SHORT COMMENTARY

Brazilian Microbiome Project: Revealing the Unexplored Microbial Diversity—Challenges and Prospects

Victor Satler Pylro · Luiz Fernando Wurdig Roesch · José Miguel Ortega ·
Alexandre Morais do Amaral · Marcos Rogério Tótola · Penny Ruth Hirsch ·
Alexandre Soares Rosado · Aristóteles Góes-Neto · Artur Luiz da Costa da Silva ·
Carlos Augusto Rosa · Daniel Kumazawa Morais · Fernando Dini Andreote ·
Gabriela Frois Duarte · Itamar Soares de Melo · Lucy Seldin · Márcio Rodrigues Lambais ·
Mariangela Hungria · Raquel Silva Peixoto · Ricardo Henrique Kruger ·
Siu Mui Tsai · Vasco Azevedo ·
The Brazilian Microbiome Project Organization Committee

Abstract The Brazilian Microbiome Project (BMP) aims to assemble a Brazilian Metagenomic Consortium/Database. At present, many metagenomic projects underway in Brazil are widely known. Our goal in this initiative is to co-ordinate and standardize these together with new projects to come. It is estimated that Brazil hosts approximately 20 % of the entire world's macroorganism biological diversity. It is 1 of the 17 countries that share nearly 70 % of the world's catalogued animal and plant species, and is recognized as one of the most megadiverse countries. At the end of 2012, Brazil has joined GBIF (Global Biodiversity Information Facility), as associated member, to improve the access to the Brazilian biodiversity data in a free and open way. This was an important step toward increasing international collaboration and clearly shows the commitment of the Brazilian government in directing national policies toward sustainable development. Despite its importance, the Brazilian microbial diversity is still considered to be largely unknown, and it is clear that to maintain ecosystem dynamics and to sustainably manage land use, it is crucial to understand the biological and functional diversity of the system. This is the first attempt to collect and collate information about Brazilian microbial genetic and functional diversity in a

Further information and people involved: Brazilian Microbiome Project—http://brmicrobiome.org/ and links therein.

V. S. Pylro (✉) · M. R. Tótola · D. K. Morais
Microbiology Department, Federal University of Vicosa, Vicosa, Minas Gerais 365700-000, Brazil
e-mail: victor.pylro@ufv.br

L. F. W. Roesch
Federal University of Pampa, Sao Gabriel, Rio Grande do Sul 97300-000, Brazil

J. M. Ortega
Biochemical and Immunology Department, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais 31270-010, Brazil

A. M. do Amaral
EMBRAPA LabEx Programme, Plant Biology and Crop Science, Rothamsted Research, Harpenden AL5 2JQ, UK

V. S. Pylro · P. R. Hirsch
Department of AgroEcology, Rothamsted Research, Harpenden AL5 2JQ, UK

A. S. Rosado · L. Seldin · R. S. Peixoto
General Microbiology Department, IMPPG, Federal University of Rio de Janeiro, Rio de Janeiro 21941-590, Brazil

A. Góes-Neto
Biological Science Department, State University of Feira de Santana, Feira de Santana, Bahia 44036-900, Brazil

A. L. da Costa da Silva
Genetic Department, Federal University of Pará, Belem, Pará 66075-900, Brazil

C. A. Rosa
Microbiology Department, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil

F. D. Andreote · M. R. Lambais
Soil Department, Luiz de Queiroz College of Agriculture, Piracicaba, Sao Paulo 13418-900, Brazil

systematic and holistic manner. The success of the BMP depends on a massive collaborative effort of both the Brazilian and international scientific communities, and therefore, we invite all colleagues to participate in this project.

It is estimated that Brazil hosts approximately 20 % of the entire world's macro-organism biological diversity. It is 1 of the17 countries that share nearly 70 % of the world's catalogued animal and plant species, and is recognized as one of the most megadiverse countries [2, 4, 6, 7]. This biodiversity is often related to the great climate variability and geomorphology, in a country covering more than 8.5 million km^2. Biological diversity is an important resource, not only for the environmental services it currently provides but also for its potential as a resource in the development of new and sustainable ecosystem management tools and opportunities for bioprospecting. At the end of 2012, Brazil joined the GBIF (Global Biodiversity Information Facility), as an associated member, to improve the access to the biodiversity data in a free and open way. This was an important step for increasing international collaboration aimed at consolidating knowledge of Brazil's biodiversity and indicates the interest of the Brazilian government in making a firm commitment to direct national policies toward sustainable development. Furthermore, currently, there is a strong effort to create a national database of biodiversity, carried out by the Brazilian Ministry of Science, Technology and Innovation, named *Sistema de Informação sobre a Biodiversidade Brasileira—SiBBr* (Information System of Brazilian Biodiversity) (http://www.sibbr.gov.br/). Its purpose is to support development of further scientific research above and beyond the public policies that are currently subsidized.

Despite its importance, Brazilian microbial diversity is still considered to be largely unknown [1], and it is clear that to maintain ecosystem dynamics and to manage land use sustainably, it is crucial to understand the biological and functional diversity of the system. The anticipated strategic and economic

G. F. Duarte
Biological Science Department, Federal University of Ouro Preto, Ouro Preto, Minas Gerais 35400-000, Brazil

I. S. de Melo
EMBRAPA Environment, Jaguariuna, Sao Paulo 13820-000, Brazil

M. Hungria
EMBRAPA Soja, Londrina, Parana 86001-970, Brazil

R. H. Kruger
Brasilia University, Brasilia, Federal District 70910-900, Brazil

S. M. Tsai
Center for Nuclear Energy in Agriculture, Piracicaba, Sao Paulo 13416-000, Brazil

V. Azevedo
General Biology, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil

benefits are related to the discovery of microorganisms for use as a source of commercially exploitable products. The characterisation of genes and metabolic pathways encoding key biocatalysts, as yet unknown, may support programs focused on the application of biosynthetic or biodegradation processes including, for example, the degradation of pollutants, synthesis of biofuels, and production of new medicines. This possibility opens up new opportunities for the exploration of molecular bioprospecting in Brazil. The knowledge generated by this process is a key resource to create a fast and efficient strategy for scientific and technological development. The big challenge for the rational use of Brazilian biodiversity is undoubtedly how to transform a huge natural genetic heritage into biotechnological richness. Targets should include communities associated with specific hosts, both plants and animal, some unique to Brazil, as well as terrestrial and aquatic environments. Recent advances in DNA/RNA extraction procedures, next generation sequencing (NGS), metagenomics, and metatranscriptomics, among others, are allowing comparative analysis of the diversity, abundance and important ecosystem functional genes of whole microbial communities at far greater depths than ever before. Due to these recent advances, we believe that there is a great opportunity to study the Brazilian microbiome using state-of-the-art molecular methodologies together with more conventional analyses.

The Brazilian Microbiome Project (BMP) aims to assemble a Brazilian Metagenomic Consortium/Database. At present, many metagenomic projects underway in Brazil are widely known (some examples are shown in Table 1) (for a better view on prior efforts, please refer to [1]). Our main goal is to co-ordinate and standardize these, together with future projects. As proposed previously for large, genome-scale community research projects [9], this paper is a marker by the BMP and is intended to clarify the scope of the initiative and help researchers decide whether to participate. The initial endeavors of the BMP have resulted in the *I BMP Meeting*, held in December 2012 in Viçosa, Minas Gerais, Brazil. The event brought together researchers from cutting-edge institutions and technical experts from companies, during a day of lectures and discussion that formulated the major guidelines, initial goals, requirements, and challenges for the project. Furthermore, one of the most important discussion points was the creation of a so-called National Institute of Science and Technology: Microbiome (*INCT* in Portuguese) initiative, as suggested by the administration of the Agricultural, Biological Sciences division of CNPq (National Council for Scientific and Technological/Brazil). *INCTs* are expected to occupy a strategic position in the National System of Science and Technology in Brazil, by developing consistent and high-priority research programs for the scientific and technological improvements. In this context, *INCTs* are planned to release calls for specific research proposals for both, microbiome-related and educational/scientific dissemination projects. Together with its International Advisory Board, the BMP

Table 1 Examples of microbiome-related project currently being carried out in Brazil

Institution	Research Group	Project
EMBRAPA Soja	Laboratory of Soil Biotechnology	Soil microbiome related to crop rotation (soy, corn, and wheat) in tillage system
		Herbicide effects
		Comparative microbiome of soils under soybean cultivation (transgenic and non-transgenic) at Paraná State
		Microbiome of Cerrado's crop systems
UFRJ	Bioinovar—Bioremediation Unity Molecular Microbial Ecology Laboratory Laboratory of Microbial Genetics	Microbiome of Brazilian corals
		Microbiome of sediments and rhizosphere at mangrove
		Freshwater, marine and hypersaline aquatic microbiomes from Southeast region
		Comparative analysis of soil microbiome, rhizosphere, and litter at many agro-forestry crop systems
		Mixed forest system microbiome (eucalyptus and legumes)
		Antarctic rhizosphere and soil microbiome
CENA/USP	Cellular and Molecular Biology Laboratory	Black soil of Amazon microbiome
		Rhizosphere microbiome—legume
		Soil microbiome under sugar cane cultivation
		Mangrove soilsmicrobiome
		Atlantic rainforest soil microbiome. Santa Catarina State
FIOCRUZ/UEFS	Bioinformatics Excellence Center	Metagenome of decaying wood at Atlantic Rainforest Biome
ESALQ/USP	Soil Microbiology Laboratory	Microbiome of mangrove sediments
		Microbiome of sugar cane cultivated soils
USP	Oceanographic Institute	Microbiome of extreme environments (Antarctic)
EMBRAPA Cerrados	Soil Microbiology and Biochemistry Laboratory	Cerrado soils microbiome under native vegetation
		Functional evaluation of Cerrado soils microbiome (Federal District)
EMBRAPA Aves e Suínos	Genomic Applied to Poultry and Swine	Metagenomic analysis of rumen
LNCC	Bioinformatics' Laboratory	Respiratory tract of pigs microbiome Metagenome Northeast
EMBRAPA Meio Ambiente	Environmental Microbiology Laboratory	Rhizosphere metagenome of *Mimosa tenuiflora* caatinga
		Matagenome of the Brazilian tropical dry forest
		Rhizosphere microbiome of Mandacaru (*Cereus jamacaru*)
		Evaluation of functional rumen microbiome
UNIPAMPA	Center for Interdisciplinary Research in Biotechnology	Microbiome of Pampa
		Microbiome of Cerrado
		Microbiome of Antartica
UFV	Laboratory of Environmental Biotechnology and Biodiversity	Microbiome of Trindade Island
		Functional analysis of Trindade Island microbiome under petroleum contamination

Organization Committee will further discuss the priorities and demands of these calls.

The BMP database assemblage will be based on seven main steps, as follows: (1) Collate information regarding all microbiome-based studies in Brazil (or Brazilian samples); (2) Evaluate all methods applied so far in each study, and further demand priorities; (3) Evaluate selected methods by using actual samples; (4) Select and standardize methods and data analyses; (5) Design the BMP database (sequences and related metadata); (6) Develop friendly tools for data analyses; and (7) Implement, monitor and adjust the database. We expect to release the first datasets as soon as the database structures are in place.

Our challenge includes the development/dissemination of a change of paradigm related to the standards for experimental

design and data analysis as well as integration with existing consortia and standards (e.g., Earth Microbiome Project—http://earthmicrobiome.org/). This will enable the comparison of diverse Brazilian microbiomes with others from around the world, to identify significant commonalities and differences. However, to estimate the "true" composition and diversity of any microbiome, it is essential to obtain sufficient coverage with appropriate sampling depth and replication. Unfortunately, this has not often been the case, and it is frequent to find papers comparing microbial diversity from single samples without proper replications [8]. It may be more important to include biological replicates than great depth of sequencing to detect major patterns of variation among microbial communities [3, 5].

One of the goals of the BMP is to understand interactions between organisms in the same niche and how communities are affected by different stresses including season, climate, or the health status of host organisms (human, animal, or plant). Moreover, the database is intended to contain all informative metadata related to changes at all taxonomic levels associated with the relevant microbiome. Clustering algorithms will be developed to relate the presence of clades within each microbiome, from species to phyla, within the various habitats that comprise a constantly enlarging sampling project. Functional characteristics are likely to be of crucial importance, thus developing functional annotation is a likely challenge for the project. A major effort is planned to propagate gene ontology terms and functional annotation along all characterized gene coding sequences, including distant orthologues, shared domain architectures and even protein structural similarities. Understanding the interactions between different cohabiting organisms and their environment is the first step in a full characterization of the system. The interaction between an organism or a given clade (e.g., genus, family, or even phylum) with the environmental/associated biota will be addressed. Currently, there is an effort focused on the further integration of both BMP and *SiBBr* databases. In a first approach, the BMP projects are intended to use SiBBr data as a metadata in the microbe-based analyses, to anticipate the influence of macro-organisms on the microbial community. The second step will be to encompass microbial analyses in all macro-organism-based studies underway in Brazil, allowing a deeper understanding regarding interactions among macro-diversity, micro-diversity, and the environment. We have already organized a categorized data, such as: (1) orthologous groups of genes that have been enriched by means of database interaction (e.g., connecting KEGG to Uniprot clusters); and (2) local databases with information of metabolites that can be either produced or obtained from the environment by an organism. Thus, one can select all enzymes from this organism and determine the reactions that it is capable of catalyzing, ending up with a list of compounds either produced or

consumed. These reactions are often grouped in pathways, so depicting those of interest either by frequency, or even by focused interest, could be a feasible task. Moreover, if a group of orthologous genes is available, it will be possible to recover the list of enzymes that handle these reactions and from the taxonomic data, determine in which clades they occur. Furthermore, from the database, a collection of curated sequences can be easily screened from the metagenomic data, and then, analysis may start even before data is collected. Simultaneous data collection and analysis will offer a breakthrough in the efficient analysis of environmental sequences.

Aside from collecting metadata together with taxonomic and functional information, a huge challenge for BMP will be to develop novel techniques and sampling approaches aimed to better characterize individuals within communities. For example, the probability of assembling *contigs* from metagenomic sequences is directly dependent on their abundance, then predicting the depth of sampling needed is a field for creative research. One powerful goal is to be able to provide real-time feedback, via a dynamic database, to the participating groups including field workers to ensure that sampling is sufficient, i.e., it will be consistent and never excessively overlapped.

Together with the sequencing effort, BMP shall bring together expertise on additional molecular characterisation of the microbiome, aiming to provide the best sampling schema for next generation sequencing as the technologies evolve. Sequential sampling of the target microbiomes in specific environments will support BMP system biology. Sharing expertise in areas ranging from 16S rRNA characterization and generation of probe genes to relevant enzyme and other functional assays will enhance the ability of the BMP crews, coupled with the interactive real-time analysis, to produce valuable sampling with continuity and focused on the goal of characterizing co-habitation. The BMP will be shaped in order to increase, disclose, and stimulate the knowledge regarding the Brazilian microbiome. For this reason, satellite laboratories of both data generation and data analysis (bioinformatics) are planned to improve the training of new BMP groups besides local infrastructure and human resources of partners' labs/groups already active. This training effort will improve the interpretation of data and support the development of subject-focused protocols and user-friendly interfaces. However, problems such as the lack of sequencing facilities and the geographical separation of researchers in a large country like Brazil provide a challenge for the BMP in developing this collaborative multidisciplinary project.

To overcome these barriers, BMP will also reinforce networking with Brazilian Societies such as the Brazilian Association on Bioinformatics and Computational Biology (AB3C), Brazilian Genetics Society (SBG), Brazilian Society of Microbiology

(SBM), Brazilian Society of Soil Science (SBCS), and others, to better develop requirements to be included in the database and collaborative agreements to provide data to the scientific community. Regular participation by BMP in meetings of these societies will improve awareness and promote collaborations.

BMP is also concerned about legislation, economic repercussion, intellectual property rights, and social issues related to bioprospecting. In this regard, we are integrating specialized companies, to help us on related issues. Moreover, all projects included in the BMP will need to be in agreement with the Authorization for Access to Genetic Resources (Bioprospecting/ Technological Development), as requested by the Brazilian legislation. Data-sharing policies are one of the major concerns for research participants, and it is now in discussion. In order to stimulate international collaboration, all microbial nucleotide sequence and associated metadata will be previously stored into a controlled-access database, allowing BMP members to get ahead start of the data intellectual exploitation. However, these data will be further released into a freely and publicly accessible database. This is important to increase the scientific utility of the data.

Currently, the BMP is organized in a committee composed by researchers from all Brazilian regions and by an International Advisory Board (for details, see http://www.brmicrobiome.org). Notwithstanding the challenges and objectives above, this is the first attempt to collect and collate information about Brazilian microbial genetic and functional diversity in a systematic and holistic manner. New sequence data have been generated from samples collected in all Brazilian regions; however, the success of the BMP depends on a massive collaborative effort of both the Brazilian and international scientific communities. Therefore, we invite all colleagues to participate in this project. There is no prioritization of specific taxonomic groups, studies could include any ecosystem, and all proposals and any help will be very welcome.

References

1. Bruce T et al (2012) Microbial diversity of Brazilian Biomes. In: Nelson KE, Jones-Nelson B (eds) Genomics applications for the developing world. Springer, Berlin, pp 217–247
2. Convention on Biological Diversity. Brazil—Country profile/ Overview. http://www.cbd.int/countries/profile/default.shtml?country=br. Accessed 13 August 2013
3. Hamady M, Knight R (2009) Microbial community profiling for human microbiome projects: tools, techniques, and challenges. Genome Res 19(7):1141–1152
4. Landim MIPF, Hingst-Zaher E (2010) Brazil's biodiversity crisis. The International Council of Museums Magazine, Paris, 63(2): p. 14–15
5. Lemos LN et al (2011) Rethinking microbial diversity analysis in the high throughput sequencing era. J Microbiol Methods 86(1):42–51
6. Mittermier RA et al (2004) Hotspots revisited: earth's biologically richest and most endangered terrestrial ecoregions. CEMEX &Agrupacion Sierra Madre, Cidade do México, p 392
7. MMA (1998) First National Report for the Convention on Biological Diversity - Brazil Clearing-house Mechanism for CDB. 283 p. Available on: http://www.mma.gov.br/component/k2/item/7927. Accessed 13 August 2013
8. Prosser JI (2010) Replicate or lie. Environ Microbiol 12(7):1806–1810
9. The Wellcome Trust (2003) Sharing Data from Large-Scale Biological Research Projects: A System of Tripartite Responsibility. Available on: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd003207.pdf. Accessed 27 August 2013