quality assessment of the human genome sequence. **Nature**, v. 429, n. 6990, p. 365-368, May 2004. DOI: https://doi.org/10.1038/nature02390.

SIMPSON, A. J.; REINACH, F. C.; ARRUDA, P.; ABREU, F. A.; ACENCIO, M.; ALVARENGA, R.; ALVES, L. M.; ARAYA, J. E.; BAIA, G. S.; BAPTISTA, C. S.; BARROS, M. H.; BONACCORSI, ED.; BORDIN, S.; BOVÉ, J. M.; BRIONES, M. R.; BUENO, M. R.; CAMARGO, A. A.; CAMARGO, L. E.; CARRARO, D. M.; CARRER, H.; COLAUTO, N. B.; COLOMBO, C.; COSTA, F. F.; COSTA, M. C.; COSTA-NETO, C. M.; COUTINHO, L. L.; CRISTOFANI, M.; DIAS-NETO, E.; DOCENA, C.; EL-DORRY, H.; FACINCANI, A. P.; FERREIRA, A. J.; FERREIRA, V. C.; FERRO, J. A.; FRAGA, J. S.; FRANÇA, S. C.; FRANCO, M. C.; FROHME, M.; FURLAN, L. R.; GARNIER, M.; GOLDMAN, G. H.; GOLDMAN, M. H.; GOMES, S. L.; GRUBER, A.; HO, P. L.; HOHEISEL, J. D.; JUNQUEIRA, M. L.; KEMPER, E. L.; KITAJIMA, J. P.; KRIEGER, J. E.; KURAMAE, E. E.; LAIGRET, F.; LAMBAIS, M. R.; LEITE, L. C.; LEMOS, E. G.; LEMOS, M. V.; LOPES, A. S.; LOPES, C. R.; MACHADO, J. A.; MACHADO, M. A.; MADEIRA, A. M.; MADEIRA, H. M.; MARINO, C. L.; MARQUES, M. V.; MARTINS, E. A.; MARTINS, E. M.; MATSUKUMA, A. Y.; MENCK, C. F.; MIRACCA, E. C.; MIYAKI, C. Y.; MONTERIRO-VITORELLO, C. B.; MOON, D. H.; NAGAI, M. A.; NASCIMENTO, A. L.; NETTO, L. E.; NHANI JUNIOR, A.; NOBREGA, F. G.; NUNES, L. R.; OLIVEIRA, M. A.; OLIVEIRA, M. C. de; OLIVEIRA, R. C. de; PALMIERI, D. A.; PARIS, A.; PEIXOTO, B. R.; PEREIRA, G. A.; PEREIRA JUNIOR, H. A.; PESQUERO, J. B.; QUAGGIO, R. B.; ROBERTO, P. G.; RODRIGUES, V.; ROSA, A. J. de M. ; ROSA JUNIOR, V. E. de; SÁ, R. G. de; SANTELLI, R. V.; SAWASAKI, H. E.; SILVA, A. C. da; SILVA, A. M. da; SILVA, F. R. da; SILVA, W. A.; SILVEIRA, J. F. da; SILVESTRI, M. L.; SIQUEIRA, W. J.; SOUZA, A. A. de; SOUZA, A. P. de; TERENZI, M. F.; TRUFFI, D.; TSAI, S. M.; TSUHAKO, M. H.; VALLADA, H.; VAN SLUYS, M. A.; VERJOVSKI-ALMEIDA, S.; VETTORE, A. L.; ZAGO, M. A.; ZATZ, M.; MEIDANIS, J.; SETUBAL, J. C. The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis. **Nature**, v. 406, n. 6792, p. 151-159, July. 2000. DOI: https://doi.org/10.1038/35018003.

TKACZ, A.; POOLE, P. Role of root microbiota in plant productivity. **Journal of Experimental Botany**, v. 66, n. 8, p. 2167-2175, Apr. 2015. DOI: https://doi.org/10.1093/jxb/erv157.

WILHELM, R. C.; VAN ES, H. M.; BUCKLEY, D. H. Predicting measures of soil health using the microbiome and supervised machine learning. **Soil Biology and Biochemistry**, v. 164, 108472, Jan. 2022. DOI: https://doi.org/10.1016/j.soilbio.2021.108472.

YUAN, J.; WEN, T.; ZHANG, H.; ZHAO, M.; PENTON, C. R.; THOMASHOW, L. S.; SHEN, Q. Predicting disease occurrence with high accuracy based on soil macroecological patterns of Fusarium wilt. **The ISME Journal**, v. 14, n. 12, p. 2936-2950, Dec. 2020. DOI: https://doi.org/10.1038/s41396-020-0720-5.

CGPE 18988